

## ADJUSTING COMPOSITE LIKELIHOOD RATIO STATISTICS

Luigi Pace, Alessandra Salvan and Nicola Sartori

*University of Udine, University of Padova and University of Padova*

*Abstract:* Composite likelihood may be useful for approximating likelihood based inference when the full likelihood is too complex to deal with. Stemming from a misspecified model, inference based on composite likelihood requires suitable corrections. Here we focus on the composite likelihood ratio statistic for a multidimensional parameter of interest, and we propose a parameterization invariant adjustment that allows reference to the usual asymptotic chi-square distribution. Two examples dealing with pairwise likelihood are analysed through simulation.

*Key words and phrases:* Equicorrelated multivariate normal distribution, first-order autoregression, Godambe information, pairwise likelihood, pseudolikelihood.

### 1. Introduction

A partially misspecified likelihood may prove useful in inference. Its usefulness is enhanced, as with composite likelihood (Lindsay (1988)), when the fully specified likelihood is computationally cumbersome, or when a fully specified model is out of reach. See Varin (2008) and Varin, Reid, and Firth (2011) for recent reviews on composite likelihood methods.

Some effects of misspecification are easy to cope with. Under regularity conditions, provided the misspecified score remains an unbiased estimating function, asymptotic normality of the maximum likelihood estimator still holds with Godambe information replacing expected information in the asymptotic covariance, see e.g., Cox and Reid (2004). This allows the construction of Wald-type test statistics and confidence regions. Score-type statistics, having the usual asymptotic null distribution, can also be defined (Molenberghs and Verbeke (2005, Chap. 9)), but seem to suffer from numerical instability. As a consequence, most routine statistical analyses employ Wald-type statistics.

As is well known, Wald-type statistics lack invariance under reparameterization, and force confidence regions to have an elliptical shape. In this respect a likelihood ratio-type statistic would be more appealing. However, under the form of misspecification we are considering, the asymptotic distribution of the likelihood ratio statistic departs from the familiar likelihood result, and involves a linear combination of independent chi-squared variates with coefficients given

by the eigenvalues of a matrix related to Godambe information (Foutz and Srivastava (1977, 1978); Kent (1982)).

We propose a parameterization invariant adjustment to the composite likelihood ratio statistic for a multidimensional parameter of interest that allows reference to the usual asymptotic chi-square distribution.

The paper is organized as follows. In Section 2, notation and background are given. Section 3 introduces the proposed adjustment and outlines its relation to alternative adjustments in the literature. Two examples dealing with pairwise likelihood are analysed in detail in Section 4; the first deals with the equicorrelated multivariate normal distribution, the second considers first-order autoregression. In both examples the overall parameter is three-dimensional and various cases of components of interest are considered. Simulation results indicate that the proposed adjustment allows quite accurate inferences. Some concluding remarks are given in Section 5.

## 2. Notation

Consider independent observations  $y_i$  of a random vector  $Y_i = (Y_{i1}, \dots, Y_{iq})$ ,  $i = 1, \dots, n$ , where  $Y_i$  has density  $f(y_i; \theta)$ ,  $\theta \in \Theta \subseteq \mathbb{R}^d$ ,  $y_i \in \mathcal{Y}$ . Let  $y = (y_1, \dots, y_n)$ . We denote by  $\ell(\theta; y) = \sum_{i=1}^n \log f(y_i; \theta)$  the full log likelihood. The corresponding score function and expected information are denoted by  $U(\theta) = U(\theta; y) = (\partial/\partial\theta)\ell(\theta; y)$  and  $I(\theta) = E_\theta\{-\partial U(\theta)/\partial\theta^\top\}$ . The maximum likelihood estimate is  $\hat{\theta}$ . The log likelihood ratio is  $w(\theta) = 2\{\ell(\hat{\theta}) - \ell(\theta)\}$ , and its Wald and score variants are  $w^e(\theta) = (\hat{\theta} - \theta)^\top I(\theta)(\hat{\theta} - \theta)$  and  $w^u(\theta) = U(\theta)^\top I(\theta)^{-1}U(\theta)$ , respectively.

Let  $\theta$  be partitioned as  $\theta = (\psi, \lambda)$ , where the  $d_0$  dimensional parameter  $\psi$  is the component of interest. Let  $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$  denote the constrained maximum likelihood estimate of  $\theta$  with  $\psi$  fixed. Under regularity conditions, the profile likelihood ratio statistic  $w_p(\psi) = 2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)\}$  has an asymptotic null  $\chi_{d_0}^2$  distribution. In order to define its Wald and score variants, partition the score vector as  $U(\theta)^\top = (U_\psi(\theta)^\top, U_\lambda(\theta)^\top)$ , where  $U_\psi(\theta) = (\partial/\partial\psi)\ell(\theta; y)$  and  $U_\lambda(\theta) = (\partial/\partial\lambda)\ell(\theta; y)$ , and consider the further partitions

$$I(\theta) = \begin{bmatrix} I_{\psi\psi} & I_{\psi\lambda} \\ I_{\lambda\psi} & I_{\lambda\lambda} \end{bmatrix}, \quad I(\theta)^{-1} = \begin{bmatrix} I^{\psi\psi} & I^{\psi\lambda} \\ I^{\lambda\psi} & I^{\lambda\lambda} \end{bmatrix},$$

where, for instance,  $I_{\psi\lambda} = E_\theta\{-\partial U_\psi(\theta)/\partial\lambda^\top\}$ . The Wald statistic is  $w_p^e(\psi) = (\hat{\psi} - \psi)^\top (I^{\psi\psi})^{-1}(\hat{\psi} - \psi)$  and the score variant is  $w_p^u(\psi) = U_\psi(\hat{\theta}_\psi)^\top I^{\psi\psi} U_\psi(\hat{\theta}_\psi)$ .

Composite likelihood is defined through  $K$  marginal or conditional events  $A_k(y_i)$  on  $\mathcal{Y}$ ,  $k = 1, \dots, K$ , giving likelihood contributions  $L_k(\theta; y_i) = L(\theta; A_k(y_i))$ .

Composite likelihood is then defined as

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

where  $w_k$ ,  $k = 1, \dots, K$ , are positive weights. We denote by  $c\ell(\theta) = c\ell(\theta; y) = \log cL(\theta; y)$  the composite log likelihood.

When the events  $A_k(y_i)$  are defined in terms of pairs of observations  $(y_{ir}, y_{is})$ , from the bivariate marginal density  $f_{rs}(y_{ir}, y_{is}; \theta)$ ,  $r, s = 1, \dots, q$ ,  $r \neq s$ , composite log likelihood is called pairwise log likelihood and is denoted by

$$p\ell(\theta; y) = \sum_{i=1}^n \sum_{\substack{r,s=1 \\ r \neq s}}^q w_{rs} \log f_{rs}(y_{ir}, y_{is}; \theta). \quad (2.1)$$

Hereafter, regularity conditions as detailed e.g., in Molenberghs and Verbeke (2005, Sec. 9.2.2) are assumed.

The composite score function  $(\partial/\partial\theta)c\ell(\theta; y)$  is denoted by  $cU(\theta; y)$  and the pairwise score  $(\partial/\partial\theta)p\ell(\theta; y)$  by  $pU(\theta; y)$ . More generally, when a pairwise likelihood is considered as a special composite likelihood, the prefix  $c$  in the notation for a given quantity is replaced by  $p$ .

