Bias reduction and robustness for gaussian longitudinal data analysis

Elena Bortolato[1] and Euloge Clovis Kenne Pagui[2]

[1]University of Padova, Department of Statistical Sciences, [2]University of Oslo, Deptartment of Biostatistics
[1]elena.bortolato.1@phd.unipd.it (corresponding author)
[2]e.c.k.pagui@medisin.uio.no

Abstract

In the analysis of clustered or longitudinal data structures, such as those coming from surveys or trials, it is desirable to have accurate estimates of parameters relative to the dependence within repeated measurements. In this context, marginal models together with the generalized estimating equation approaches allow to consistently estimate regression coefficients. Inference may be problematic for small and moderate sample sizes, or with models involving many covariates. Moreover, the correlation structure may be misspecified and even the empirical sandwich covariance matrix estimator has limited effectiveness when the number of subjects is small. In this paper, we focus on gaussian with exchangeable correlation matrix model and we propose to use a suitable adjustment of the score function aiming at mean and median bias reduction of maximum likelihood estimates. Extensive simulation studies show a remarkable performance of the proposed methods. In addition, we show that the bias reduction methods maintain an appreciable robustness with respect to the traditional maximum likelihood and generalized estimating equations when the covariance structure is misspecified.

Some key words: Adjusted scores, Bias reduction; Exchangeable correlation matrix; Generalized estimating equations; Misspecified covariance.

1 Introduction

To conduct inference with multivariate responses models, that is when several observations ($K > 1$) , are linked to each unit $y_i = (y_{i_1}, ...y_{i_K})$, $i = 1, \ldots, n$, it is natural to allow for possible dependence structure among components. In this framework, one popular approach is based on Generalized Estimating Equations (GEE) by Liang and Zeger (1986). In general, observations taken repeatedly on each unit are assumed to be correlated and observations on different subjects are considered independent. It is well known that with this approach, the regression coefficient estimators are consistent and unbiased while their covariance matrix depends on variance and correlation parameters that in turn may be subject to the bias in situation of moderate sample size. As a result, standard errors may be biased and lead to misleading inferential conclusions.

In many applications, when the data are collected on a real and continuous scale, a multivariate Gaussian distribution could often be a suitable candidate. Hence, the GEE models coincides with the marginal model and the standard approach to inference can be fully based on maximum likelihood (ML).The method has desirable asymptotic properties even though it may deteriorate when the sample size is finite, or in situation in which the model involves many covariates. Despite the fact that the ML estimators of the mean parameters are unbiased, the variance and correlation parameters may result in substantial bias and therefore compromise the accuracy of the inference. This affects not only the conclusions on the covariance and correlation parameters theirselves, but especially the standard errors and confidence Wald type intervals of regression coefficients.

We point out that, a delicate aspect to take into consideration concerns the specification of the correlation matrix structure. In the literature, the most often assumed forms for the "working" correlation matrix are independence, exchangeable, first order autoregressive (AR-1) and unconstrained. Usually, it is not an easy task to identify a priori which form of correlation structure is appropriate for the underlying model. That difficulty is overcome by using the sandwich information to compute the standard errors. The latter are expected to be robust under misspecification of the correlation matrix. Some works have highlighted weaknesses of this approach as well. For instance Wang and Carey (2003) observed that the sandwich estimator is indeed not robust when the sample size is not large enough. Guo et al. (2005) compared the robust Wald test and the robust score test empirically noticing that one tends to be too liberal, whereas the other is conservative. The authors proposed a suitable modification of the score test for the special case of comparison of two groups. To achieve better efficiency and coverage performance with small sample size, Wang and Long (2011) combined the ideas of Pan (2001), who considers under mild conditions some pooling strategy for the variance estimation, and of Mancl and DeRouen (2001), who propose an alternative covariance estimator to the robust covariance estimator of GEE. More recently, Ford and Westgate (2018) compared a series of proposed methodologies to deal with the problem.

In this paper, we focus on the improved estimation of the Gaussian multivariate model with exchangeable correlation structure, based on adjustments of the likelihood equation that have been proposed starting from the contributions of Firth (1993) and Kenne Pagui et al. (2017), resulting in mean or median bias reduction, respectively. While mean bias reduction yields an estimator with reduced bias, median bias reduction is such that each component of the estimator is, with high accuracy, median unbiased, that is, it has the same probability of underestimating and overestimating the corresponding parameter component We also propose the exploration of the traditional bias corrected (BC) approach that subtracts an estimate of the bias from the maximum likelihood estimate, see e.g. Section 9.2 of Cox and Hinkley (1974). Moreover, under misspecification of the correlation matrix, we propose a suitable adjustment of the likelihood equation by the improved Jeffreys penalization (Lima and Cribari-Neto, 2019). For the underlying model, we obtain the quantities required for mean and median BR, mean BC and the improved Jeffreys penalization, as well as a development of `R` code, which is not always straightforward, but is necessary in order to make these methods available to practitioners.

The numerical studies indicate that, overall, bias correction and reduction are both preferable to standard likelihood inference, especially with moderate sample sizes and large number of covariates. Improvement trough the Jeffreys penalized likelihood provides comparable results to mean and median BR, both in terms of coverages of Wald confidence intervals and robustness under misspecification of the correlation structure when the sample size is relatively small. Moreover, the results indicate that inference based on the adjusted scores for the multivariate Gaussian with exchangeable structure is in general robust under different possible misspecification of the correlation matrix. The methods have been applied on a real dataset.

The rest of this paper is organized as follows. In the next section, we introduce the notation and review the basic background on bias reduction methods. Section 3 illustrates our contribution in obtaining the required quantities in the Gaussian multivariate model with exchangeable correlation structure. Extensive simulation studies are presented in Section 4. Section 5 focuses on an application. Finally, a brief discussion is given in Section 6.

2 Bias-reducing adjusted scores

Consider a regular model with d-dimensional parameter θ, log-likelihood $\ell(\theta)$ and score function $U(\theta)$. We let $U_r(\theta)$ be a component of $U(\theta)$, $j(\theta) = -\partial^2\ell(\theta)/\partial\theta\partial\theta^T$ be the observed information and $i(\theta) = E_\theta[j(\theta)]$ the expected information. Moreover, the maximum likelihood estimate denoted with $\hat{\theta}$ is a solution of the equation $U(\theta) = 0$. Under random sampling, the bias expansion of the ML estimator has form $E_\theta[\hat{\theta} - \theta] = b(\theta) + O(n^{-2})$, where $b(\theta) = i(\theta)^{-1}A^*(\theta)$ with $A^*(\theta)$ having components

$$A_r^*(\theta) = \frac{1}{2}\text{tr}\{i(\theta)^{-1}[P_r(\theta) + Q_r(\theta)]\},$$

where $P_r(\theta)$ and $Q_r(\theta)$ are $d \times d$ matrices defined as

$$P_r(\theta) = E[U(\theta)U(\theta)^T U_r(\theta)], \quad Q_r(\theta) = E[-j(\theta)U_r(\theta)],\ r = 1, \ldots, d,$$

The quantities $P_r(\theta)$ and $Q_r(\theta)$ are first used in Kosmidis and Firth (2010).

