Working Paper Series, N. 2, March 2014

Empirical and Simulated Adjustments of Composite Likelihood Ratio Statistics

Manuela Cattelan
Department of Statistical Sciences
University of Padua
Italy

Nicola Sartori
Department of Statistical Sciences
University of Padua
Italy

Abstract: Composite likelihood inference has gained much popularity thanks to its computational manageability and its theoretical properties. Unfortunately, performing composite likelihood ratio tests is inconvenient because of their awkward asymptotic distribution. There are many proposals for adjusting composite likelihood ratio tests in order to recover an asymptotic chi square distribution, but they all depend on the sensitivity and variability matrices. The same is true for Wald-type and score-type counterparts. In realistic applications sensitivity and variability matrices usually need to be estimated, but there are no comparisons of the performance of composite likelihood based statistics in such an instance. We compare the accuracy of inference based on the statistics considering two methods typically employed for estimation of sensitivity and variability matrices, namely an empirical method that exploits independent observations, and Monte Carlo simulation. The results in two examples involving the pairwise likelihood show that a very large number of independent observations should be available in order to obtain accurate coverages using empirical estimation, while simulation from the full model provides accurate results regardless of the availability of independent observations.

Keywords: composite likelihood, equicorrelated multivariate normal, multivariate probit, pairwise likelihood.

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Manuela Cattelan
tel: +39 049 827 4124
manuela.cattelan@unipd.it
http://www.stat.unipd.it/~cattelan

Empirical and Simulated Adjustments of Composite Likelihood Ratio Statistics

Manuela Cattelan
Department of Statistical Sciences
University of Padua
Italy

Nicola Sartori
Department of Statistical Sciences
University of Padua
Italy

Abstract: Composite likelihood inference has gained much popularity thanks to its computational manageability and its theoretical properties. Unfortunately, performing composite likelihood ratio tests is inconvenient because of their awkward asymptotic distribution. There are many proposals for adjusting composite likelihood ratio tests in order to recover an asymptotic chi square distribution, but they all depend on the sensitivity and variability matrices. The same is true for Wald-type and score-type counterparts. In realistic applications sensitivity and variability matrices usually need to be estimated, but there are no comparisons of the performance of composite likelihood based statistics in such an instance. We compare the accuracy of inference based on the statistics considering two methods typically employed for estimation of sensitivity and variability matrices, namely an empirical method that exploits independent observations, and Monte Carlo simulation. The results in two examples involving the pairwise likelihood show that a very large number of independent observations should be available in order to obtain accurate coverages using empirical estimation, while simulation from the full model provides accurate results regardless of the availability of independent observations.

Keywords: composite likelihood, equicorrelated multivariate normal, multivariate probit, pairwise likelihood.

1 Introduction

The use of the likelihood function to perform inference in statistical models is becoming more and more cumbersome for diverse reasons, as for example the availability of huge datasets and the implementation of complex models developed to reproduce natural phenomena. This problem is often overcome through the definition of pseudo-likelihood functions that are computationally manageable, but retain some nice properties of the likelihood function. Many of the pseudo-likelihood functions proposed in the literature belong to the class of composite likelihoods (Lindsay, 1988; Varin, 2008; Varin *et al.*, 2011). Indeed, the definition of composite likelihood given by Lindsay (1988) is quite general and encompasses any function which is a product of marginal or conditional probabilities for subsets of events. Composite likelihoods share some nice properties of the ordinary likelihood, as the unbiasedness of the composite likelihood score function and the asymptotic normal distribution of the maximum composite likelihood estimator (Molenberghs and Verbeke, 2005). The simplifications of both computational issues and model assumptions that derive from this type of pseudo-likelihood led to a considerable diffusion of composite likelihood estimation and hence the investigation of its theoretical properties and the development of further inferential techniques based on composite likelihood.

In this paper, we focus on hypothesis testing and confidence regions construction when a composite likelihood is employed. There are composite likelihood versions of the tests developed in the full likelihood context. Hence, Wald-type, score-type and likelihood ratios statistics based on the composite likelihood can be specified. However, as with the full likelihood, the Wald-type statistic lacks invariance under reparameterisations of the model and forces confidence regions to have an elliptical shape. On the other hand, score-type statistics are often numerically unstable (Rotnitzky and Jewell, 1990; Molenberghs and Verbeke, 2005; Pace *et al.*, 2011), while composite likelihood ratio statistics do not have the usual asymptotic chi square distribution.

There are different proposals to overcome the problem of the awkward asymptotic distribution of the composite likelihood ratio statistic. All such proposals, as well as the Wald-type and score-type statistics, depend on sensitivity and variability matrices, which are, respectively, the expected value of minus the hessian of the composite log likelihood and the variance of the composite score function. The computation of these matrices is generally cumbersome and approximations are typically used (Varin *et al.*, 2011, §5.1). The main purpose of this paper is to compare the behavior of the various statistics when they are based on estimated sensitivity and variability matrices. In particular, empirical and Monte Carlo estimates are considered. Two simulation studies are implemented to compare the performance of adjusted composite likelihood ratio statistics when pairwise likelihood is used for inferential purposes.

The paper is organized as follows. Section 2 reviews composite likelihood based statistics and the proposals to overcome the problem of the asymptotic distribution of the composite likelihood ratio statistics. Section 3 presents the methods commonly employed to estimate the sensitivity and variability matrices. Section 4 shows the results of simulation studies that compare the different statistics in two model settings, namely equicorrelated multivariate normal data and a multivariate probit model, and Section 5 concludes with a brief discussion.