Let  $H = H(\theta) = E_{\theta}\{-\partial cU(\theta)/\partial\theta^{\top}\}$  and  $J = J(\theta) = E_{\theta}\{cU(\theta)cU(\theta)^{\top}\}$ . Godambe information matrix is given by  $G = G(\theta) = H(\theta)J(\theta)^{-1}H(\theta)$ . The maximum composite likelihood estimate, i.e., the maximizer of  $cL(\theta; y)$ , is denoted by  $\hat{\theta}_c$ . Similarly,  $\hat{\theta}_p$  denotes the maximum pairwise likelihood estimate. Asymptotically, the distribution of  $\hat{\theta}_c$  is multivariate normal with covariance matrix  $G(\theta)^{-1}$ , briefly  $\hat{\theta}_c \sim N_d(\theta, G(\theta)^{-1})$ . Therefore, the asymptotic distribution of the Wald-type statistic  $cw^e(\theta) = (\hat{\theta}_c - \theta)^{\top}G(\theta)(\hat{\theta}_c - \theta)$  is chi-squared on  $d$  degrees of freedom,  $\chi_d^2$ . The same result holds for the score-type statistic  $cw^u(\theta) = cU(\theta)^{\top}J(\theta)^{-1}cU(\theta)$ .

Let  $cw(\theta) = 2\{c\ell(\hat{\theta}_c) - c\ell(\theta)\}$  be the composite likelihood ratio statistic. Its asymptotic null distribution is the distribution of  $\sum_{a=1}^d \mu_a(\theta)Z_a^2$ , where  $\mu_1(\theta), \dots, \mu_d(\theta)$  are eigenvalues of  $J(\theta)H(\theta)^{-1} = H(\theta)G(\theta)^{-1}$  and  $Z_1, \dots, Z_d$  are independent standard normal variates.

With the partition  $\theta = (\psi, \lambda)$ , partition the composite score vector as  $cU(\theta)^{\top} = (cU_{\psi}(\theta)^{\top}, cU_{\lambda}(\theta)^{\top})$ , where  $cU_{\psi}(\theta) = (\partial/\partial\psi)c\ell(\theta; y)$  and  $cU_{\lambda}(\theta) = (\partial/\partial\lambda)c\ell(\theta; y)$ , and consider the further partitions

$$H = \begin{bmatrix} H_{\psi\psi} & H_{\psi\lambda} \\ H_{\lambda\psi} & H_{\lambda\lambda} \end{bmatrix}, \quad H^{-1} = \begin{bmatrix} H^{\psi\psi} & H^{\psi\lambda} \\ H^{\lambda\psi} & H^{\lambda\lambda} \end{bmatrix},$$

and similarly for  $G$  and  $G^{-1}$ . Above, for instance,  $H_{\psi\lambda} = E_{\theta}\{-\partial cU_{\psi}(\theta)/\partial\lambda^{\top}\}$ .

Denoting by  $\hat{\theta}_{c\psi}$  the constrained composite maximum likelihood estimate of  $\theta$  for fixed  $\psi$ , we have that the asymptotic distribution of the profile composite likelihood ratio statistic for  $\psi$ ,

$$cw_p(\psi) = 2\{c\ell(\hat{\theta}_c) - c\ell(\hat{\theta}_{c\psi})\}, \quad (2.2)$$

is  $\sum_{a=1}^{d_0} \nu_a(\theta) Z_a^2$ , where  $\nu_1(\theta), \dots, \nu_{d_0}(\theta)$  are the eigenvalues of  $(H^{\psi\psi})^{-1}G^{\psi\psi}$ . This follows from Kent (1982, Thm. 3.1). It is possible to replace  $\theta$  with  $\hat{\theta}_{c\psi}$  when evaluating the eigenvalues of  $(H^{\psi\psi})^{-1}G^{\psi\psi}$ .

A Wald-type statistic for the  $\psi$  component based on  $c\ell(\theta)$  may be defined having an asymptotic  $\chi_{d_0}^2$  null distribution. In particular, denoting by  $\hat{\psi}_c$  the  $\psi$  component of  $\hat{\theta}_c$ , we have  $cw_p^e(\psi) = (\hat{\psi}_c - \psi)^\top (G^{\psi\psi})^{-1}(\hat{\psi}_c - \psi)$ .

Finally, using the asymptotic result (Rotnitzky and Jewell (1990)),

$$cU_\psi(\hat{\theta}_{c\psi}) \sim N_{d_0}(0, H^{\psi\psi}(\hat{\theta}_{c\psi})^{-1}G^{\psi\psi}(\hat{\theta}_{c\psi})H^{\psi\psi}(\hat{\theta}_{c\psi})^{-1}), \quad (2.3)$$

we consider the score type statistic

$$cw_p^u(\psi) = cU_\psi(\hat{\theta}_{c\psi})^\top H^{\psi\psi}(\hat{\theta}_{c\psi})\{G^{\psi\psi}(\hat{\theta}_{c\psi})\}^{-1}H^{\psi\psi}(\hat{\theta}_{c\psi})cU_\psi(\hat{\theta}_{c\psi}), \quad (2.4)$$

again with an asymptotic  $\chi_{d_0}^2$  null distribution. See Molenberghs and Verbeke (2005, Sec. 9.3) for more details.

### 3. Adjustments to Composite Likelihood Ratio

The asymptotic null distribution of the composite likelihood ratio statistic depends both on the statistical model and on the definition of the parameter of interest. This calls for adjustments to  $cw(\theta)$  and to  $cw_p(\psi)$  such that the reference sampling distributions depend only on the dimension of the parameter of interest. Problems without nuisance parameters are addressed first. For a scalar parameter of interest, most proposed adjustments agree and lead to the exact asymptotic reference. Beyond the scalar parameter case, some adjustments proposed in the literature are not parameterization invariant or only match some moments of the asymptotic reference. The adjustment we propose, free of both predicaments, is given at (3.8) and (3.9).

#### 3.1. No nuisance parameters

In the special case  $d = 1$ , we have  $\mu_1(\theta) = J(\theta)/H(\theta)$  in the asymptotic null distribution of  $cw(\theta)$ , so that the adjusted likelihood ratio statistic

$$cw(\theta)_{adj} = (\mu_1(\theta))^{-1} cw(\theta) \quad (3.1)$$

is asymptotically  $\chi_1^2$ .

For  $d > 1$  simple adjustments of  $cw(\theta)$  based on moment conditions have been considered by several authors. First order moment matching (Rotnitzky and Jewell (1990); Molenberghs and Verbeke (2005, Sec. 9.3.3)) gives the adjustment

$$cw(\theta)_1 = (\bar{\mu}(\theta))^{-1} cw(\theta), \quad (3.2)$$

with  $\bar{\mu}(\theta) = \sum_{a=1}^d \mu_a(\theta)/d = \text{tr}(J(\theta)H(\theta)^{-1})/d$ . A  $\chi_d^2$  approximation is used for the null distribution of  $cw(\theta)_1$ .

First and second moment matching gives the Satterthwaite type (Satterthwaite (1946)) adjustment suggested in Varin (2008),

$$cw(\theta)_2 = \kappa^{-1} cw(\theta), \quad (3.3)$$

where a  $\chi_\nu^2$  approximation is used for the null distribution of  $cw(\theta)_2$ . Here,

$$\begin{aligned} \kappa &= \kappa(\theta) = \frac{\sum_{a=1}^d \mu_a(\theta)^2}{\sum_{a=1}^d \mu_a(\theta)}, \\ \nu &= \nu(\theta) = \frac{(\sum_{a=1}^d \mu_a(\theta))^2}{\sum_{a=1}^d \mu_a(\theta)^2}. \end{aligned}$$

Matching of moments up to higher order can also be considered, as in Wood (1989). See also Lindsay, Pilla, and Basak (2000).

Chandler and Bate (2007) propose the vertical scaling of  $cw(\theta)$ ,

$$cw(\theta)_{CB} = \frac{(\hat{\theta}_c - \theta)^\top G(\hat{\theta}_c)(\hat{\theta}_c - \theta)}{(\hat{\theta}_c - \theta)^\top H(\hat{\theta}_c)(\hat{\theta}_c - \theta)} cw(\theta), \quad (3.4)$$

having the usual  $\chi_d^2$  asymptotic null distribution. The same result holds for the asymptotically equivalent form

$$cw(\theta)_{CB}^* = \frac{(\hat{\theta}_c - \theta)^\top G(\theta)(\hat{\theta}_c - \theta)}{(\hat{\theta}_c - \theta)^\top H(\theta)(\hat{\theta}_c - \theta)} cw(\theta). \quad (3.5)$$

When  $d = 1$  adjustments (3.2), (3.3), and (3.5) coincide with (3.1) and are parameterization invariant. Even (3.4) is parameterization invariant for a scalar  $\theta$ . When  $d > 1$ , among the above adjustments only those based on the eigenvalues of  $J(\theta)H(\theta)^{-1}$  or of  $J(\hat{\theta}_c)H(\hat{\theta}_c)^{-1}$  are invariant under reparameterization.