The mean bias corrected estimator, see e.g. Section 9.2 of Cox and Hinkley (1974), Section 5.3 of Barndorff-Nielsen and Cox (1994) and Cordeiro and McCullagh (1991), is obtained as $\theta^\dagger = \hat{\theta} - b(\hat{\theta})$. Firth (1993) proposed an adjusted score of form

$$U^*(\theta) = U(\theta) + A^*(\theta),$$

with $A^*(\theta)$ a model-dependent adjustment term of order $O(1)$, which is built in such a way that $b(\theta)$ is implicitly removed. The resulting mean BR estimator, θ^*, solution of $U^*(\theta) = 0$, has bias of order $O(n^{-2})$, smaller than that of $\hat{\theta}$, which is $O(n^{-1})$.

When θ is the canonical parameter of a full exponential family, the mean BR estimator θ^* corresponds to the mode of the posterior distribution obtained using Jeffreys prior (Jeffreys, 1946), which is a well known improper prior with invariance property. On the other hand, Jeffreys invariant prior is used in the literature to penalize the likelihood function with aims of solving issues related to boundary estimates that can occur with positive probability in models for discrete data (Kosmidis and Firth (2018)) and in Cox regression (Heinze and Schemper, 2001; Kenne Pagui and Colosimo, 2020). Some authors also proposed small modifications to the Jeffreys adjustment to achieve desirable sampling estimator properties, such as unbiasedness outside of exponential families. In that direction, Lima and Cribari-Neto (2019) suggested to use the penalized log-likelihood of form

$$\ell_{adj}(\theta) = \ell(\theta) + \eta \log |i(\theta)|, \eta \in \mathbb{R},$$

where η ends up being a power to which the prior is elevated. With $\eta = 1$ the original penalization is obtained, while $\eta = 0$ gives the usual log-likelihood. We derive the corresponding estimating equation as

$$\bar{U}(\theta) = \frac{\partial \ell_{adj}(\theta)}{\partial \theta} = U(\theta) + \bar{A}(\theta, \eta).$$

Following a similar idea of Firth (1993) and Kosmidis et al. (2010), Kenne Pagui et al. (2017) developed an adjusted score of form

$$\tilde{U}(\theta) = U(\theta) + \tilde{A}(\theta).$$

The resulting median BR estimator, $\tilde{\theta}$, satisfies in the continuous case the improved median centering property,$Pr_\theta(\tilde{\theta}_r < \theta_r) = 1/2 + O(n^{-3/2}), r = 1, \ldots, d$, in contrast with the corresponding $O(n^{-1/2})$ order of error for the ML estimator. The quantity $\tilde{A}(\theta) = A^*(\theta) - i(\theta)F(\theta)$, where $F(\theta)$ is a vector of components $F_r = [i(\theta)^{-1}]_r^T \tilde{F}_r$, $r = 1, \ldots, p+2$. The vector $\tilde{F}_r$ has elements $\tilde{F}_{r,t} = \text{tr}\{h_r[(1/3)P_t + (1/2)Q_t]\}$, $t = 1, \ldots, p+2$ and the matrix h_r is defined as $h_r = \{[i(\theta)^{-1}]_r[i(\theta)^{-1}]_r^T\}/i^{rr}(\theta)$, where $[i(\theta)^{-1}]_r$ is the r-th column of $i(\theta)^{-1}$ and $i^{rr}(\theta)$ its r-th element. The estimators $\tilde{\theta}$ and θ^* have the same asymptotic distribution as $\hat{\theta}$. In practice, standard errors are computed using diagonal elements of the inverse Fisher information, evaluated at the corresponding estimates, i.e $i(\tilde{\theta})^{-1}$, $i(\theta^*)^{-1}$, $i(\hat{\theta})^{-1}$. We point out that mean and median bias reduction have the disadvantage of being directly applicable only when the score functions, Fisher information and the first order bias term ($b(\theta)$) of the maximum likelihood estimator are available in closed form.

3 Exchangeable multivariate Gaussian model

Longitudinal analyses often involve multiple outcomes measured repeatedly from the same subject or cluster. When the outcome responses are continuous, the multivariate gaussian model with a suitable correlation structure is often assumed. One of the main issues is the correct definition of the dependence among repeated measures. In the literature, most often assumed forms for the "working" correlation matrix are independence, exchangeable, first order autoregressive (AR-1) and unconstrained. Here, we focus on the Gaussian multivariate model with exchangeable correlation structure. In this case the mean and median BR adjustment terms are obtainable in a closed form. We also assume that we have the same fixed number of responses (K) per subjects.

Consider n independent observations from a q-variate normal, $Y_i \sim N_q(\mu_i, V)$, $i = 1, ..., n$, with $\mu_i = X_i\beta$, where X_i is a $q \times p$ design matrix and $\beta =$

$(\beta_1, \dots, \beta_p)$. Let $N = n \times q$, and $Y = (Y_1, \dots, Y_n)^T$, then $Y \sim N_N(\mu, \mathcal{V})$, with $\mu = (\mu_1, \dots, \mu_n)^T \in \mathbb{R}^N$.
In the above, the $N \times N$ block diagonal matrix $\mathcal{V}$ has form

$$\mathcal{V} = \begin{pmatrix} V & 0 & \dots & 0 \\ 0 & V & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & V \end{pmatrix}, \quad \text{with } V = \sigma^2 \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}.$$

The model has in total $p + 2$ parameters: $\theta = (\beta_1, \dots, \beta_p, \sigma^2, \rho)^T$. Denoting by Ω the inverse of $\mathcal{V}$, the log-likelihood is

$$\begin{aligned} \ell(\theta; y) = & -\frac{n}{2}[(q-1)\log(1-\rho) + q\log\sigma^2 + \log(q\rho - \rho + 1)] \\ & - \frac{1}{2}(y - X\beta)^T \Omega (y - X\beta). \end{aligned}$$

The derivatives of the log-likelihood with respect to the components β, σ^2, ρ are

$$\begin{aligned} U_\beta &= X^T \Omega (y - X\beta), \\ U_{\sigma^2} &= \frac{N}{2\sigma^2} - \frac{1}{2}(y - X\beta)^T \Omega_{\sigma^2} (y - X\beta), \\ U_\rho &= -\frac{N(q-1)\rho}{2(\rho-1)(q\rho-\rho+1)} - \frac{1}{2}(y - X\beta)^T \Omega_\rho (y - X\beta), \end{aligned}$$

where $\Omega_{\sigma^2} = \frac{\partial\Omega}{\partial\sigma^2}$ and $\Omega_\rho = \frac{\partial\Omega}{\partial\rho}$. The information matrix has form

$$i(\theta) = \begin{pmatrix} X^T\Omega X & 0 & 0 \\ 0 & \frac{N}{2\sigma^4} & \frac{N(q-1)\rho}{2(\rho-1)(q\rho-\rho+1)\sigma^2} \\ 0 & \frac{N(q-1)\rho}{2(\rho-1)(q\rho-\rho+1)\sigma^2} & \frac{3N(q-1)(q\rho^2-\rho^2+1)}{2(\rho-1)^2(q\rho-\rho+1)^2} \end{pmatrix},$$

with the corresponding inverse given by

$$i(\theta)^{-1} = \begin{pmatrix} (X^T\Omega X)^{-1} & 0 & 0 \\ 0 & \frac{2\sigma^4(q\rho^2-\rho^2+1)}{N} & \frac{2(\rho-1)\rho(q\rho-\rho+1)\sigma^2}{N} \\ 0 & \frac{2(\rho-1)\rho(q\rho-\rho+1)\sigma^2}{N} & \frac{2(q\rho-\rho+1)^2(\rho-1)^2}{N(q-1)} \end{pmatrix}.$$

The maximum likelihood estimator is obtained as a solution of $U_\beta = 0,\ U_{\sigma^2} = 0,\ U_\rho = 0$. It is straightforward to derive the estimator for the β as follows

$$\hat{\beta} = (X^T\Omega X)^{-1} X^T \Omega Y.$$

The ML estimator $\hat{\beta}$ is asymptotically normal, with variance given by $(X^T\Omega X)^{-1}$. The estimates of β, σ^2 and ρ are obtained iteratively. As usual, standard errors are computed using diagonal elements of Fisher information inverse evaluated at the corresponding estimates. Under a specific condition for which the vector subspace generated by the columns of X is the same of that generated by the columns of ΩX, the regression coefficient estimates simplify to $\hat{\beta} = (X^T X)^{-1} X^T Y$. The latter corresponds to the estimates obtained under full independence, see Salvan et al. (2020, Section 1.6.4) for more details.