2 Adjusting composite likelihood ratio statistics

Let $y_1, \ldots, y_n$ be independent realizations of a q-dimensional random vector $Y_i = (Y_{i1}, \ldots, Y_{iq})$, with density or probability function $f(y_i; \theta)$ depending on a d-dimensional parameter θ. If the full likelihood is computationally cumbersome, or the model cannot be fully specified, composite likelihood may offer a valid alternative. A composite likelihood is a combination of likelihoods for conditional or marginal events (Lindsay, 1988). Assume there are K marginal or conditional events $A_k(y_i)$ involving elements of y_i, $k = 1, \ldots, K$, for which we can compute the likelihood $L_k(\theta; y_i) \propto f(Y_i \in A_k; \theta)$, then the composite likelihood is

$$cL(\theta; y) = \prod_{i=1}^{n} \prod_{k=1}^{K} L_k(\theta; y_i)^{w_k},$$

where w_k are non negative weights and $y = (y_1, \ldots, y_n)$. The composite log likelihood is $cl(\theta; y) = \log cL(\theta; y)$ and the composite score function is $cU(\theta; y) = \nabla_\theta \, cl(\theta; y)$. The maximizer of $cl(\theta; y)$, $\hat{\theta}_c$, is the maximum composite likelihood estimate. The maximum composite likelihood estimator is asymptotically normally distributed, $\hat{\theta}_c \dot{\sim} N_d(\theta, G(\theta)^{-1})$, where $G(\theta)$ denotes the Godambe information matrix. Specifically, the asymptotic covariance matrix is $G(\theta)^{-1} = H(\theta)^{-1} J(\theta) H(\theta)^{-1}$, where $H(\theta) = \mathrm{E}\{-\nabla_\theta cU(\theta; y)\}$ is called the sensitivity matrix and $J(\theta) = \mathrm{E}\{cU(\theta) \, cU(\theta)^T\}$ is called the variability matrix. Composite likelihood is not a proper likelihood, but it can be interpreted as the likelihood for a misspecified model, hence the second Bartlett identity does not hold and typically $J(\theta) \neq H(\theta)$.

A type of composite likelihood often used in applications is the pairwise likelihood, which is the product of marginal bivariate probabilities,

$$pL(\theta; y) = \prod_{i=1}^{n} \prod_{j=1}^{q-1} \prod_{k=j+1}^{q} f(y_{ij}, y_{ik}; \theta)^{w_{ij,ik}},$$

and the pairwise log likelihood is $pl(\theta; y) = \log pL(\theta; y)$.

Assume that interest lies in a p-dimensional parameter γ, where $\theta = (\gamma, \delta)$ and δ is a nuisance parameter of dimension $d - p$. It is possible to define test statistics based on the composite likelihood which are analogous to those based on the full likelihood. Denote by $\hat{\theta}_{c\gamma}$ the constrained maximum composite likelihood estimate of θ for a fixed γ, and let $\hat{\theta}_c = (\hat{\gamma}_c, \hat{\delta}_c)$. The Wald-type statistic for the parameter of interest is

$$cW(\gamma) = (\hat{\gamma}_c - \gamma)^T \{G^{\gamma\gamma}(\hat{\theta}_{c\gamma})\}^{-1} (\hat{\gamma}_c - \gamma), \quad (1)$$

where $G^{\gamma\gamma}(\hat{\theta}_{c\gamma})$ denotes the $p \times p$ submatrix of the inverse of $G(\hat{\theta}_{c\gamma})$ pertaining to γ. The statistic $cW(\gamma)$ has an asymptotic χ^2_p distribution. Unfortunately, this quantity is not invariant to reparameterisations of the model.

The score-type statistic based on the composite likelihood is

$$cS(\gamma) = cU_\gamma(\hat{\theta}_{c\gamma}) H^{\gamma\gamma}(\hat{\theta}_{c\gamma}) \{G^{\gamma\gamma}(\hat{\theta}_{c\gamma})\}^{-1} H^{\gamma\gamma}(\hat{\theta}_{c\gamma}) cU_\gamma(\hat{\theta}_{c\gamma}), \quad (2)$$

where $cU_\gamma(\theta) = \nabla_\gamma cl(\theta; y)$ denotes the derivative of the composite log likelihood with respect to the parameter of interest and $H^{\gamma\gamma}(\hat{\theta}_{c\gamma})$ denotes the submatrix of the inverse of $H(\hat{\theta}_{c\gamma})$ pertaining to γ. The asymptotic distribution of $cS(\gamma)$ is χ^2_p, but this statistic is often numerically unstable (Molenberghs and Verbeke, 2005).

Finally, it is possible to define also a composite likelihood ratio statistic

$$cLR(\gamma) = 2\{cl(\hat{\theta}_c) - cl(\hat{\theta}_{c\gamma})\},$$

but its asymptotic distribution is a weighted sum of p independent chi square random variables with one degree of freedom, precisely $\sum_{i=1}^{p} \omega_i \chi^2_{1i}$, where $\omega_1, \ldots, \omega_p$ are the eigenvalues of $\{H^{\gamma\gamma}(\hat\theta_{c\gamma})\}^{-1} G^{\gamma\gamma}(\hat\theta_{c\gamma})$. This awkward distribution prevents the use of the composite likelihood ratio statistic when the dimension of the parameter of interest is larger than one, hence various adjustments have been proposed, mainly in order to recover an approximate χ^2_p distribution.

A first proposal for the adjustment of composite likelihood ratio statistics suggests to match the first order moment of the composite likelihood ratio statistic with that of a χ^2_p random variable (Molenberghs and Verbeke, 2005)

$$cLR(\gamma)_1 = \overline{\omega}(\gamma)^{-1} cLR(\gamma),$$

where $\overline{\omega}(\gamma) = \sum_{i=1}^{p} \omega_i(\hat\theta_{c\gamma})/p$, and then use a χ^2_p as approximate distribution. A better approximation can be obtained through first and second order moment matching (Varin, 2008), which gives a Satterthwaite type adjustment (Satterthwaite, 1946)

$$cLR(\gamma)_2 = \kappa^{-1} cLR(\gamma),$$

where $\kappa = \kappa(\gamma) = \sum_{i=1}^{p} \omega_i(\hat\theta_{c\gamma})^2 / \sum_{i=1}^{p} \omega_i(\hat\theta_{c\gamma})$. This quantity has an asymptotic χ^2_ν distribution, where the degrees of freedom are $\nu = \nu(\gamma) = (\sum_{i=1}^{p} \omega_i(\hat\theta_{c\gamma}))^2 / \sum_{i=1}^{p} \omega_i(\hat\theta_{c\gamma})^2$. Unfortunately, this adjustment yields an asymptotic distribution with number of degrees of freedom depending on the parameter.