Indeed, consider an alternative parameterization  $\omega = \omega(\theta)$ , i.e., a one to one smooth function of  $\theta$  with inverse  $\theta(\omega)$ . Let  $\theta'(\omega) = (\partial\theta(\omega)/\partial\omega^\top)^\top$ . Use the

superscript  $\Omega$  to denote composite likelihood quantities under the  $\omega$  parameterization and get

$$\begin{aligned} c\ell^\Omega(\omega) &= c\ell(\theta(\omega)), \\ \hat{\omega}_c &= \omega(\hat{\theta}_c), \\ cw^\Omega(\omega) &= cw(\theta(\omega)), \\ cU^\Omega(\omega) &= \theta'(\omega)cU(\theta(\omega)), \\ H^\Omega(\omega) &= \theta'(\omega)H(\theta(\omega))\theta'(\omega)^\top, \\ J^\Omega(\omega) &= \theta'(\omega)J(\theta(\omega))\theta'(\omega)^\top, \\ G^\Omega(\omega) &= \theta'(\omega)G(\theta(\omega))\theta'(\omega)^\top, \\ J^\Omega(\omega)H^\Omega(\omega)^{-1} &= \theta'(\omega)J(\theta(\omega))H(\theta(\omega))^{-1}\theta'(\omega)^{-1}. \end{aligned}$$

The matrices  $J^\Omega(\omega)H^\Omega(\omega)^{-1}$  and  $J(\theta(\omega))H(\theta(\omega))^{-1}$  are similar, so that eigenvalues  $\mu_a(\theta)$ ,  $a = 1, \dots, d$ , are invariant under reparameterization.

On the other hand,

$$cw^\Omega(\omega)_{CB} = \frac{(\hat{\omega}_c - \omega)^\top \theta'(\hat{\omega}_c) G(\theta(\hat{\omega}_c)) \theta'(\hat{\omega}_c)^\top (\hat{\omega}_c - \omega)}{(\hat{\omega}_c - \omega)^\top \theta'(\hat{\omega}_c) H(\theta(\hat{\omega}_c)) \theta'(\hat{\omega}_c)^\top (\hat{\omega}_c - \omega)} cw(\theta(\omega)), \quad (3.6)$$

which does not, in general, coincide with  $cw(\theta(\omega))_{CB}$  (see Section 4.1 for a numerical example).

A parameterization invariant vertical scaling of  $cw(\theta)$  can be obtained as follows. Consider the expansion

$$\hat{\theta}_c - \theta = H(\theta)^{-1}cU(\theta) + O_p(n^{-1}). \quad (3.7)$$

Substituting (3.7) into (3.5) we obtain the asymptotically equivalent version

$$cw(\theta)_{INV} = \frac{cU(\theta)^\top J(\theta)^{-1}cU(\theta)}{cU(\theta)^\top H(\theta)^{-1}cU(\theta)} cw(\theta), \quad (3.8)$$

which is easily seen to be parameterization invariant, i.e., such that  $cw^\Omega(\omega)_{INV} = cw(\theta(\omega))_{INV}$ .

The denominator of the adjustment in (3.8) is a quadratic approximation to  $cw(\theta)$  such that  $cw(\theta) = cU(\theta)^\top H(\theta)^{-1}cU(\theta) + O_p(n^{-1/2})$ . As  $cU(\theta) \overset{\sim}{\sim} N_d(0, J(\theta))$ , the numerator of the adjustment has an asymptotic  $\chi_d^2$  null distribution. The asymptotic  $\chi_d^2$  null distribution of  $cw(\theta)_{INV}$  is therefore directly visible.

### 3.2. Nuisance parameter case

Adjustments  $cw_P(\psi)_1$  and  $cw_P(\psi)_2$  to  $cw_P(\psi)$  analogous to (3.2) and (3.3), respectively, can be defined with eigenvalues  $\nu_a(\theta)$  of  $(H^{\psi\psi})^{-1}G^{\psi\psi}$  evaluated at  $\hat{\theta}_{c\psi}$ .

The extension of (3.8) to the nuisance parameter case is obtained by modifying  $cw_P(\psi)$  using a factor defined as the ratio of an asymptotically  $\chi_{d_0}^2$  quadratic form in the profile composite score  $cU_\psi(\hat{\theta}_{c\psi})$  to a quadratic approximation of  $cw_P(\psi)$  expressed as a quadratic form in  $cU_\psi(\hat{\theta}_{c\psi})$ . Using the asymptotic result (2.3), leading to (2.4), and the expansion

$$cw_P(\psi) = cU_\psi(\hat{\theta}_{c\psi})^\top H^{\psi\psi}(\hat{\theta}_{c\psi})cU_\psi(\hat{\theta}_{c\psi}) + O_p(n^{-1/2}),$$

we obtain

$$cw_P(\psi)_{INV} = \frac{cw_P^u(\psi)}{cU_\psi(\hat{\theta}_{c\psi})^\top H^{\psi\psi}(\hat{\theta}_{c\psi})cU_\psi(\hat{\theta}_{c\psi})} cw_P(\psi). \quad (3.9)$$

It is straightforward to check that  $cw_P(\psi)_{INV}$  is invariant under interest-preserving reparameterizations, that is one-one functions that map  $\theta = (\psi, \lambda)$  to  $\omega = (\eta(\psi), \zeta(\psi, \lambda))$ . When  $d_0 = 1$  expression (3.9) simplifies as

$$cw_P(\psi)_{INV} = \frac{H^{\psi\psi}(\hat{\theta}_{c\psi})}{G^{\psi\psi}(\hat{\theta}_{c\psi})} cw_P(\psi)$$

and coincides with the nuisance parameter version of (3.2) and (3.3).

#### 4. Examples

In this section we provide simulation results to assess coverage probabilities of confidence regions based on the adjustments considered in Section 3. Two examples are studied, both dealing with multivariate normal distributions with structured covariance matrix. They are chosen so that we can easily do closed form calculations both of complete and pairwise likelihood quantities, not for their direct interest in application of composite likelihood. In both examples we have  $d = 3$  and situations with  $d_0 = 1$  and  $d_0 = 2$  are considered.

##### 4.1. Equicorrelated multivariate normal data

Consider a one-way normal-theory random effects model in which a component  $Y_{ir}$  of the  $i$ th vector has the form  $Y_{ir} = \mu + \xi_i + \varepsilon_{ir}$ ,  $i = 1, \dots, n$ ,  $r = 1, \dots, q$ , where  $\xi_i$  and  $\varepsilon_{ir}$  are independently normally distributed with zero mean and variances respectively  $\sigma_\xi^2$  and  $\sigma_\varepsilon^2$ . The model can be reparameterized in various ways. For instance, the problem can be reformulated by writing  $Y_i$  as a multivariate normal with components having mean  $\mu$  and variance  $\sigma^2 = \sigma_\xi^2 + \sigma_\varepsilon^2$ , and with correlation  $\rho = \sigma_\xi^2 / (\sigma_\xi^2 + \sigma_\varepsilon^2)$  between any two components of the same vector.

The special case with  $\mu = 0$  and  $\sigma^2 = 1$ , where  $\theta = \rho$ , has been treated in detail by Cox and Reid (2004). Here we examine the case with  $\theta = (\mu, \rho, \sigma^2)$ .