For the underlying model with exchangeable correlation matrix, after algebra calculations, we provide the expressions of the adjustment terms in a closed form for the proposed approaches. The expectations involved in the estimating equations are available in the Appendix. In particular, the adjustment term components for mean BR are given by $A^*(\theta)$

$$
\begin{aligned}
A^*_\beta &= 0,\\
A^*_{\sigma^2} &= -\frac{2q\rho^2 - 2\rho^2 - p}{2\sigma^2},\\
A^*_\rho &= \frac{-(q-1)(2q\rho^3 - 2\rho^3 - \rho)}{2(\rho-1)(q\rho-\rho+1)} - \frac{1}{2} tr\{(X^T\Omega X)^{-1} X^T \Omega_\rho X\}.
\end{aligned}
$$

The adjustment term components for the median BR is

$$
\begin{aligned}
\tilde{A}_\beta &= 0,\\
\tilde{A}_{\sigma^2} &= A^*_{\sigma^2} + \frac{3q^2\rho^4 - 6q\rho^4 + 3\rho^4 + 6q\rho^2 - 6\rho^2 - q\rho + 2\rho + 1}{3(q\rho^2 - \rho^2 + 1)\sigma^2},\\
\tilde{A}_\rho &= A^*_\rho + \frac{3q^3\rho^5 - 9q\rho^5(q-1) - 3\rho^5 + 9\rho^3(q^2+1)}{3(\rho-1)(q\rho-\rho+1)(q\rho^2-\rho^2+1)}\\
&\quad + \frac{-18q\rho^3 - 2q^2\rho^2 + 6q\rho^2 - 4\rho(\rho-q+1) - q + 2}{3(\rho-1)(q\rho-\rho+1)(q\rho^2-\rho^2+1)}.
\end{aligned}
$$

With a similar derivation, the components of the adjustment term using Jeffreys penalization are as follows

$$
\begin{aligned}
\bar{A}_\beta(\eta) &= 0,\\
\bar{A}_{\sigma^2}(\eta) &= -\eta\frac{(2+p)}{2\sigma^2},\\
\bar{A}_\rho(\eta) &= \eta\frac{4q\rho - 4\rho - 2q + 4}{2(1-\rho)(q\rho-\rho+1)} + tr\{(X^T\Omega\rho X)^{-1}(X^T\Omega_\rho X)\}.
\end{aligned}
$$

We remark that the above results can be generalized only with respect to the case of independence correlation structure, with $\rho = 0$. For a more complex correlation structure, the needed algebra in obtaining the corresponding adjustment terms may be tedious and closed forms even not available. For this reason, we devote major effort in evaluating the robustness of the proposed approaches under alternative specifications of the correlation structure.

4 Simulation study

To illustrate the properties of the proposed estimators, a simulation study was conducted under two different scenarios. In the former (scenario 1), the regression coefficient estimates are equal for all methods, while they are different in the latter (scenario 2). We check the robustness of the proposed methods under possible misspecification of the correlation matrix structure. A total of seven methods are compared: maximum likelihood (ML), bias corrected (BC), mean bias reduction (BR), median bias reduction (MBR), maximum penalized likelihood with Jeffreys power prior (JEF), generalized estimating equations (GEE) and GEE with robust estimation of standard errors (ROB) (Huber et al., 1967). For the maximum penalized likelihood with Jeffreys power prior, we only present results with $\eta = -1/2$, which is the one showing better performance.

The properties of the estimators are assessed by simulating 10 000 samples with different sample sizes $n = 20, 50$ with vector sizes $q = 5, 10$. Performance of estimetors are evaluated in terms of relative bias (RB), $100 \times B/|\beta|$; relative increase in mean squared error from its absolute minimum due to bias (IBMSE), $100 \times B^2/SD^2$; empirical percentage of underestimation (PU), and coverage of nominally 95% Wald-type confidence intervals (WALD). Here, B denotes the absolute bias and SD, its standard deviation. The measure PU is calculated as the proportion of times that the estimate is smaller than the corresponding true parameter value. We consider five covariates with true coefficient values set to $\beta = (2, 0.3, -1, 3, -0.5)$ and $\sigma^2 = 5$ in all scenarios. We fixed $\rho = 0.9$, for both the exchangeable and first order autoregressive correlation matrices. The unstructured (unstr) and cluster (clus) correlation matrices are provided in the Appendix. We underline that in all the settings, even though the correlation matrix is misspecified, the fitting model is based on the exchangeable normal defined in Section 2.

Scenario 1. Consider a regression model with mean as follows

$$\mu_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5}, i = 1, \dots, n,$$

x_{i2} are independent realizations of uniform random variables on (-10,10); x_{i3} independent realizations of an exponential distribution with rate 0.5; x_{i4}

and x_{i5} are independent realizations of Bernoulli random variables with success probabilities of 0.5 and 0.2, respectively. The values of the explanatory variables x_{i2}, x_{i3}, x_{i4} and x_{i5} were held constant throughout the simulations.

In this case, the ML estimator of the regression coefficients is unbiased and does not depend on σ^2 and ρ. Therefore, all the proposed methods provide the same estimates for the regression coefficients. The difference appears on the estimates of σ^2 and ρ. As a result, we only look at the WALD as a performance measure although we anyway report all the others.

The simulation results are displayed in Figures 1 and 2. We note that the MBR, BR, JEF are generally better when the model is correctly specified. Indeed, the empirical coverages are close to 94%, against ML's 89%. When the correlation matrix is misspecified instead, the JEF slightly seems to be more robust. The simulation results about the variance and correlation parameters are presented in Table 1 and fulfill the expected properties. Both mean and median BR estimators perform better than the ML, with respect to the four performance measures. This appears more evident for small sample size.