Other two adjustments of the composite likelihood ratio statistics are proposed by Chandler and Bate (2007) and Pace *et al.* (2011). The former suggest the following adjustment

$$cLR(\gamma)_{CB} = \frac{(\hat\gamma_c - \gamma)^T \{G^{\gamma\gamma}(\hat\theta_c)\}^{-1} (\hat\gamma_c - \gamma)}{(\hat\gamma_c - \gamma)^T H_{\gamma\gamma}(\hat\theta_c) (\hat\gamma_c - \gamma)^T} cLR(\gamma), \tag{3}$$

which has asymptotic χ^2_p distribution. In a simulation study Chandler and Bate (2007) show that their proposal behaves well, and at least it does not perform worse than statistics (1) and (2) in all settings considered. However, Pace *et al.* (2011) show that $cLR(\gamma)_{CB}$ is not parameterisation invariant, and therefore propose a different rescaling that preserves the parameterisation invariance of likelihood ratio statistic, that is

$$cLR(\gamma)_I = \frac{cS(\gamma)}{cU_\gamma(\hat\theta_{c\gamma}) H^{\gamma\gamma}(\hat\theta_{c\gamma}) cU_\gamma(\hat\theta_{c\gamma})} cLR(\gamma), \tag{4}$$

which is again asymptotically χ^2_p distributed. Despite being partially based on the score statistic $cS(\gamma)$, $cLR(\gamma)_I$ usually does not inherit its numerical instability.

The performance of the different adjustments is compared in a simulation study in Pace *et al.* (2011) that consider two different model settings: equicorrelated multivariate normal data and first order autoregression. In both cases, the authors use pairwise likelihood for making inference on model parameters and compare the results with those produced by maximum likelihood based statistics. Moreover, in both settings it is possible to compute analytically the Fisher information matrix and the matrices $H(\theta)$ and $J(\theta)$ for the pairwise likelihood. In general, the statistic (4) seems to behave well in all settings considered, while in some instances the empirical coverage of adjustment (3) is much lower than the nominal value. These results are obtained when the quantities of interest can be computed analytically. This rarely occurs in applications where composite likelihood is employed. Indeed, composite likelihood is often used in complex models where not only it is not possible to deal with the full likelihood, but also the analytical computation of $H(\theta)$ and $J(\theta)$ is typically unfeasible. The main concern of this paper is to investigate the behavior of the different proposals when the quantities involved in the computation of the statistics have to be estimated.

3 Estimation of $H(\theta)$ and $J(\theta)$

Estimation of the matrices $H(\theta)$ and $J(\theta)$ is a typical concern in applications in which composite likelihood is employed since they are necessary also to report the standard errors of the maximum composite likelihood estimates. While $H(\theta)$ can be reasonably estimated through the observed hessian, the estimation of the variability matrix $J(\theta)$ poses major difficulties.

Matrices $H(\theta)$ and $J(\theta)$ are usually estimated either empirically, exploiting groups of independent or almost independent data, or through simulation. When there are groups of independent observations, as for example when data are divided into clusters, it is possible to estimate $J(\theta)$ as

$$\hat{J}^E(\theta) = \frac{1}{n}\sum_{i=1}^{n} cU(\theta; y_i)cU(\theta; y_i)^T,$$

where $cU(\theta; y_i)$ denotes the elements of the composite score involving only observations of the vector y_i. For example, $cU(\theta; y_i) = \sum_{j=1}^{q-1}\sum_{k=j+1}^{q} \nabla_\theta \log f(y_{ij}, y_{ik}; \theta)$ if pairwise likelihood is employed. When independent repetitions of the data are not available, as in time series or spatial data, but it is possible to identify groups of data with low dependence, this method may be applied to groups of slightly dependent data. For example, when dealing with time series with dependence decreasing in time, a window subsampling method may be employed (Varin, 2008). The empirical estimate of the sensitivity matrix is

$$\hat{H}^E(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \nabla_\theta cU(\theta; y_i),$$

which corresponds to minus the Hessian matrix. However, since the second Bartlett identity holds for single subsets of the data (Varin, 2008), the sensitivity matrix can also be estimated as

$$\hat{H}^E(\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} cU(\theta; y_i \in A_k)cU(\theta; y_i \in A_k)^T,$$

which avoids the computation of the second derivative. When pairwise likelihood is employed, this corresponds to

$$\hat{H}^E(\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{q-1}\sum_{k=j+1}^{q} cU(\theta; y_{ij}, y_{ik})cU(\theta; y_{ij}, y_{ik})^T.$$

The empirical estimation of $H(\theta)$ and $J(\theta)$ does not require any further assumptions than those made for the composite likelihood function, which consist only in the specification of low order marginal or conditional probabilities.

An alternative method to estimate the Godambe information matrix is through simulation, which requires assumptions about the full distribution of the data. Although this may appear an important limitation of this method, in most of the applications of composite likelihood a full model is assumed for the data, but the difficulties in computing the likelihood function lead to the use of composite likelihood. In these cases the likelihood function is difficult to evaluate, but it may be quite straightforward to simulate from the full model. This is the same kind of situations where modern Approximate Bayesian Computation methods are nowadays widely used (Marin *et al.*, 2012). Let y^m, $m = 1, \dots, M$, denote the mth dataset simulated from $f(y;\theta)$, the full distribution of the data. Then, the Monte Carlo estimates of $J(\theta)$ and $H(\theta)$ are

$$\hat{J}^S(\theta) = \frac{1}{M} \sum_{m=1}^{M} cU(\theta; y^m) cU(\theta; y^m)^T,$$

and

$$\hat{H}^S(\theta) = -\frac{1}{M} \sum_{m=1}^{M} \nabla_\theta cU(\theta; y^m).$$

Again, in the estimation of $H(\theta)$ it is possible to exploit the second Bartlett identity, this may be convenient especially if analytical first derivatives can be computed. Usually a few hundreds simulated datasets are sufficient for reasonable accuracy.