The model is a full exponential family with log likelihood given by

$$\ell(\theta) = -\frac{nq}{2} \log \sigma^2 - \frac{n(q-1)}{2} \log(1-\rho) - \frac{n}{2} \log\{1 + \rho(q-1)\} - \frac{SS_W}{2\sigma^2(1-\rho)} - \frac{qSS_B + nq(\bar{y} - \mu)^2}{2\sigma^2\{1 + \rho(q-1)\}}, \quad (4.1)$$

where, denoting by  $\bar{y}_i = \sum_{r=1}^q y_{ir}/q$ ,

$$\bar{y} = \frac{1}{nq} \sum_{i=1}^n \sum_{r=1}^q y_{ir}, \quad SS_B = \sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \bar{y}_i)^2, \quad SS_W = \sum_{i=1}^n (\bar{y}_i - \bar{y})^2$$

give the sufficient statistic.

Given the equicorrelation within any possible pair of components of  $Y_i$ , we consider the pairwise log likelihood (2.1) with all weights equal to 1. The generic pair  $(Y_{ir}, Y_{is})$ ,  $r \neq s$ , has a bivariate normal distribution with the two components having mean  $\mu$ , variance  $\sigma^2$ , and correlation  $\rho$ . Therefore,

$$p\ell(\theta) = -\frac{nq(q-1)}{2} \log \sigma^2 - \frac{nq(q-1)}{4} \log(1-\rho^2) - \frac{q-1+\rho}{2\sigma^2(1-\rho^2)} SS_W - \frac{q(q-1)SS_B + nq(q-1)(\bar{y} - \mu)^2}{2\sigma^2(1+\rho)}. \quad (4.2)$$

This model is rather peculiar in the sense that the estimate from the pairwise likelihood coincides with full maximum likelihood estimate,  $\hat{\theta}_p = \hat{\theta}$  (Mardia et al. (2009)). Moreover E. C. Kenne Pagui, in an unpublished work, showed that  $pU(\theta) = J(\theta)H(\theta)^{-1}U(\theta)$ . Hence, even though  $H(\theta) \neq J(\theta)$ , we have that  $I(\theta) = G(\theta)$ . The two score functions,  $U(\theta)$  and  $pU(\theta)$ , and the matrices  $I(\theta)$ ,  $J(\theta)$ , and  $H(\theta)$  are reported in the Appendix.

As a consequence of this particular property of the model, the Wald and score statistics based on the pairwise likelihood coincide with the analogous quantities based on the full likelihood, at least without nuisance parameters. This is not the case for the pairwise likelihood ratio statistic and its adjustments.

In order to assess the quality of the proposed approximations, we ran an illustrative simulation experiment, with  $n = 5$  and  $q = 30$ , and with three values of  $\rho$ , ranging from a moderate to a strong correlation. Such a setting, with small  $n$  relative to  $q$ , is likely to show problems for the pairwise likelihood, in particular when  $\rho$  is one of the components of interest. We first considered the case in which the overall parameter  $\theta$  is of interest. Table 1 reports the empirical coverages of confidence regions based on several statistics: likelihood ratio, score and Wald statistics from the full model, and adjustments (3.2), (3.3), (3.4) and (3.8) of the pairwise likelihood ratio statistic.

Table 1. Equicorrelated multivariate normal model. Empirical coverage of three-dimensional confidence regions based on different statistics in three simulations, each with 100,000 replications, with  $n = 5$ ,  $q = 30$ ,  $\mu = 0$ ,  $\sigma^2 = 1$ , and  $\rho = 0.2, 0.5, 0.9$ . Note that  $pw^e(\theta) = w^e(\theta)$  and  $pw^u(\theta) = w^u(\theta)$ .

	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.9$		
	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
$w(\theta)$	0.856	0.920	0.978	0.856	0.921	0.981	0.855	0.918	0.980
$w^e(\theta)$	0.921	0.962	0.991	0.746	0.807	0.890	0.488	0.525	0.580
$w^u(\theta)$	0.903	0.941	0.976	0.904	0.940	0.975	0.902	0.940	0.976
$pw(\theta)_{INV}$	0.910	0.952	0.987	0.896	0.947	0.988	0.838	0.903	0.969
$pw(\theta)_1$	0.886	0.931	0.975	0.886	0.936	0.981	0.832	0.896	0.968
$pw(\theta)_{CB}$	0.691	0.746	0.819	0.699	0.761	0.847	0.528	0.578	0.657
$pw(\theta)_2$	0.910	0.953	0.989	0.910	0.956	0.991	0.859	0.926	0.986

We note that the proposed adjustment (3.8) shows a reasonable performance in terms of coverage, even outperforming the likelihood ratio from the full model except when there is very large correlation. Adjustments (3.2) and (3.3) based on moments also gave quite good results, with the latter being overall very accurate. On the contrary, Wald and Chandler and Bate's adjustment (3.4) showed generally rather poor coverages.

A different aspect of confidence regions based on the pairwise likelihood, not reflecting on coverage properties, is their shape. We might think of the shape of the confidence region based on the likelihood ratio from the full model as a gold standard. Figure 1 displays the confidence regions with nominal level 0.95 based on the statistics considered in Table 1, for a simulated sample with  $n = 5$ ,  $q = 30$ ,  $\rho = 0.5$ ,  $\sigma^2 = 1$ , and  $\mu = 0$ , giving  $\bar{y} = -0.065$ ,  $SS_W = 69.869$ , and  $SS_B = 1.226$ . The parameter  $\mu$  is considered as known in this analysis. From the plots we see that the score statistic has a degenerate shape in one direction. On the contrary,  $pw(\theta)_{INV}$  seems to mitigate this problem and, overall, is the solution that gives closest agreement with  $w(\theta)$ . Unreported simulation results in this two-parameter case showed similar results to those for the three-parameter case, although with better performances of all statistics.

As a numerical example of the lack of parameterization invariance of (3.4), using the same data as those used for producing plots in Figure 1, consider testing  $H_0 : \theta = \theta_0$ , with  $\theta_0 = (0, 0.7, 1)$ , in the three parameter model. We obtained  $pw(\theta_0)_{CB} = 7.725$  with approximate  $p$ -value 0.052. With the alternative parameterization  $\omega = (\mu, 0.5 \log((1+\rho)/(1-\rho)), \sigma^2)$  the problem becomes testing  $H_0 : \omega = \omega_0$ , with  $\omega_0 = (0, 0.867, 1)$  and we obtained  $cw^\Omega(\omega_0)_{CB} = 8.767$  with approximate  $p$ -value 0.033.

We now consider the nuisance parameter case. First take  $\psi = (\rho, \sigma^2)$  as a two-dimensional parameter of interest, with  $\mu$  treated as a nuisance parameter.