			q=5				q=10			
			PU	RB	WALD	IBMSE	PU	RB	WALD	IBMSE
$n = 20$		MBR	49.92	4.05	88.22	1.34	49.36	4.50	88.47	1.66
		JEF	51.11	2.76	89.46	0.72	50.53	3.23	89.52	0.99
	σ^2	BR	54.23	0.01	86.58	0	53.89	0.50	86.79	0.02
		BC	61.86	-5.93	83.04	3.59	61.54	-5.43	83.16	3.03
		MLE	82.95	-23.14	64.61	85.42	82.37	-22.55	65.77	81.40
		MBR	49.81	-0.93	91.23	3.87	49.54	-0.85	89.63	3.47
		JEF	52.18	-1.01	93.55	5.45	51.44	-0.87	92.08	4.31
	ρ	BR	44.92	-0.38	89.06	0.71	44.62	-0.36	87.51	0.67
		BC	43.56	-0.28	87.69	0.36	43.58	-0.28	85.96	0.39
		MLE	79.74	-4.93	91.77	67.39	80.29	-4.83	89.34	69.27
$n = 50$		MBR	49.43	1.41	92.70	0.52	50.05	1.25	92.43	0.42
		JEF	48.20	2.03	93.74	1.12	48.50	1.88	93.43	1.00
	σ^2	BR	52.31	0.06	92.03	0	53.11	-0.08	91.72	0
		BC	54.25	-0.89	91.64	0.22	54.96	-1.02	91.22	0.30
		MLE	71.69	-9.18	82.99	28.00	71.74	-9.20	83.17	29.02
		MBR	49.61	-0.27	93.34	1.16	49.80	-0.29	93.62	1.49
		JEF	48.76	-0.19	94.07	0.66	48.69	-0.20	94.16	0.79
	ρ	BR	45.51	-0.02	92.47	0.01	45.82	-0.06	92.79	0.07
		BC	45.11	0	92.33	0	45.38	-0.04	92.59	0.03
		MLE	69.05	-1.55	94.30	31.68	70.59	-1.57	93.31	36.97

Table 1: Scenario 1: simulation results for σ^2, ρ when data are generated using the exchangeable correlation matrix.

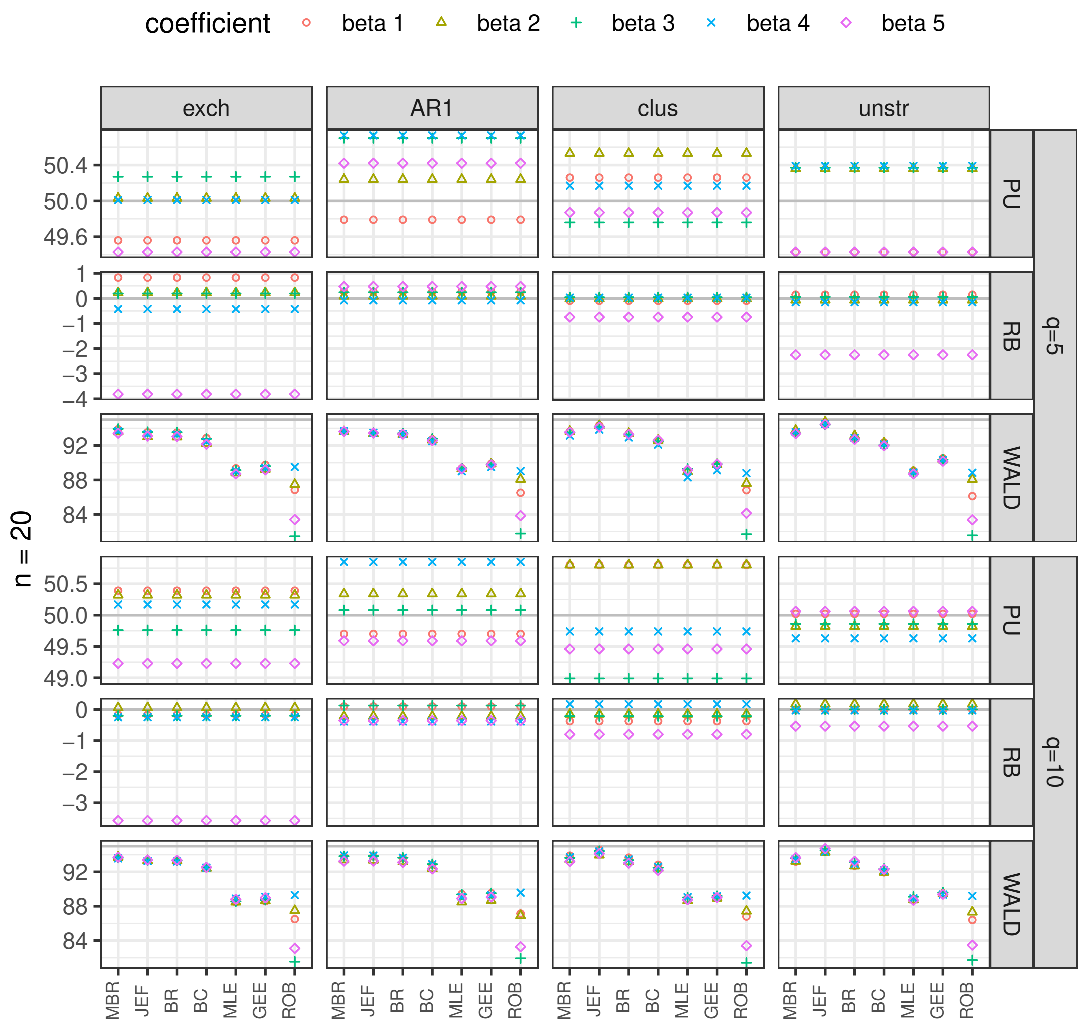

Figure 1: Scenario 1. Simulation study about the regression coefficients $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ with $n = 20$. PU, empirical percentage of underestimation; RB, empirical relative bias; WALD, 95% Wald-type confidence intervals.