Even when it is possible to compute $J(\theta)$ exactly, it may be computationally more convenient to use $\hat{J}^S(\theta)$. Indeed, consider a single observation ($n = 1$) of a q-dimensional multivariate normal random vector, as for instance in spatial statistics. The computational cost of the likelihood is of order q^3, while that of the pairwise likelihood and score functions is of order q^2. On the other hand, the computational cost of $J(\theta)$ is of order q^4, while that of $\hat{J}^S(\theta)$ is Mq^2.

The main interest here is in investigating whether there are differences in the performances of the various composite likelihood based statistics when $H(\theta)$ and $J(\theta)$ have to be estimated with respect to cases in which they are available analytically, and which of the two estimating methods yields better results. Such an investigation has an important practical relevance since the estimation of $H(\theta)$ and $J(\theta)$ is the only option in most realistic applications. The proposed solutions are explored in simulation studies.

4 Simulation studies

Simulation studies are performed considering two different model settings and using pairwise likelihood for inferential purposes. The first setting assumes equicorrelated multivariate normal data. In this rather toyish case it is possible to compute analytically the sensitivity and the variability matrices (Pace *et al.*, 2011), thus allowing a comparison of the performance of analytical, empirical and simulation based quantities. The second model considered is a multivariate probit model, in which the analytical form of $H(\theta)$ and $J(\theta)$ is not available.

4.1 Equicorrelated multivariate normal

Assume that Y_i, $i = 1, \ldots, n$, has a q-dimensional normal distribution in which all components have mean μ, variance σ^2 and $\text{cor}(Y_{ij}, Y_{ik}) = \rho$, $\forall j \neq k$. Analytical expressions for $H(\theta)$ and $J(\theta)$ when pairwise likelihood is employed are reported in Pace *et al.* (2011). Several different settings are considered with increasing values of the correlation parameter ρ and increasing number of independent observations n. The values of the model parameters in the simulations are $\mu = 0$, $\sigma^2 = 1$ and $\rho = 0.2, 0.5, 0.9$. The dimension of the datasets considered is $n = 5, 30$ and 100, while $q = 30$. For each setting 100,000 datasets were simulated and the various statistics were computed using analytical, empirical and simulated versions of $H(\theta)$ and $J(\theta)$. Although a few hundreds repetitions are enough, as will be shown in a further simulation study reported later, in order to obtain an accurate approximation of the sensitivity and the variability matrices, $M = 1,000$ simulations are employed for the Monte Carlo estimation of $H(\theta)$ and $J(\theta)$. Table 1 shows the empirical coverages of the statistics when the nominal levels are 0.95 and 0.99. The superscripts A, E, and S denote that the statistic is computed employing analytical, empirical and Monte Carlo versions of matrices $H(\theta)$ and $J(\theta)$, respectively. The empirical coverages of statistics based on analytical quantites are analogous to those obtained in Pace *et al.* (2011). Given the poor performance of the adjustment (3) and of that based on first order matching, they are not reported here. The Wald-type statistic based on analytical quantities shows poor coverages that worsen for increasing values of the correlation parameter. Even though its coverage improves as n gets larger, it remains unsatisfactory even for $n = 100$ when $\rho = 0.9$. The score-type statistic behaves quite well, its coverage is not influenced by the strength of the correlation parameter, and it improves as n increases. On the other hand, shapes of confidence regions based on the score statistics may be quite irregular as shown in Pace *et al.* (2011). The adjustment of the composite likelihood ratio statistic based on second order matching and the one proposed by Pace *et al.* (2011) provide quite good results. The latter has a coverage somewhat lower than the nominal values when $n = 5$ and $\rho = 0.9$, but the problem disappears when n increases to 30.

Now consider the coverages when the statistics are computed using empirical or simulation based sensitivity and variability matrices. When empirical estimation is employed, the dimension of the dataset n appears of great importance. When $n = 5$ the coverages are worse than those obtained with analytical quantities. They seem to approach those obtained using analytical quantities as the dimension n increases, but they are not as accurate even for $n = 100$. On the contrary, the coverages of the statistics based on simulated quantities are almost identical to those of the statistics based on analytical calculations. In this case, the use of simulation based sensitivity and variability matrices appears definitely preferable.

Table 2 reports the empirical coverages of the statistics when μ is considered as a nuisance parameter. The Wald-type statistic does not perform well for small