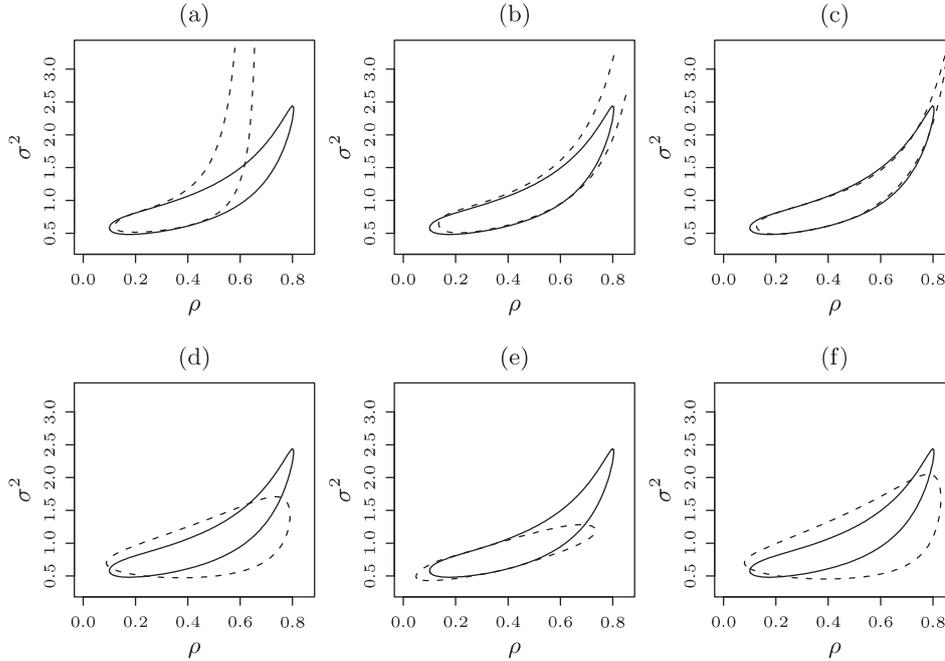


Figure 1. Equicorrelated multivariate normal model. Confidence regions with level 0.95 for  $(\rho, \sigma^2)$ , with  $\mu$  known and equal to zero, for a simulated sample with  $n = 5$ ,  $q = 30$ , and true parameter value  $\mu = 0$ ,  $\rho = 0.5$ , and  $\sigma^2 = 1$ . In each plot, the solid line corresponds to  $w(\theta)$ , while the dashed line corresponds to: (a)  $w^e(\theta)$ ; (b)  $w^u(\theta)$ ; (c)  $pw(\theta)_{INV}$ ; (d)  $pw(\theta)_1$ ; (e)  $pw(\theta)_{CB}$ ; (f)  $pw(\theta)_2$ .

The constrained estimate of  $\mu$  for fixed  $\psi$  does not depend on  $\psi$  and is equal to  $\bar{y}$ , this for both the full and pairwise likelihood. For this reason, the shapes of confidence regions, not displayed here, are very close to those in Figure 1 for the case with  $\mu$  known. On the other hand, coverages are slightly affected by the presence of the nuisance parameter. Table 2 reports the results of a simulation in the same setting as the one leading to the results in Table 1. The same conclusions emerge. We did not consider Chandler and Bate's adjustment, given its poor coverage shown in Table 1 and its lack of invariance.

Finally, we restrict our attention to  $\rho$  as a scalar parameter of interest, with  $\lambda = (\mu, \sigma^2)$  treated as a nuisance parameter. With a scalar parameter of interest, all adjustments of the pairwise likelihood ratio coincide with (3.1). On the contrary, the constrained estimate of  $\sigma^2$  for fixed  $\psi$  computed from the full likelihood does not coincide with the one from the pairwise likelihood and, therefore, Wald and score statistics no longer coincide in the two cases. Table 3 reports the empirical coverage probabilities of confidence intervals for  $\rho$  in a simulation

Table 2. Equicorrelated multivariate normal model. Empirical coverage of profile confidence regions for  $\psi = (\rho, \sigma^2)$  when  $\mu$  is treated as a nuisance parameter, based on different statistics. Results of three simulations, each with 100,000 replications, with  $n = 5$ ,  $q = 30$ ,  $\mu = 0$ ,  $\sigma^2 = 1$ , and  $\rho = 0.2, 0.5, 0.9$ . Note that  $pw_P^e(\theta) = w_P^e(\theta)$  and  $pw_P^u(\theta) = w_P^u(\theta)$ .

	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.9$		
	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
$w_P(\psi)$	0.846	0.912	0.975	0.847	0.913	0.977	0.846	0.912	0.976
$w_P^e(\psi)$	0.935	0.969	0.992	0.735	0.793	0.874	0.480	0.515	0.569
$w_P^u(\psi)$	0.925	0.960	0.988	0.926	0.961	0.988	0.925	0.960	0.988
$pw_P(\psi)_{INV}$	0.921	0.962	0.992	0.890	0.946	0.988	0.823	0.891	0.961
$pw_P(\psi)_1$	0.908	0.952	0.987	0.862	0.929	0.989	0.803	0.863	0.943
$pw_P(\psi)_2$	0.925	0.967	0.994	0.898	0.966	0.997	0.837	0.908	0.981

Table 3. Equicorrelated multivariate normal model. Empirical coverage of profile confidence intervals for  $\psi = \rho$  when  $(\mu, \sigma^2)$  is treated as a nuisance parameter, based on different statistics. Results of three simulations, each with 100,000 replications, with  $n = 5$ ,  $q = 30$ ,  $\mu = 0$ ,  $\sigma^2 = 1$  and  $\rho = 0.2, 0.5, 0.9$ .

	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.9$		
	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
$w_P(\psi)$	0.827	0.896	0.968	0.826	0.896	0.969	0.826	0.895	0.969
$w_P^e(\psi)$	0.926	0.989	0.998	0.810	0.881	0.970	0.684	0.729	0.795
$w_P^u(\psi)$	0.963	0.976	0.990	0.963	0.976	0.990	0.962	0.976	0.990
$pw_P(\psi)_{INV}$	0.935	0.987	0.997	0.862	0.940	0.999	0.798	0.864	0.946
$pw_P^e(\psi)$	0.926	0.989	0.998	0.810	0.881	0.970	0.684	0.729	0.795
$pw_P^u(\psi)$	0.940	0.988	0.997	0.879	0.959	1.000	0.825	0.899	0.985

study with the same setting as that leading to Table 1. All confidence intervals seem to suffer the particular setting with small  $n$  and large  $q$ , with most statistics showing a poor coverage. An exception is given by the two score statistics, with the one from the full model always being strongly conservative. This behaviour can be explained by the plots of the confidence intervals in Figure 2 corresponding to the same simulated sample as that used for Figure 1. Indeed, the score statistic from the full model is increasing very slowly for large values of  $\rho$ , thus giving a very large interval. This feature is not shared by the pairwise score. Finally, we note the close agreement between the adjusted pairwise likelihood ratio and the full likelihood ratio, still considered as the gold standard.

#### 4.2. First order autoregression

Consider a normal autoregressive process of order one. We use the parameterization given in Davison (2003, Example 6.24), such that  $Y_{ir} - \mu =$

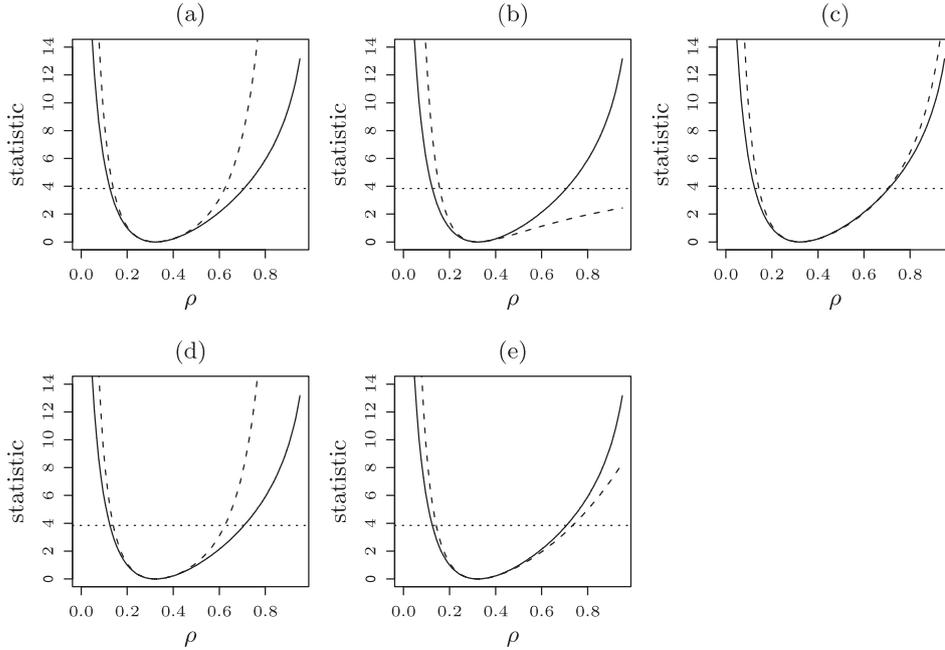


Figure 2. Equicorrelated multivariate normal model. Confidence intervals with level 0.95 for  $\rho$  when  $(\mu, \sigma^2)$  is considered as a nuisance parameter, for a simulated sample with  $n = 5$ ,  $q = 30$  and true parameter value  $\mu = 0$ ,  $\rho = 0.5$ , and  $\sigma^2 = 1$ . In each plot, the solid line corresponds to  $w_P(\theta)$ , while the dashed line corresponds to: (a)  $w_P^e(\rho)$ ; (b)  $w_P^u(\rho)$ ; (c)  $pw_P(\rho)_{INV}$ ; (d)  $pw_P^e(\rho)$ ; (e)  $pw_P^u(\rho)$ .