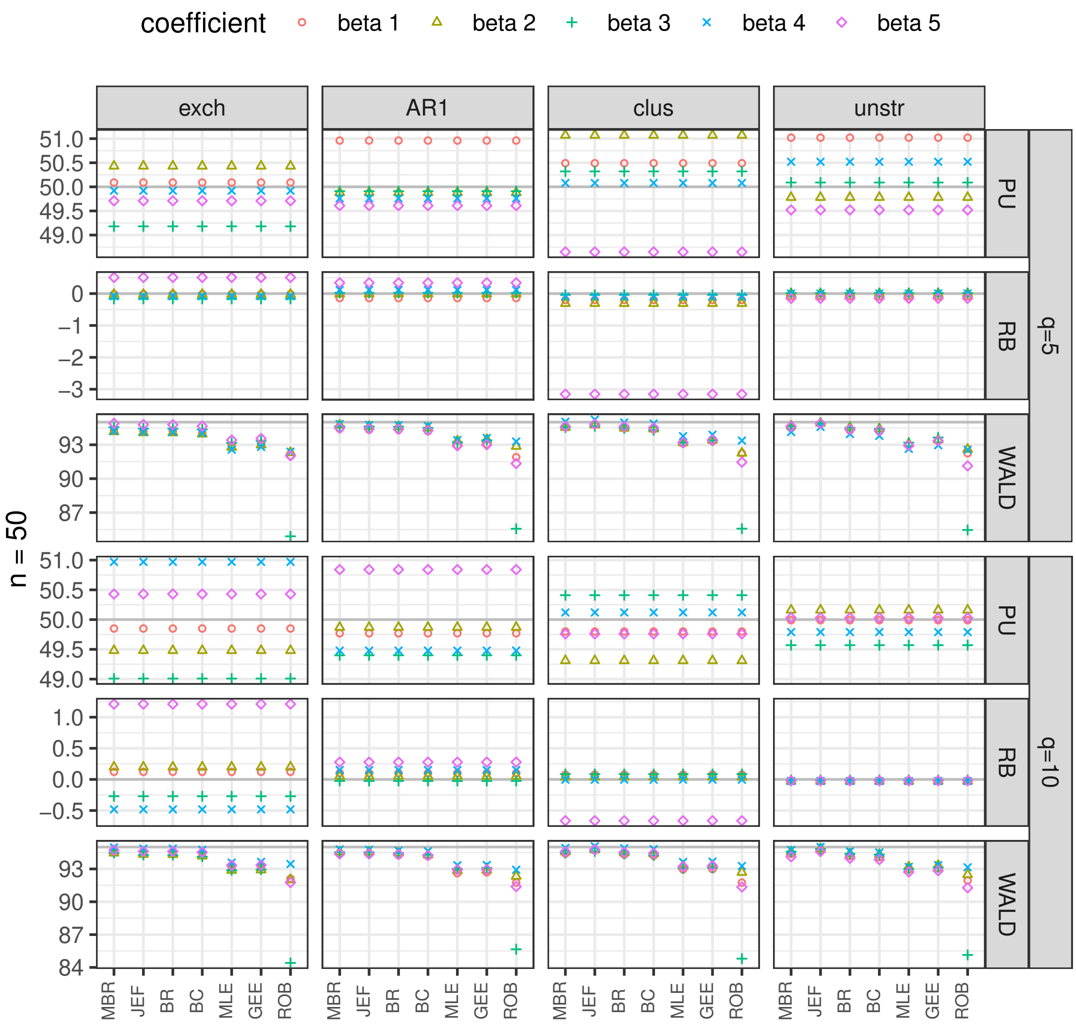

Figure 2: Scenario 1. Simulation results about the regression coefficients $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ with $n = 50$. PU, empirical percentage of underestimation; RB, empirical relative bias; WALD, 95% Wald-type confidence intervals.

Scenario 2. Consider a regression model with mean as follows

$$\mu_{ij} = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{ij4} + \beta_5 x_{ij5},$$

where x_{i2} and x_{i3} are independent realizations of Bernoulli random variables with success probabilities of 0.5 and 0.2, respectively; x_{ij4} are independent realizations of $N(0, 10)$ with non fixed observations within subjects. The covariate $x_{ij5} = j$ is the fixed observation time, for each subject $(j = 1, \ldots, q)$. The true parameter values are the same as in the previous scenario. The choice of the covariates and the true values of the parameters allows the heterogeneity between the subjects. In this case, the estimator of the regression coefficients depend on those of σ^2, ρ. Results are reported in figure 3 and 4. Even here, we note that the WALD confidence intervals provided by the MBR, BR, JEF methods achieve remarkable empirical coverage, closer to the nominal 95are better with respect to those of ML and GEE. The same conclusion is obtained in case of misspecification of the correlation matrix. The proposed methods outperform even the coverages obtained by ROB, for small sample size in all the cases, with exception of the the AR-1 type of misspecification. In the latter, the worst performance of the proposed methods is only on β_5, a coefficient relative to the time varying covariate, $x_{ij5} = j$. In fact, assuming an exchangeable correlation structure instead of the autoregressive type, lead to the underestimation of the corresponding standard error. On the other hand, the robust method works better in correcting that standard error. A summary of the performance on the estimators for σ^2, ρ with the correctly specified model, is given in Table 2.

5 Application to a real dataset: Stroke

We consider the `Stroke` dataset (Dobson e Barnett, 2008), available in the `R` package MLGdata (Sartori et al., 2020) on CRAN. This was collected with the aim of study post-heart attack rehabilitation therapies. Patients were assigned to three experimental groups: A, treated with the innovative therapy; B, treated with traditional therapy in the same hospital as the patients of group A; C, treated with traditional therapy in a different hospital. For each of the 24 patients, 8 measures of functional ability were obtained in consecutive weeks. The study aimed to verify whether treatment A was more effective than the others. The model considered is

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5},$$

where $x_{i1} = 1$ if the subject belongs to the group B, $x_{i2} = 1$ if the subject belongs to the group C, x_{i3} refers to the week, while x_{i4} and x_{i5} are the indicators resulting from the interaction terms between groups and weeks. The

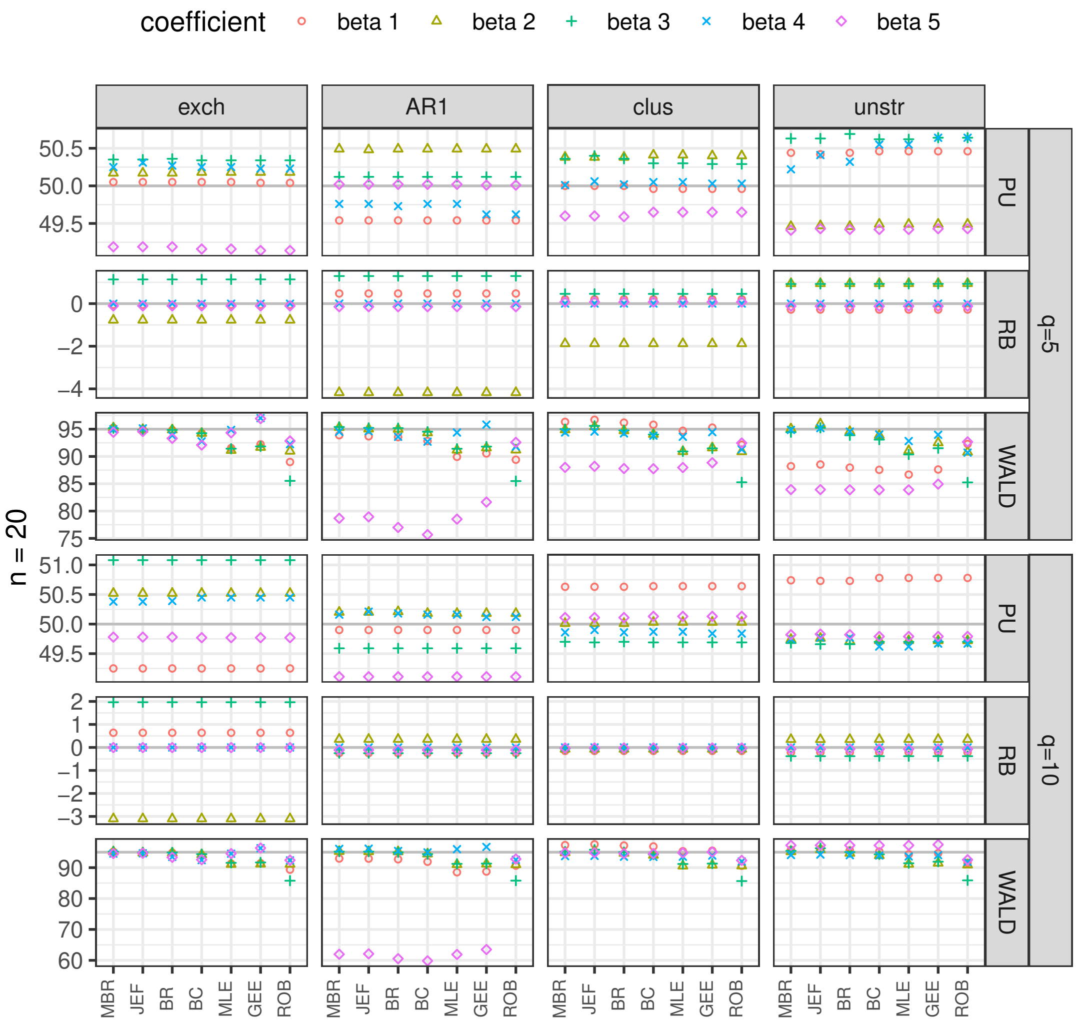

Figure 3: Scenario 2. Simulation results about the regression coefficients $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ with $n = 20$. PU, empirical percentage of underestimation; RB, empirical relative bias; WALD, 95% Wald-type confidence intervals.