n	ρ	W^A	W^E	W^S	S^A	S^E	S^S	LR_2^A	LR_2^E	LR_2^S	LR_I^A	LR_I^E	LR_I^S
							95.0						
	0.2	96.1	90.3	96.0	93.9	100.0	93.8	95.3	100.0	95.3	95.1	100.0	95.0
5	0.5	80.6	67.6	80.5	94.1	100.0	94.0	95.7	100.0	95.7	94.6	99.9	94.5
	0.9	52.4	41.5	52.3	94.0	100.0	93.9	92.8	98.9	92.7	90.4	99.3	90.2
	0.2	95.3	93.8	95.1	94.8	88.8	94.8	95.2	94.3	95.1	95.2	91.3	95.1
30	0.5	91.6	89.8	91.5	94.9	88.8	94.8	95.1	94.0	95.1	94.9	95.0	94.8
	0.9	77.8	75.4	77.8	94.8	89.0	94.8	94.6	97.7	94.6	94.3	97.6	94.2
	0.2	95.1	94.0	95.0	95.0	92.5	94.9	95.2	94.8	95.1	95.1	93.2	95.0
100	0.5	93.8	92.9	93.8	95.1	92.6	95.0	95.1	94.4	95.1	95.0	94.2	94.9
	0.9	87.0	85.7	86.9	94.9	92.5	94.8	94.8	95.0	94.8	94.7	95.6	94.6
							99.0						
	0.2	99.1	94.7	99.1	97.6	100.0	97.5	98.9	100.0	98.9	98.7	100.0	98.7
5	0.5	88.8	75.8	88.8	97.6	100.0	97.5	99.1	100.0	99.1	98.8	100.0	98.8
	0.9	57.9	47.0	57.9	97.6	100.0	97.6	98.7	99.3	98.7	96.9	99.9	96.8
	0.2	99.1	98.5	99.0	98.7	95.5	98.6	98.9	98.8	98.8	99.0	97.6	99.0
30	0.5	96.7	95.1	96.7	98.6	95.4	98.6	99.0	98.9	99.0	99.0	99.5	98.9
	0.9	84.4	81.8	84.3	98.6	95.5	98.6	98.9	99.9	98.9	98.7	99.8	98.7
	0.2	99.1	98.5	99.0	98.9	97.5	98.8	98.9	98.7	98.8	99.0	98.0	99.0
100	0.5	98.2	97.6	98.2	98.9	97.6	98.8	99.0	98.7	99.0	99.0	98.7	98.9
	0.9	93.0	91.7	92.9	98.8	97.5	98.8	99.0	99.2	98.9	98.9	99.3	98.8

Table 1: Empirical coverages of the statistics: Wald-type (W), score-type (S), composite likelihood ratio using second order matching adjustment (LR_2) and composite likelihood ratio adjustment by Pace *et al.* (2011) (LR_I) for nominal values 95% and 99% in an equicorrelated multivariate normal model for parameters of interest (μ, σ^2, ρ), with $\rho = 0.2, 0.5, 0.9$ and $n = 5, 30, 100$, using analytical (^{A}), empirical (^{E}) and Monte Carlo (^{S}) versions of $H(\theta)$ and $J(\theta)$.

values of n when correlation is high. The score-type statistic, the adjustments of the composite likelihood ratio statistic proposed by Pace *et al.* (2011) and that based on second order matching behave quite well, even though the latter two improve their accuracy as n gets larger. Again, the coverages of the statistics computed with empirical estimation of the matrices $H(\theta)$ and $J(\theta)$ are quite different from those deriving from the analytical quantities, and they clearly improve as the number of independent repetitions of observations increases. On the contrary, the statistics based on Monte Carlo simulations of the sensitivity and variability matrices are almost equal to those based on analytical quantities.

When the only parameter of interest is ρ, we have qualitatively the same results, the empirical coverages of the different statistics are given in Table A.1 in the Appendix.

In order to obtain accurate estimates of $H(\theta)$ and $J(\theta)$ in the Monte Carlo procedure we employed $M = 1,000$ replications. In some instances, this number of replications may require considerable computational time. We therefore investigate whether it is possible to obtain accurate coverages with fewer replications. In particular, we consider the situation in which all the parameters of the equicorrelated multivariate normal model are of interest and compare the coverages of the statistics based on simulated sensitivity and variability matrices when $M = 100$, 250 and 500.

n	ρ	W^A	W^E	W^S	S^A	S^E	S^S	LR_2^A	LR_2^E	LR_2^S	LR_I^A	LR_I^E	LR_I^S
							95.0						
5	0.2	95.9	95.2	96.2	95.9	100.0	95.9	96.6	98.6	96.6	96.0	99.6	96.0
	0.5	96.0	80.1	86.1	96.0	100.0	96.0	96.5	98.8	96.4	94.4	99.3	94.3
	0.9	62.1	53.7	62.1	96.0	100.0	95.9	90.7	98.9	90.6	89.0	98.4	88.9
30	0.2	95.2	92.8	95.1	95.3	90.0	95.2	95.4	92.6	95.3	95.3	91.6	95.2
	0.5	95.3	91.8	93.3	95.4	89.9	95.3	95.1	92.4	95.0	94.8	94.2	94.8
	0.9	85.5	83.3	85.5	95.4	89.9	95.3	94.3	97.7	94.3	94.1	97.0	94.0
100	0.2	95.2	94.1	95.1	95.2	93.2	95.2	95.3	94.1	95.3	95.2	93.7	95.1
	0.5	95.1	93.8	94.4	95.2	93.1	95.1	95.0	93.8	95.0	94.9	94.2	94.8
	0.9	95.1	90.6	91.4	95.2	93.3	95.1	94.9	95.1	94.9	94.8	95.7	94.7
							99.0						
5	0.2	98.7	99.2	99.3	98.8	100.0	98.7	99.4	100.0	99.3	99.2	100.0	99.2
	0.5	98.8	88.2	94.1	98.8	100.0	98.8	99.7	100.0	99.7	98.8	100.0	98.7
	0.9	67.6	60.0	67.6	98.8	100.0	98.7	98.0	99.9	97.9	96.1	99.8	96.1
30	0.2	98.9	98.0	99.1	99.0	96.1	98.9	99.1	98.0	99.0	99.1	97.4	99.1
	0.5	98.9	96.8	98.0	98.9	96.1	98.9	99.1	98.4	99.0	99.0	99.1	98.9
	0.9	91.1	89.3	91.1	98.9	96.0	98.9	98.7	99.8	98.7	98.6	99.5	98.6
100	0.2	99.0	98.6	99.0	99.0	98.0	99.0	99.0	98.5	98.9	99.1	98.4	99.0
	0.5	99.0	98.1	98.6	99.0	97.8	99.0	99.0	98.3	99.0	99.0	98.7	98.9
	0.9	99.0	95.5	96.3	99.1	97.9	99.0	99.0	99.3	98.9	98.9	99.3	98.9

Table 2: Empirical coverages of the statistics: Wald-type (W), score-type (S), composite likelihood ratio using second order matching adjustment (LR_2) and composite likelihood ratio adjustment by Pace *et al.* (2011) (LR_I) for nominal values 95% and 99% in an equicorrelated multivariate normal model for parameters of interest (σ^2, ρ), with $\rho = 0.2, 0.5, 0.9$ and $n = 5, 30, 100$, using analytical (^{A}), empirical (^{E}) and Monte Carlo (^{S}) versions of $H(\theta)$ and $J(\theta)$.