$\rho(Y_{ir-1} - \mu) + \varepsilon_{ir}$ ,  $i = 1, \dots, n$ ,  $r = 2, \dots, q$ , where  $\varepsilon_{ir}$  are independently normally distributed with zero mean and variance  $\sigma^2$ . This implies that we have  $n$  observations, with the  $i$ th vector  $Y_i$  having a multivariate normal distribution with components having mean  $\mu$  and covariance between  $Y_{ir}$  and  $Y_{is}$  equal to  $\sigma^2 \rho^{|r-s|} / (1 - \rho^2)$ ,  $r, s = 1, \dots, q$ .

In the following, we consider inference on the parameter  $\theta = (\mu, \rho, \sigma^2)$ , restricting our attention to a single series. Dropping the subscript  $i$  from the notation, the log likelihood from the full model is

$$\ell(\theta) = -\frac{1}{2\sigma^2} \left\{ \sum_{r=1}^q (y_r - \mu)^2 + \rho^2 \sum_{r=2}^{q-1} (y_r - \mu)^2 - 2\rho \sum_{r=2}^q (y_r - \mu)(y_{r-1} - \mu) \right\} - \frac{q}{2} \log \sigma^2 + \frac{1}{2} \log(1 - \rho^2). \quad (4.3)$$

In this model, we construct the pairwise log likelihood (2.1) using only pairs of contiguous components; hence, the only non-zero weights are  $w_{rs} = 1$  when

Table 4. First order autoregression. Empirical coverage of three-dimensional confidence regions based on different statistics in three simulations, each with 100,000 replications, with  $q = 30$ ,  $\mu = 0$ ,  $\sigma^2 = 1$ , and  $\rho = 0.2, 0.5, 0.9$ .

	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.9$		
	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
$w(\theta)$	0.889	0.942	0.988	0.881	0.938	0.986	0.819	0.898	0.974
$w^e(\theta)$	0.866	0.921	0.974	0.832	0.893	0.958	0.582	0.650	0.753
$w^u(\theta)$	0.905	0.942	0.977	0.903	0.939	0.975	0.902	0.932	0.967
$pw(\theta)_{INV}$	0.894	0.946	0.988	0.892	0.945	0.987	0.860	0.921	0.978
$pw(\theta)_1$	0.885	0.937	0.985	0.882	0.935	0.982	0.879	0.933	0.981
$pw(\theta)_{CB}$	0.879	0.933	0.982	0.854	0.914	0.972	0.681	0.768	0.883
$pw(\theta)_2$	0.892	0.944	0.988	0.895	0.947	0.988	0.904	0.955	0.991

Table 5. First order autoregression. Empirical coverage of profile confidence regions for  $\psi = (\rho, \sigma^2)$  when  $\mu$  is treated as a nuisance parameter, based on different statistics. Results of three simulations, each with 100,000 replications, with  $q = 30$ ,  $\mu = 0$ ,  $\sigma^2 = 1$ , and  $\rho = 0.2, 0.5, 0.9$ .

	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.9$		
	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
$w_P(\psi)$	0.887	0.942	0.987	0.876	0.935	0.985	0.796	0.882	0.969
$w_P^e(\psi)$	0.863	0.918	0.972	0.826	0.885	0.952	0.560	0.628	0.733
$w_P^u(\psi)$	0.921	0.955	0.984	0.927	0.960	0.986	0.950	0.973	0.992
$pw_P(\psi)_{INV}$	0.893	0.946	0.989	0.887	0.944	0.989	0.833	0.906	0.973
$pw_P(\psi)_1$	0.884	0.937	0.983	0.874	0.930	0.981	0.834	0.920	0.990
$pw_P(\psi)_2$	0.891	0.942	0.987	0.885	0.940	0.987	0.875	0.959	0.999

$r - s = 1$ ,  $r, s = 1, \dots, q$ . In general, for autoregressive processes of higher order, pairs at larger distances should be included in the pairwise likelihood or, alternatively, a composite likelihood based on higher dimensional margins could be used. A generic pair of successive components has a bivariate normal distribution with the two components having mean  $\mu$ , variance  $\sigma^2/(1 - \rho^2)$ , and correlation  $\rho$ . Hence, the pairwise log likelihood is

$$p\ell(\theta) = -\frac{1}{2\sigma^2} \left\{ \sum_{r=2}^q (y_r - \mu)^2 + \sum_{r=2}^q (y_{r-1} - \mu)^2 - 2\rho \sum_{r=2}^q (y_r - \mu)(y_{r-1} - \mu) \right\} - (q-1) \log \sigma^2 + \frac{q-1}{2} \log(1 - \rho^2). \quad (4.4)$$

All results reported below refer to the  $(\mu, \rho, \log \sigma^2)$  parameterization. Of course, this choice is only relevant for the non-invariant statistics.

Simulation results with the overall parameter  $\theta$  of interest are summarized in Table 4. The picture does not depart remarkably from that of the previous example. Results for pairwise Wald and score-type statistics are not reported, being

Table 6. First order autoregression. Empirical coverage of profile confidence intervals for  $\psi = \rho$  when  $(\mu, \sigma^2)$  is treated as a nuisance parameter, based on different statistics. Results of three simulations, each with 100,000 replications, with  $q = 30$ ,  $\mu = 0$ ,  $\sigma^2 = 1$ , and  $\rho = 0.2, 0.5, 0.9$ .

	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.9$		
	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
$w_P(\psi)$	0.892	0.945	0.988	0.876	0.934	0.984	0.760	0.855	0.959
$w_P^e(\psi)$	0.889	0.943	0.987	0.844	0.899	0.960	0.545	0.608	0.713
$w_P^u(\psi)$	0.911	0.961	0.995	0.926	0.973	0.996	0.993	0.997	0.999
$pw_P(\psi)_{INV}$	0.893	0.945	0.989	0.878	0.936	0.986	0.785	0.893	0.987
$pw_P^e(\psi)$	0.891	0.944	0.987	0.844	0.899	0.960	0.521	0.591	0.704
$pw_P^u(\psi)$	0.901	0.954	0.993	0.890	0.949	0.993	0.866	0.975	1.000

undistinguishable from those for the analogous statistics for the full model, although the statistics do not coincide. The proposed adjustment (3.8) shows again a reasonable performance in terms of coverage, and is uniformly more accurate than the likelihood ratio from the full model. Moment matching adjustments, especially (3.3), perform very well. On the contrary, Wald statistic and Chandler and Bate's adjustment (3.4) show generally rather poor coverages.