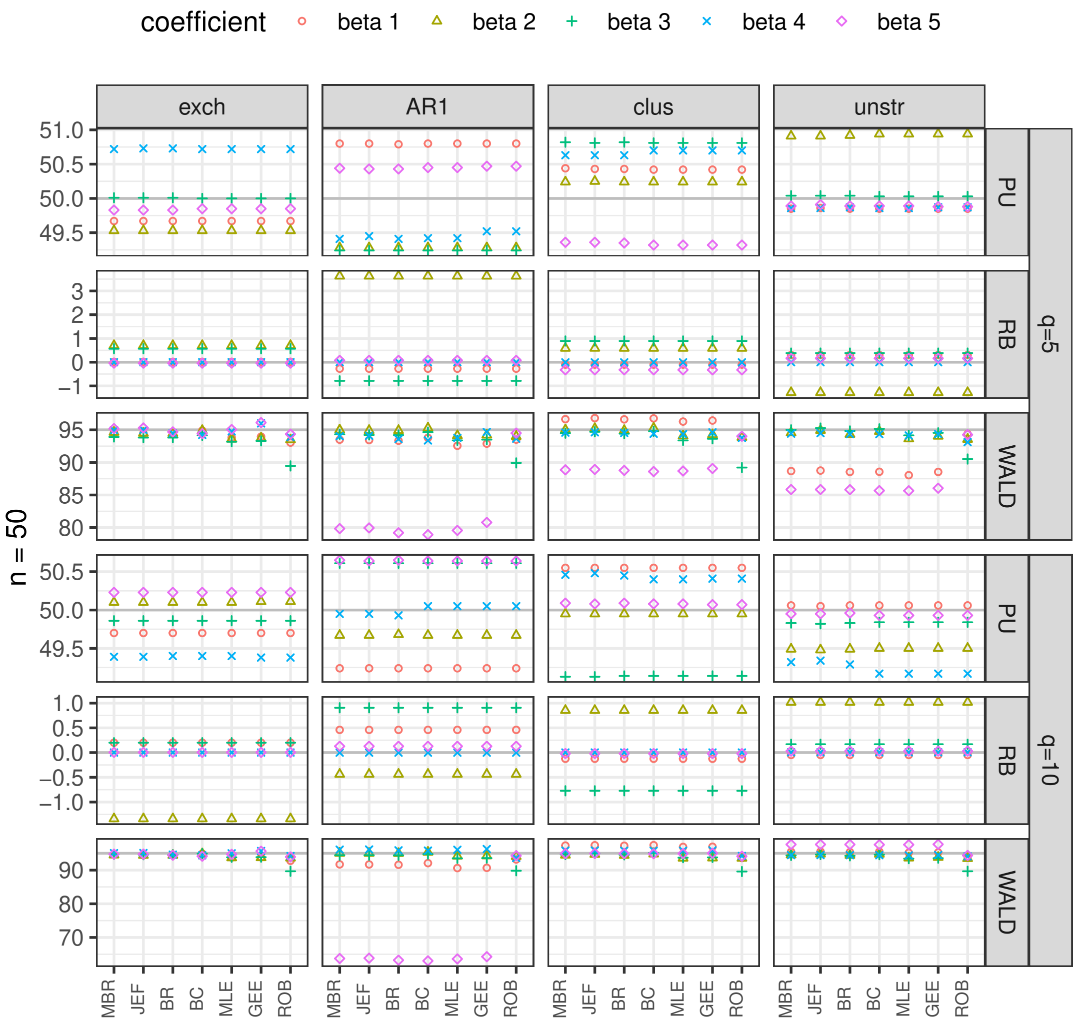

Figure 4: Scenario 2. Simulation results about the regression coefficients $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ with $n = 50$. PU, empirical percentage of underestimation; RB, empirical relative bias; WALD, 95% Wald-type confidence intervals.

				q=5				q=10	
		PU	RB	WALD	IBMSE	PU	RB	WALD	IBMSE
$n = 20$	MBR	51.02	3.10	89.74	0.90	49.92	3.67	89.72	1.26
	JEF	29.96	21.52	96.87	35.91	27.11	23.84	97.20	43.91
σ^2	BR	55.20	-0.52	88.06	0.03	54.51	0.08	88.25	0
	BC	49.01	4.86	90.91	2.12	47.66	5.55	90.87	2.77
	MLE	73.71	-14.56	77.19	29.77	72.70	-13.78	77.43	26.72
	MBR	50.20	-0.86	92.16	3.79	50.38	-0.78	91.85	3.47
	JEF	28.45	1.38	83.23	18.05	26.38	1.45	83.60	22.60
ρ	BR	44.30	-0.26	89.60	0.36	44.49	-0.24	89.95	0.36
	BC	29.41	1.33	79.27	11.92	29.69	1.19	80.24	10.28
	MLE	67.87	-2.82	94.81	31.05	70.33	-2.92	93.71	36.69
$n = 50$	MBR	50.43	1.17	93.30	0.37	50.53	1.08	93.02	0.32
	JEF	36.26	8.09	96.28	16.80	35.48	8.49	96.14	18.67
σ^2	BR	53.14	-0.14	92.63	0.01	53.38	-0.21	92.35	0.01
	BC	46.93	2.88	94.22	2.20	47.03	2.82	93.84	2.13
	MLE	64.90	-5.75	87.74	10.58	65.07	-5.71	87.79	10.57
	MBR	50.76	-0.33	94.55	1.81	50.11	-0.28	94.03	1.55
	JEF	35.14	0.57	91.20	6.95	34.63	0.59	90.85	8.44
ρ	BR	46.41	-0.07	93.60	0.09	46.03	-0.05	93.20	0.05
	BC	36.70	0.49	90.27	4.44	37.45	0.45	90.31	4.24
	MLE	61.51	-1.00	95.24	15.13	63.32	-1.03	94.94	18.08

Table 2: Scenario 2. Simulation results for σ^2, ρ when data are generated using the exchangeable correlation matrix.

results of the fitted model is presented in Table 3. We note that the point estimates of the coefficients are the same for all methods while mean and median BR have comparable standard errors that are different to the rest of the approaches. Table 4 instead reports the results of a simulation study from a maximum likelihood fit. For what concerns estimators of the regression coefficients, these differ just for WALD performance. We recall that $\beta_1, \beta_2, \beta_3$ are coefficients related to time invariant covariates, while $\beta_4, \beta_5, \beta_6$ deal the time varying ones. As it was already pointed out in Section 4, the Wald intervals based on ML improve with the time varying covariates. As a result, the BC estimator has coverages closer to the nominal value for first block, whereas for the second one they are worse than those of ML. On the other hand, BR estimator instead is more stable, with coverages ranging from 93.8 to 94.6. The JEF has coverages closer to the nominal level with range $93.9-95.3$, and it has a similar effectiveness as MBR estimator, whose empirical coverages falls in (94.0 , 95.2). The GEE method, together with the robust one, provide coverages far from the nominal value. We can appreciate the positive effect of BR and BC on the estimators of σ^2 and ρ. Overall the proposed methods are preferable than the traditional approaches and robust version.