Table 3 reports the coverages of the statistics for increasing values of replications M and considering dimension $n = 5, 30$ in the most extreme case with $\rho = 0.9$. The results for the other values of the correlation parameter are included in Tables A.2 and A.3 in the Appendix. In all cases, the results with $M = 500$ are almost identical to those obtained with $M = 1,000$, and even $M = 250$ seems to provide very accurate results. Other numbers of Monte Carlo simulations between 500 and 1,000 yield the same results obtained with $M = 500$. The relatively low value of M sufficient for reasonable accuracy may be explained by the fact that matrices $H(\theta)$ and $J(\theta)$ are expected values and they are only a part of the adjusted statistics.

4.2 Multivariate probit

Consider a multivariate probit model in which Y_{ij} is a binary random variable that can assume values either 0 or 1. We use the latent variable representation

$$Y_{ij} = 1 \Leftrightarrow Z_{ij} > 0, \qquad i = 1, \ldots, n, \;\; j = 1, \ldots, q,$$

with $Z_{ij} = x_{ij}^T \beta + U_i + \epsilon_{ij}$, where x_{ij} is an r-dimensional vector of covariates, β is a vector of regression parameters, $U_i \overset{iid}{\sim} N(0, \sigma^2)$, are independent zero-mean random effects and ϵ_{ij} are independent normally distributed errors with mean 0. The errors

	W^S		S^S		LR_2^S		LR_I^S	
n	5	30	5	30	5	30	5	30
M				95.0				
100	51.8	77.0	93.1	94.0	92.6	94.5	89.0	93.3
250	52.2	77.5	93.7	94.5	92.7	94.6	89.9	93.9
500	52.3	77.7	93.8	94.7	92.7	94.6	90.2	94.2
				99.0				
100	57.4	83.8	97.1	98.2	98.5	98.8	96.2	98.2
250	57.7	84.1	97.4	98.5	98.6	98.9	96.6	98.5
500	57.8	84.3	97.5	98.5	98.6	98.9	96.7	98.6

Table 3: Comparison of coverages of the statistics: Wald-type (W), score-type (S), composite likelihood ratio using second order matching adjustment (LR_2) and composite likelihood ratio adjustment by Pace *et al.* (2011) (LR_I) based on Monte Carlo simulation as M increases in an equicorrelated normal model with $\rho = 0.9$ and $n = 5, 30$.

are independent of the random effects and their variance is set to 1 for identification purposes. Hence, the latent variables Z_{ij} and Z_{kl} are independent if $i \neq k$, while Z_{ij} and Z_{ik} have correlation $\rho = \sigma^2/(1+\sigma^2)$, $\forall\, j \neq k$. The full likelihood is cumbersome since it entails calculation of multiple integrals of a q-variate multivariate normal distribution. In this instance pairwise likelihood is a valid alternative (Le Cessie and Van Houwelingen, 1994), indeed the pairwise log-likelihood is

$$pl(\beta, \rho; y) = \sum_{i=1}^{n} \sum_{j=1}^{q-1} \sum_{k=j+1}^{q} \log f(Y_{ij} = y_{ij}, Y_{ik} = y_{ik}; \beta, \rho),$$

where, for instance, $f(Y_{ij} = 1, Y_{ik} = 1; \beta, \rho) = \Phi_2(\lambda_{ij}, \lambda_{ik}; \rho)$ is the standard bivariate normal distribution with correlation ρ, computed in $(\lambda_{ij}, \lambda_{ik})$ with $\lambda_{ij} = x_{ij}^T \beta \sqrt{1-\rho}$.

However, in this case it is not possible to compute analytically the matrices $H(\theta)$ and $J(\theta)$ deriving from the pairwise likelihood, so only statistics based on estimated quantities can be compared. In the model, an intercept term and one covariate are included. The covariate is simulated from a uniform distribution in $[-1, 1]$, while model parameters $(\beta_0, \beta_1, \sigma^2)$ are set to $(0.5, 1, 1)$. The length of the multivariate binary observations is set to $q = 30$, and increasing dimension of the dataset is considered, namely $n = 10, 30$ and 100. For each setting 10,000 datasets are simulated and the Monte Carlo estimates of $H(\theta)$ and $J(\theta)$ are based on $M = 1,000$ replications.

Table 4 shows the empirical coverages when only two parameters are of interest, namely (β_1, ρ). As expected, for small n the simulation based statistics have better coverages than the empirical based ones, and with $n = 100$ the difference is still evident for the Wald-type statistic. Coverages of the statistics based on Monte Carlo simulation of $H(\theta)$ and $J(\theta)$ are always quite good.