The analogous results for two problems with nuisance parameters are summarized in Tables 5 and 6. For results in Table 5, we consider  $\psi = (\rho, \sigma^2)$  and  $\mu$  as a nuisance. Results for pairwise Wald and score-type statistics are still very close to those for the analogous statistics for the full model and are therefore omitted. The best adjustments are (3.9) and the nuisance parameter version of (3.3), with the latter giving higher accuracy under strong correlation. Confidence regions from a simulated sample with  $q = 50$  and true parameter value  $\mu = 0$ ,  $\rho = 0.9$ , and  $\sigma^2 = 1$  are displayed in Figure 3. Score-type confidence regions have quite an unusual shape and depart remarkably from the one based on the full likelihood ratio. On the contrary, the region based on (3.9) largely overlaps with the one based on the full likelihood ratio, although it inherits some anomaly from the score. Finally, results in Table 6 summarize coverage estimates for confidence intervals for  $\rho$  when  $(\mu, \sigma^2)$  is treated as a nuisance parameter. The pairwise score interval shows the highest accuracy. On the other hand, the scale adjusted  $pw_P(\psi)_{INV} = pw_P(\psi)_1$  improves on  $w_P(\psi)$ . Confidence intervals for  $\rho$ , for the same simulated sample as in Figure 3, are displayed in Figure 4. We note that  $pw_P(\psi)_{INV}$  matches closely to  $w_P(\psi)$ , while the pairwise score exhibits an anomalous behavior for large values of  $\rho$ .

## 5. Concluding Remarks

The simulation studies in Section 4 indicate that the parameterization invariant adjustments of composite likelihood ratio statistics (3.8) and (3.9) yield

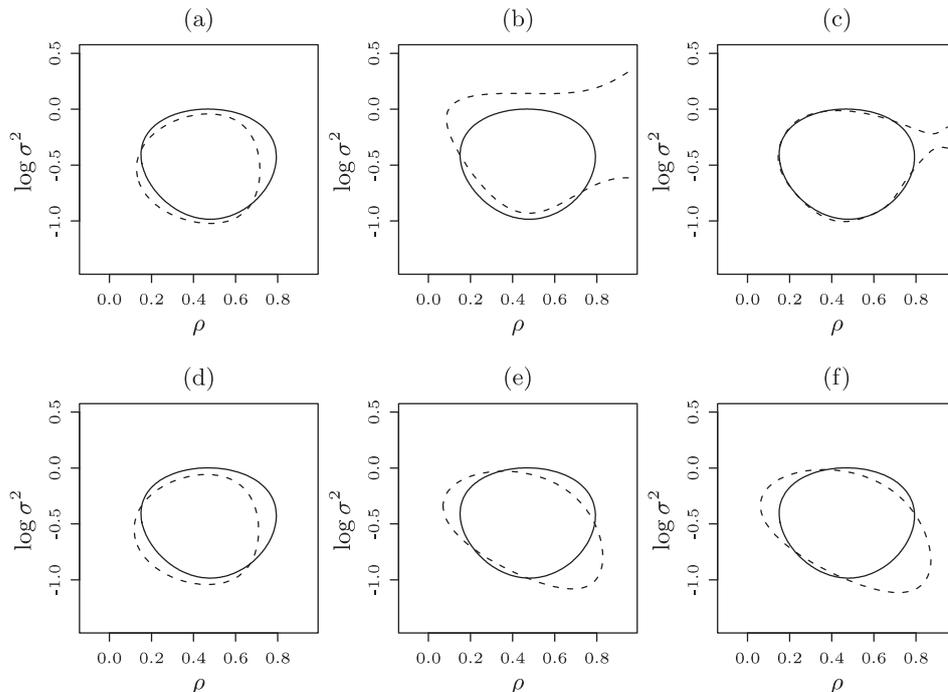


Figure 3. First order autoregression. Confidence regions with level 0.95 for  $(\rho, \sigma^2)$  when  $\mu$  is treated as a nuisance parameter, for a simulated sample with  $q = 50$  and true parameter value  $\mu = 0$ ,  $\rho = 0.9$ , and  $\sigma^2 = 1$ . In each plot, the solid line corresponds to  $w(\theta)$ , while the dashed line corresponds to: (a)  $w_P^e(\theta)$ ; (b)  $w_P^u(\theta)$  and  $p w_P^u(\rho)$ ; (c)  $p w_P(\theta)_{INV}$ ; (d)  $p w_P(\theta)^e$ ; (e)  $p w_P(\theta)_1$ ; (f)  $p w_P(\theta)_2$ .

confidence regions whose coverage and shape properties are comparable to, or even better than those based on moment-matching corrections. Moreover, statistics based on the composite likelihood ratio are to be preferred to score and Wald-type statistics. These results agree with those in Geys and Molenberghs (2002), focused on power estimation and check of type I error probabilities in problems without nuisance parameters.

In more realistic applications, analytic expressions for  $H(\theta)$  and  $J(\theta)$  may be unavailable, and evaluation of  $H(\hat{\theta}_c)$  and  $J(\hat{\theta}_c)$  seems to offer the most straightforward solution. In this way, however, exact invariance is lost. Moreover, simulation results, not reported here, dealing with the same models as in Section 4, show that accuracy in coverage decays. On the other hand, use of (3.5) in place of (3.4) does not generally yield a remarkable improvement in coverage, especially when the parameter is close to the boundary of the parameter space.

We also note that evaluation of  $H(\theta)$  and  $J(\theta)$  at various  $\theta$  values is not an

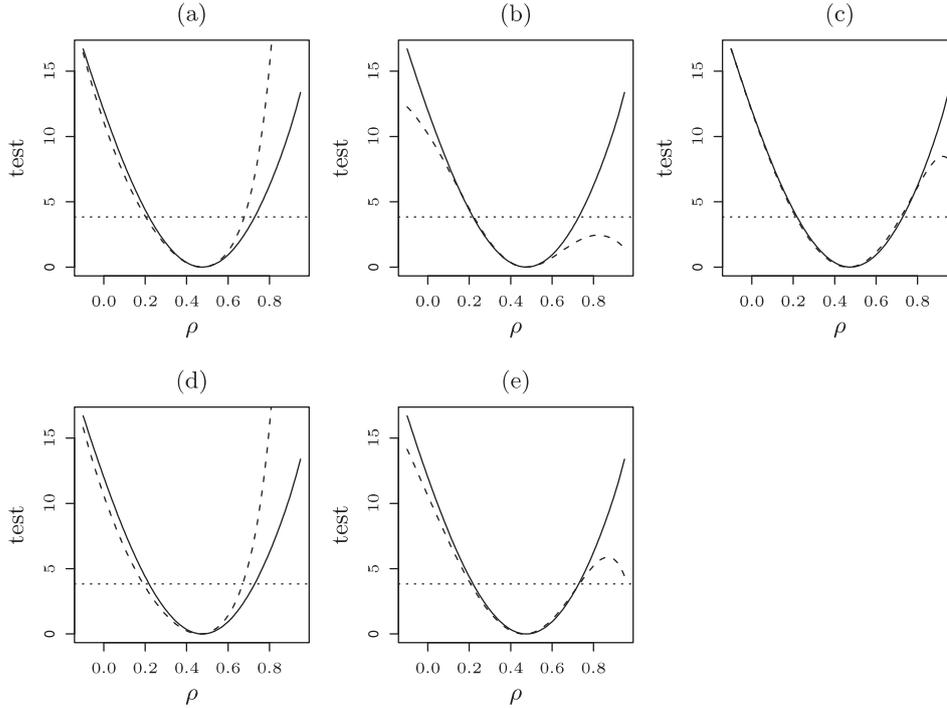


Figure 4. First order autoregression. Confidence intervals with level 0.95 for  $\rho$  when  $(\mu, \sigma^2)$  is considered as a nuisance parameter, for a simulated sample with  $q = 50$  and true parameter value  $\mu = 0$ ,  $\rho = 0.9$ , and  $\sigma^2 = 1$ . In each plot, the solid line corresponds to  $w_P(\rho)$ , while the dashed line corresponds to: (a)  $w_P^e(\rho)$ ; (b)  $w_P^u(\rho)$ ; (c)  $pw_P(\rho)_{INV}$ ; (d)  $pw_P^e(\rho)$ ; (e)  $pw_P^u(\rho)$ .

issue in hypothesis testing on a specific value of the parameter of interest, but is only required for the construction of confidence regions. These in turn are of interest, particularly when  $d_0 = 1$  or  $d_0 = 2$ . In such cases, estimation of matrices  $H(\theta)$ , and  $J(\theta)$  over a grid of values of the parameter of interest seems to be the strategy leading to the most accurate coverage, while preserving parameterization invariance. For a recent overview of computational aspects related to matrices  $H(\theta)$  and  $J(\theta)$  see Varin, Reid, and Firth (2011, Sec. 5.1).