	Intercept	Group B	Group C	Week	Group B: Week	Group C: Week	σ^2	ρ
MLE	29.82	3.35	-0.02	6.32	-1.99	-2.69	425.57	0.83
	(7.05)	(9.97)	(9.97)	(0.46)	(0.65)	(0.65)	(104.88)	(0.04)
BC	29.82	3.35	-0.02	6.32	-1.99	-2.69	516.20	0.88
	(7.84)	(11.08)	(11.08)	(0.44)	(0.62)	(0.62)	(133.08)	(0.03)
BR	29.82	3.35	-0.02	6.32	-1.99	-2.69	478.97	0.86
	(7.54)	(10.67)	(10.67)	(0.47)	(0.66)	(0.66)	(123.64)	(0.04)
MBR	29.82	3.35	-0.02	6.32	-1.99	-2.69	492.01	0.85
	(7.61)	(10.76)	(10.76)	(0.47)	(0.66)	(0.66)	(123.64)	(0.04)
JEF	29.82	3.35	-0.02	6.32	-1.99	-2.69	529.59	0.86
	(7.91)	(11.19)	(11.19)	(0.47)	(0.66)	(0.66)	(134.57)	(0.04)
GEE	29.82	3.35	-0.02	6.32	-1.99	-2.69	439.29	0.81
	(7.13)	(10.09)	(10.09)	(0.50)	(0.70)	(0.70)	-	-
ROB	29.82	3.35	-0.02	6.32	-1.99	-2.69	439.29	-
	(10.18)	(11.63)	(10.90)	(1.13)	(1.48)	(1.47)	-	-

Table 3: `Stroke` data: estimates and standard errors in parenthesis.

6 Discussion

For the exchangeable Gaussian model, we developed inference based on adjusted score equations aiming to improve the properties of the traditional maximum likelihood such as mean and median bias. We also proposed a suitable adjustment of the Jeffreys penalization. We derived the required adjustment terms in a closed form for all the proposed methods. In this work, we consider only the model with exchangeable correlation matrix, since other widely used structure of correlation may required cumbersome expectations leading to a not closed form of the resulting estimating equation. In fact, the main drawback of mean and median bias reduction is their limited applicability to models whose the likelihood and related quantities are available in a closed forms. For the underlying model, we studied the behaviour of the proposed methods under misspecification of the correlation structure of the data.

Simulation results confirm the theoretical properties of the methods and indicate that they are effective in improving over standard likelihood inference, especially in situation of small or moderate sample size. In particular, when there are no time varying covariates (as in scenario 1), the proposed approaches are preferable over traditional methods both in terms of bias reduction and robustness When the mean structure involves time varying covariates and the model is misspecified by the autoregressive correlation structure, we observe that the proposed methods fail in terms of robustness for the corresponding associated coefficients. In the latter case, Huber's robust method implemented in gee `R` package shows better results in terms of coverages. As expected, all the methods improve as the sample size increases.

Finally, we note that the adjusted score with Jeffreys penalization acts similarly as mean and median bias reduction. An advantage is due to its simplicity since the adjustment term involves only the Fisher information. On the other side, the important issue concerns the determination of the optimal value for η which is not straightforward. In practice, we suggest to run a parametric bootstrap for different values of η using maximum likelihood estimates as true values and then, choose the one with smaller mean bias. This aspect is subject of the future research.

		MLE	BC	JEF	BR	MBR	GEE	ROB
PU	β_1	50.34	50.34	50.34	50.34	50.34	50.34	50.34
	β_2	50.40	50.40	50.40	50.40	50.40	50.40	50.40
	β_3	49.19	49.19	49.19	49.19	49.19	49.19	49.19
	β_4	50.17	50.17	50.17	50.17	50.17	50.17	50.17
	β_5	49.22	49.22	49.22	49.22	49.22	49.22	49.22
	β_6	48.80	48.80	48.80	48.80	48.80	48.80	48.80
	σ^2	70.64	42.98	51.65	53.50	49.82	66.49	66.49
	ρ	66.68	27.31	52.68	44.42	49.66	80.64	
RB	β_1	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15
	β_2	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09
	β_3	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10
	β_4	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
	β_5	-0.30	-0.30	-0.30	-0.30	-0.30	-0.30	-0.30
	β_6	-0.18	-0.18	-0.18	-0.18	-0.18	-0.18	-0.18
	σ^2	-10.78	7.88	1.50	0.25	2.92	-7.91	-7.91
	ρ	-3.61	2.53	-1.34	-0.22	-0.94	-5.78	
WALD	β_1	92.36	94.75	93.90	93.79	94.02	92.62	89.31
	β_2	92.89	95.35	94.62	94.53	94.75	93.20	91.97
	β_3	92.75	95.29	94.54	94.40	94.70	93.02	91.57
	β_4	95.02	93.61	95.27	94.58	95.21	96.25	89.54
	β_5	94.82	93.28	95.12	94.34	95.07	96.26	91.24
	β_6	94.70	93.26	95.00	94.33	94.93	96.10	91.11
	σ^2	80.51	92.78	90.15	89.47	90.44		
	ρ	93.03	78.96	92.93	90.38	91.90		
IBMSE	β_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	β_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	β_3	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	β_4	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	β_5	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	β_6	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	σ^2	21.50	7.36	0.32	0.01	1.13	10.85	10.85
	ρ	27.75	20.11	4.56	0.13	2.25	74.47	

Table 4: Stroke data. Simulation results with true parameter values fixed to the ML estimates. The estimators considered are those from ML, BC, JEF, BR, MBR, GEE and ROB. PU, empirical percentage of underestimation; RB, empirical relative bias; WALD, 95% Wald-type confidence intervals and IBMSE, relative increase in mean squared error from its absolute minimum due to bias.

Appendix

The expectations involved in the adjustment terms for mean and median BR are

$$\begin{aligned}
&E[(y-\mu)] = E[(y-\mu)(y-\mu)(y-\mu)] = 0,\\
&E[(y-\mu)^T\Omega_{\sigma^2}(y-\mu)) = \frac{q}{\sigma^2},\\
&E[(y-\mu)^T\Omega_{\rho}(y-\mu)) = -\frac{(q-1)q\rho}{(\rho-1)(q\rho-\rho+1)},\\
&E((y-\mu)^T\Omega_{\sigma^2\sigma^2}(y-\mu))] = \frac{2q}{\sigma^4},\\
&E[(y-\mu)^T\Omega_{\sigma^2\rho}(y-\mu))] = \frac{(q-1)q\rho}{(\rho-1)(q\rho-\rho+1)\sigma^2},\\
&E[(y-\mu)^T\Omega_{\rho\rho}(y-\mu))] = \frac{2(q-1)q(q\rho^2-\rho^2+1)}{(\rho-1)^2(q\rho\rho+1)^2},\\
&E[(y-\mu)^T\Omega_{\sigma^2}(y-\mu))((y-\mu)^T\Omega_{\sigma^2\sigma^2}(y-\mu))] = \frac{2q(q+2)}{\sigma^6},
\end{aligned}$$

$$\begin{aligned}
&E[(y-\mu)^T\Omega_{\sigma^2}(y-\mu))((y-\mu)^T\Omega_{\rho\rho}(y-\mu))] = -\frac{2(q-1)q(q+2)(q\rho^2-\rho^2+1)}{(\rho-1)^2(q\rho-\rho+1)^2\sigma^2},\\
&E[(y-\mu)^T\Omega_{\rho}(y-\mu))((y-\mu)^T\Omega_{\rho\rho}(y-\mu))]\\
&= -\frac{2(q-1)q(q^3\rho^3-3q\rho^3+2\rho^3+q^2\rho+5q\rho-6\rho-2q+4)}{(\rho-1)^3(q\rho-\rho+1)^3},\\
&E[(y-\mu)^T\Omega_{\rho}(y-\mu))((y-\mu)^T\Omega_{\sigma^2\sigma^2}(y-\mu))] = -\frac{2(q-1)q(q+2)\rho}{(\rho-1)(q\rho-\rho+1)\sigma^4}.
\end{aligned}$$

We give below the forms of the correlation matrices used in the simulations of Section 4 for $q=5$ and $q=10$, respectively.