Finally, Table 5 reports the empirical coverages when $\rho = \sigma^2/(1+\sigma^2)$ is the only parameter of interest. Again, the coverage of simulation based statistics are more accurate for small values of n, but when the number of independent repetitions is large, even statistics based on empirical quantities provide good results. The results when all parameters are of interest are reported in Table A.4 in the Appendix.

n	W^E	W^S	S^E	S^S	LR_2^E	LR_2^S	LR_{IN}^E	LR_{IN}^S
				95.0				
10	86.0	92.8	93.6	95.1	97.8	95.6	96.1	95.1
30	90.6	93.7	93.2	94.7	97.2	95.0	96.0	94.7
100	92.8	94.6	94.7	95.1	95.3	94.9	95.4	95.1
				99.0				
10	90.9	96.5	99.8	99.1	99.5	99.1	99.0	99.1
30	94.8	97.8	98.4	98.9	99.4	98.9	99.1	98.9
100	96.8	98.5	98.8	98.9	99.2	98.9	98.9	98.9

Table 4: Empirical coverages of the statistics: Wald-type (W), score-type (S), composite likelihood ratio using second order matching adjustment (LR_2) and composite likelihood ratio adjustment by Pace *et al.* (2011) (LR_I) for nominal values 95% and 99% for the parameter of interest (β_1, ρ) in a multivariate probit model with $q = 30$ and $n = 10, 30, 100$, using empirical (^{E}) and Monte Carlo (^{S}) versions of $H(\theta)$ and $J(\theta)$.

n	W^E	W^S	S^E	S^S	LR_2^E	LR_2^S	LR_{IN}^E	LR_{IN}^S
				95.0				
10	89.9	93.7	89.9	94.7	97.0	95.0	97.0	95.0
30	92.8	95.2	93.2	95.1	96.4	95.1	96.4	95.1
100	94.3	94.7	94.5	95.0	95.1	95.1	95.1	95.1
				99.0				
10	93.3	96.4	98.1	99.2	99.0	99.2	99.0	99.2
30	96.1	98.0	97.9	99.0	99.2	99.1	99.2	99.1
100	97.4	98.6	98.8	99.0	99.0	99.0	99.0	99.0

Table 5: Empirical coverages of the statistics: Wald-type (W), score-type (S), composite likelihood ratio using second order matching adjustment (LR_2) and composite likelihood ratio adjustment by Pace *et al.* (2011) (LR_I) for nominal values 95% and 99% for the parameter of interest ρ in a multivariate probit model with $q = 30$ and $n = 10, 30, 100$, using empirical (^{E}) and Monte Carlo (^{S}) versions of $H(\theta)$ and $J(\theta)$.

5 Discussion

This paper considers hypothesis testing using likelihood based statistics when a composite likelihood is employed for inferential purposes. Hypothesis testing presents some difficulties since Wald-type tests lack invariance to reparameterisations of the model, score-type tests are often numerically unstable, while composite likelihood ratio statistics do not follow the usual asymptotic chi square distribution. Many different adjustments of the composite likelihood ratio statistic have been proposed to overcome the problem of its awkward asymptotic distribution. The proposal by Pace *et al.* (2011) seems an interesting alternative, however its performance has been considered so far only in examples in which the sensitivity and the variability matrices can be computed analytically. This rarely happens in applications in which composite likelihood is employed and typically those matrices need to be estimated. We considered the performance of the different statistics when $H(\theta)$ and $J(\theta)$ are estimated either empirically, or through Monte Carlo simulation. The score-type statistic, the adjustment of the composite likelihood ratio statistic based on second order moment matching and the adjustment proposed by Pace *et al.* (2011) seem to perform quite well in all situations considered. However, score-type tests can be numerically unstable, while the adjustment based on the second order moment matching has an asymptotic distribution which depends on the parameters of the model.

The results show that empirical estimation of the sensitivity and variability matrices requires a large number of independent repetitions of the observations and in our simulations it is not very accurate even with a dataset with as much as 100 independent replications. In many applications, as in time series or in spatial statistics, subsets of independent data are not available and the empirical method is applied to subsets of data with low dependence, using for example window subsampling. In these instances we may expect that the performance of the statistics based on empirical quantities will be even worse. The coverages of the statistics based on Monte Carlo simulation are almost identical to those of the statistics based on analytically computed quantities in the equicorrelated multivariate normal setting. In the multivariate probit model it is not possible to compute the sensitivity and variability matrices analytically, but the statistics based on simulation provide coverages closer to the nominal values than the empirically estimated ones. A further simulation study shows that the computational burden deriving from the simulation of the matrices $H(\theta)$ and $J(\theta)$ can be reduced since $M = 500$ repetitions, or even $M = 250$, may be enough. Moreover, such moderate number of repetitions can be done in parallel, thus substantially reducing computational time. In general, it seems that simulation based quantities are preferable, even when the number of independent repetitions of the data is quite large. Therefore, even considering the computational cost of exact calculation of matrix $J(\theta)$ in complex models, the simulation approach should be the default choice whenever simulation from the full model is feasible.

References

Chandler, R. E. and Bate, S. (2007). Inference for clustered data using the independence likelihood. *Biometrika* **94**, 167–183.

Le Cessie, S. and Van Houwelingen, J.C. (1994). Logistic regression for correlated binary data. *Journal of the Royal Statistical Society Series C - Applied Statistics* **43**, 95–108.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics* **80**, 221–239.

Marin, J.-M., Pudlo, P., Robert, C. P. and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing* **22**, 1167–1180.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer: New York.

Pace, L., Salvan, A. and Sartori, N. (2011). Adjusting composite likelihood ratio statistics. *Statistica Sinica* **21**, 129–148.

Rotnitzky, A. and Jewell, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* **77**, 485–497.

Satterthwaite, F. E. (1946). Approximate distribution of estimates of variance components. *Biometrics Bullettin* **2**, 110–114.

Varin, C. (2008). On composite marginal likelihoods. *AStA Advances in Statistical Analysis* **95**, 1–28.

Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* **21**,5–42.