- unstructured (unstr):

$$\begin{pmatrix}
1 & 0.59 & -0.30 & 0.02 & -0.37\\
0.59 & 1 & -0.57 & -0.68 & -0.71\\
-0.30 & -0.57 & 1 & 0 & 0.86\\
0.02 & -0.68 & 0 & 1 & 0.20\\
-0.37 & -0.71 & 0.86 & 0.20 & 1
\end{pmatrix}$$

and

$$\begin{pmatrix} 1 & -0.17 & 0.03 & 0.07 & 0.16 & 0.11 & 0.11 & 0.06 & -0.06 & 0.03 \\ -0.17 & 1 & 0.18 & -0.37 & -0.02 & -0.01 & 0.20 & 0.25 & 0.06 & -0.01 \\ 0.03 & 0.18 & 1 & -0.21 & 0.22 & -0.15 & 0.22 & -0.24 & 0.17 & -0.15 \\ 0.07 & -0.37 & -0.21 & 1 & 0.28 & -0.07 & -0.08 & 0.09 & 0 & -0.23 \\ 0.16 & -0.02 & 0.22 & 0.28 & 1 & -0.11 & 0.17 & 0.08 & -0.02 & -0.18 \\ 0.11 & -0.01 & -0.15 & -0.07 & -0.11 & 1 & -0.06 & 0.13 & 0.12 & -0.04 \\ 0.11 & 0.20 & 0.22 & -0.08 & 0.17 & -0.06 & 1 & -0.31 & -0.08 & -0.16 \\ 0.06 & 0.25 & -0.24 & 0.09 & 0.08 & 0.13 & -0.31 & 1 & 0.11 & 0.13 \\ -0.06 & 0.06 & 0.17 & 0 & -0.02 & 0.12 & -0.08 & 0.11 & 1 & -0.32 \\ 0.03 & -0.01 & -0.15 & -0.23 & -0.18 & -0.04 & -0.16 & 0.13 & -0.32 & 1 \end{pmatrix}$$

- clustered (clus):

$$\begin{pmatrix} 1 & -0.20 & 0 & 0 & 0 \\ -0.20 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.70 & 0.70 \\ 0 & 0 & 0.70 & 1 & 0.70 \\ 0 & 0 & 0.70 & 0.70 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & -0.20 & 0 & 0 & 0 & 0.83 & -0.17 & 0 & 0 & 0 \\ -0.20 & 1 & 0 & 0 & 0 & -0.17 & 0.83 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.70 & 0.70 & 0 & 0 & 0.83 & 0.58 & 0.58 \\ 0 & 0 & 0.70 & 1 & 0.70 & 0 & 0 & 0.58 & 0.83 & 0.58 \\ 0 & 0 & 0.70 & 0.70 & 1 & 0 & 0 & 0.58 & 0.58 & 0.83 \\ 0.83 & -0.17 & 0 & 0 & 0 & 1 & -0.20 & 0 & 0 & 0 \\ -0.17 & 0.83 & 0 & 0 & 0 & -0.20 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.83 & 0.58 & 0.58 & 0 & 0 & 1 & 0.70 & 0.70 \\ 0 & 0 & 0.58 & 0.83 & 0.58 & 0 & 0 & 0.70 & 1 & 0.70 \\ 0 & 0 & 0.58 & 0.58 & 0.83 & 0 & 0 & 0.70 & 0.70 & 1 \end{pmatrix}$$

References

Barndorff-Nielsen, O. E. and D. R. Cox (1994). *Inference and Asymptotics.* London: Chapman & Hall

Cordeiro, G. M. and P. McCullagh (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society: Series B (Methodological) 53*(3), 629–643.

Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics*. London: Chapman & Hall

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 27–38.

Ford, W. P. and P. M. Westgate (2018). A comparison of bias-corrected empirical covariance estimators with generalized estimating equations in small-sample longitudinal study settings. *Statistics in medicine 37*(28), 4318–4329.

Guo, X., W. Pan, J. E. Connett, P. J. Hannan, and S. A. French (2005). Small-sample performance of the robust score test and its modifications in generalized estimating equations. *Statistics in medicine 24*(22), 3479–3495.

Heinze, G. and M. Schemper (2001). A solution to the problem of monotone likelihood in cox regression. *Biometrics 57*(1), 114–119.

Huber, P. J. et al. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Volume 1, pp. 221–233. University of California Press.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 186*(1007), 453–461.

Kenne Pagui, E. C. and E. A. Colosimo (2020). Adjusted score functions for monotone likelihood in the cox regression model. *Statistics in Medicine 39*(10), 1558–1572.

Kenne Pagui, E. C., A. Salvan, and N. Sartori (2017). Median bias reduction of maximum likelihood estimates. *Biometrika 104*(4), 923–938.

Kosmidis, I. (2014). Bias in parametric estimation: reduction and useful side-effects. *Wiley Interdisciplinary Reviews: Computational Statistics 6*(3), 185–196.

Kosmidis, I. and D. Firth (2010). A generic algorithm for reducing bias in parametric estimation. *Electronic Journal of Statistics 4*, 1097–1112.

Kosmidis, I. and D. Firth (2018). Jeffreys-prior penalty, finiteness and shrinkage in binomial-response generalized linear models. *arXiv preprint arXiv:1812.01938*.

Kosmidis, I., D. Firth, et al. (2010). A generic algorithm for reducing bias in parametric estimation. *Electronic Journal of Statistics 4*, 1097–1112.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

Lima, V. and F. Cribari-Neto (2019). Penalized maximum likelihood estimation in the modified extended weibull distribution. *Communications in Statistics-Simulation and Computation 48*(2), 334–349.

Mancl, L. A. and T. A. DeRouen (2001). A covariance estimator for gee with improved small-sample properties. *Biometrics 57*(1), 126–134.

Pan, W. (2001). On the robust variance estimator in generalised estimating equations. *Biometrika 88*(3), 901–906.

Salvan, A., N. Sartori, and L. Pace (2020). *Modelli per risposte correlate*, pp. 269–318. Milano: Springer Milan.

Sartori, N., A. Salvan, L. Pace, and M. N. Sartori (2020). Package 'mlgdata'.

Wang, M. and Q. Long (2011). Modified robust variance estimator for generalized estimating equations with improved small-sample performance. *Statistics in medicine 30*(11), 1278–1291.

Wang, Y.-G. and V. Carey (2003). Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. *Biometrika 90*(1), 29–41.