A Appendix

n	ρ	W^A	W^E	W^S	S^A	S^E	S^S	LR_2^A	LR_2^E	LR_2^S	LR_I^A	LR_I^E	LR_I^S
						95.0							
5	0.2	98.7	74.0	98.7	98.7	78.6	98.7	98.7	76.8	98.6	98.7	76.8	98.6
	0.5	95.7	65.6	87.9	95.8	81.2	95.7	93.9	73.8	93.9	93.9	73.8	93.9
	0.9	72.7	56.6	72.7	90.1	83.7	90.0	86.5	69.1	86.4	86.5	69.1	86.4
30	0.2	95.4	89.3	95.3	95.5	89.8	95.4	95.4	89.6	95.4	95.4	89.6	95.4
	0.5	95.0	87.7	94.0	95.0	90.6	95.0	94.8	89.9	94.8	94.8	89.9	94.8
	0.9	90.6	83.3	90.5	94.1	91.0	94.1	93.7	89.3	93.7	93.7	89.3	93.7
100	0.2	95.1	93.0	95.1	95.2	93.1	95.1	95.2	93.1	95.1	95.2	93.1	95.1
	0.5	94.9	92.5	94.6	94.9	93.4	94.9	94.9	93.2	94.9	94.9	93.2	94.9
	0.9	93.6	90.4	93.5	94.7	93.6	94.7	94.6	93.1	94.6	94.6	93.1	94.6
						99.0							
5	0.2	99.7	100.0	99.8	99.7	100.0	97.7	99.7	100.0	99.6	99.7	100.0	99.6
	0.5	100.0	77.0	96.9	100.0	100.0	100.0	100.0	95.8	99.9	100.0	95.8	99.9
	0.9	79.5	62.4	79.5	98.5	100.0	98.4	94.6	81.5	94.6	94.6	81.5	94.6
30	0.2	99.3	95.1	99.3	99.3	95.8	99.3	99.3	95.5	99.3	99.3	95.5	99.3
	0.5	99.1	93.4	98.4	99.1	96.6	99.1	99.0	95.7	99.0	99.0	95.7	99.0
	0.9	95.3	88.7	95.3	98.6	97.1	98.6	98.4	94.6	98.4	98.4	94.6	98.4
100	0.2	99.1	97.6	99.1	99.1	97.9	99.1	99.1	97.8	99.0	99.1	97.8	99.0
	0.5	99.0	96.9	98.8	99.0	98.1	99.0	99.0	97.8	99.0	99.0	97.8	99.0
	0.9	97.8	94.9	97.8	98.9	98.4	98.9	98.8	97.6	98.8	98.8	97.6	98.8

Table A.1: Empirical coverages of the statistics: Wald-type (W), score-type (S), composite likelihood ratio using second order matching adjustment (LR_2) and composite likelihood ratio adjustment by Pace *et al.* (2011) (LR_I) for nominal values 95% and 99% in an equicorrelated multivariate normal model for parameter of interest ρ, with $\rho = 0.2, 0.5, 0.9$ and $n = 5, 30, 100$, using analytical (^{A}), empirical (^{E}) and Monte Carlo (^{S}) versions of $H(\theta)$ and $J(\theta)$.

	W^S		S^S		LR_2^S		LR_I^S	
n	5	30	5	30	5	30	5	30
M				95.0				
100	95.2	94.2	93.1	93.9	95.1	95.0	94.2	94.1
250	95.7	94.9	93.6	94.5	95.2	95.1	94.7	94.8
500	95.9	95.1	93.8	94.7	95.2	95.1	95.0	95.0
				99.0				
100	98.8	98.7	97.1	98.3	98.8	98.7	98.3	98.6
250	99.0	98.9	97.4	98.5	98.8	98.8	98.5	98.8
500	99.1	99.0	97.4	98.6	98.9	98.8	98.6	98.9

Table A.2: Comparison of coverages of the statistics: Wald-type (W), score-type (S), composite likelihood ratio using second order matching adjustment (LR_2) and composite likelihood ratio adjustment by Pace *et al.* (2011) (LR_I) based on Monte Carlo simulation as M increases in an equicorrelated normal model with $\rho = 0.2$ and $n = 5, 30$.

	W^S		S^S		LR_2^S		LR_I^S	
n	5	30	5	30	5	30	5	30
M				95.0				
100	79.5	90.7	93.3	94.0	95.5	94.9	93.5	93.9
250	80.2	91.3	93.8	94.5	95.7	95.0	94.2	94.5
500	80.3	91.4	94.0	94.7	95.7	95.1	94.4	94.7
				99.0				
100	88.0	96.2	97.2	98.2	99.0	98.9	98.3	98.5
250	88.5	96.5	97.4	98.5	99.1	99.0	98.6	98.8
500	88.6	96.7	97.5	98.5	99.1	99.0	98.7	98.9

Table A.3: Comparison of coverages of the statistics: Wald-type (W), score-type (S), composite likelihood ratio using second order matching adjustment (LR_2) and composite likelihood ratio adjustment by Pace *et al.* (2011) (LR_I) based on Monte Carlo simulation as M increases in an equicorrelated normal model with $\rho = 0.5$ and $n = 5, 30$.

n	W^E	W^S	S^E	S^S	LR_2^E	LR_2^S	LR_{IN}^E	LR_{IN}^S
				95.0				
10	68.1	91.8	91.5	94.9	99.0	94.7	98.0	94.9
30	82.7	93.5	91.7	94.7	99.0	94.6	97.7	94.7
100	89.6	94.6	94.3	95.3	95.8	95.0	96.1	95.1
				99.0				
10	75.7	96.1	100.0	98.7	99.2	98.8	99.5	98.9
30	88.6	97.8	97.2	98.9	100.0	98.7	99.6	99.0
100	94.8	98.6	98.6	99.0	99.4	98.8	99.3	99.1

Table A.4: Empirical coverages of the statistics: Wald-type (W), score-type (S), composite likelihood ratio using second order matching adjustment (LR_2) and composite likelihood ratio adjustment by Pace *et al.* (2011) (LR_I) for nominal values 95% and 99% for the parameter of interest (β_0, β_1, ρ) in a multivariate probit model with $q = 30$ and $n = 10, 30, 100$, using empirical (^{E}) and Monte Carlo (^{S}) versions of $H(\theta)$ and $J(\theta)$.

Acknowledgements

The first author acknowledges the “Progetto Giovani Ricercatori” of the University of Padova for financial support of a research project carried out at the Department of Statistical Sciences, University of Padova. The second author acknowledges the financial support of the CARIPARO Foundation Excellence - grant 2011/2012.