Parameterization invariance, although in itself an appealing property, does not perhaps fully explain the satisfactory sampling behaviour of the proposed adjustments. A theoretical explanation is an open issue for further research.

Under the general setting of model misspecification entailing failure of the second Bartlett identity, Stafford (1996) and Royall and Tsou (2003) consider adjusting a working likelihood function in order to recover the familiar asymptotic chi-squared distribution for the likelihood ratio statistic. Their treatment requires a scalar parameter of interest and leads to a robust adjustment that

may be seen as an empirical version of adjustments considered in this paper when  $d = 1$ , or  $d_0 = 1$ . An extension to a multidimensional parameter of interest of robust adjusted likelihood functions could be based on (3.8) and (3.9), with empirical versions of  $H(\theta)$  and  $J(\theta)$ .

### Acknowledgements

We thank the referees and an Associate Editor for their useful comments.

### Appendix

In the equicorrelated example of Section 4.1 the score function for the full log likelihood (4.1) has elements

$$\begin{aligned} U_\mu(\theta) &= \frac{nq(\bar{y} - \mu)}{\sigma^2\{1 + \rho(q-1)\}}, \\ U_\rho(\theta) &= -\frac{SS_W}{2\sigma^2(1-\rho)^2} + \frac{q(q-1)\{SS_B + n(\bar{y} - \mu)^2\}}{2\sigma^2\{1 + \rho(q-1)\}^2} + \frac{n(q-1)}{2(1-\rho)} \\ &\quad - \frac{n(q-1)}{2\{1 + \rho(q-1)\}}, \\ U_{\sigma^2}(\theta) &= \frac{SS_W}{2(\sigma^2)^2(1-\rho)} + \frac{q\{SS_B + n(\bar{y} - \mu)^2\}}{2(\sigma^2)^2\{1 + \rho(q-1)\}} - \frac{nq}{2\sigma^2}, \end{aligned}$$

while the expected information matrix has elements

$$\begin{aligned} I_{\mu\mu}(\theta) &= \frac{nq}{\sigma^2\{1 + \rho(q-1)\}}, \quad I_{\rho\rho} = \frac{n(q-1)}{2} \left[ \frac{1}{(1-\rho)^2} + \frac{q-1}{\{1 + \rho(q-1)\}^2} \right], \\ I_{\sigma^2\sigma^2}(\theta) &= \frac{nq}{2(\sigma^2)^2}, \quad I_{\mu\rho}(\theta) = I_{\mu\sigma^2}(\theta) = 0, \quad I_{\rho\sigma^2} = -\frac{nq(q-1)\rho}{2\sigma^2(1-\rho)\{1 + \rho(q-1)\}}. \end{aligned}$$

The score function for the pairwise log likelihood (4.2) has elements

$$\begin{aligned} pU_\mu(\theta) &= \frac{nq(q-1)(\bar{y} - \mu)}{\sigma^2(1-\rho)}, \\ pU_\rho(\theta) &= -\frac{\rho^2 + 2\rho(q-1) + 1}{2\sigma^2(1-\rho^2)^2} SS_W + \frac{q(q-1)\{SS_B + n(\bar{y} - \mu)^2\}}{2\sigma^2(1+\rho)^2} + \frac{nq(q-1)\rho}{2(1-\rho^2)}, \\ pU_{\sigma^2}(\theta) &= \frac{q-1+\rho}{2(\sigma^2)^2(1-\rho^2)} SS_W + \frac{q(q-1)\{SS_B + n(\bar{y} - \mu)^2\}}{2(\sigma^2)^2(1+\rho)^2} - \frac{nq(q-1)}{2\sigma^2}. \end{aligned}$$

The matrix  $H(\theta)$  has elements

$$\begin{aligned} H_{\mu\mu}(\theta) &= \frac{nq(q-1)}{\sigma^2(1+\rho)}, \quad H_{\rho\rho} = \frac{nq(q-1)(1+\rho^2)}{2(1-\rho^2)^2}, \\ H_{\sigma^2\sigma^2}(\theta) &= \frac{nq(q-1)}{2(\sigma^2)^2}, \quad H_{\mu\rho}(\theta) = H_{\mu\sigma^2}(\theta) = 0, \quad H_{\rho\sigma^2} = -\frac{nq(q-1)\rho}{2\sigma^2(1-\rho^2)}, \end{aligned}$$

while the matrix  $J(\theta)$  has

$$\begin{aligned} J_{\mu\mu}(\theta) &= \frac{nq(q-1)^2\{1+\rho(q-1)\}}{\sigma^2(1+\rho)^2}, \quad J_{\mu\rho}(\theta) = J_{\mu\sigma^2}(\theta) = 0, \\ J_{\rho\rho}(\theta) &= \frac{n(q-1)}{2(1+\rho)^4} \left[ \frac{\{\rho^2+2\rho(q-1)+1\}^2}{(1-\rho)^2} + (q-1)\{1+\rho(q-1)\}^2 \right], \\ J_{\rho\sigma^2}(\theta) &= \frac{nq(q-1)\rho}{2\sigma^2(\rho-1)(\rho+1)^3} \{q^2\rho^2 - 3q\rho^2 + 3\rho^2 - q^2\rho + 5q\rho - 4\rho + 1\}, \\ J_{\sigma^2\sigma^2}(\theta) &= \frac{n(q-1)}{2(\sigma^2)^2(1+\rho)^2} \left[ (q-1+\rho)^2 + (q-1)\{1+\rho(q-1)\}^2 \right]. \end{aligned}$$

We now consider the score functions and the matrices  $I(\theta)$ ,  $H(\theta)$ , and  $J(\theta)$  for the first order autoregressive model in the example of Section 4.2. We start with the full likelihood (4.3), whose score function has the following elements

$$\begin{aligned} U_{\mu}(\theta) &= \frac{1}{\sigma^2} \left\{ \sum_{r=1}^q (y_r - \mu) + \rho^2 \sum_{r=2}^{q-1} (y_r - \mu) - \rho \sum_{r=2}^q (y_{r-1} - \mu) - \rho \sum_{r=2}^q (y_r - \mu) \right\}, \\ U_{\rho}(\theta) &= -\frac{\rho}{1-\rho^2} - \frac{1}{\sigma^2} \left\{ \rho \sum_{r=2}^{q-1} (y_r - \mu)^2 - \sum_{r=2}^q (y_{r-1} - \mu)(y_r - \mu) \right\}, \\ U_{\sigma^2}(\theta) &= \frac{1}{2(\sigma^2)^2} \left\{ \sum_{r=1}^q (y_r - \mu)^2 + \rho^2 \sum_{r=2}^{q-1} (y_r - \mu)^2 - 2\rho \sum_{r=2}^q (y_r - \mu)(y_{r-1} - \mu) \right\} \\ &\quad - \frac{q}{2\sigma^2}, \end{aligned}$$

while the expected information matrix has elements

$$\begin{aligned} I_{\mu\mu}(\theta) &= \frac{1}{\sigma^2} \{(q-1)(1-\rho)^2 + 1 - \rho^2\}, \quad I_{\rho\rho} = \frac{1 + \rho^2 + (q-2)(1-\rho^2)}{(1-\rho^2)^2}, \\ I_{\sigma^2\sigma^2}(\theta) &= \frac{q}{2(\sigma^2)^2}, \quad I_{\mu\rho}(\theta) = I_{\mu\sigma^2}(\theta) = 0, \quad I_{\rho\sigma^2} = \frac{\rho}{\sigma^2(1-\rho^2)}. \end{aligned}$$

The score function for the pairwise log likelihood (4.4) has elements

$$\begin{aligned} pU_{\mu}(\theta) &= \frac{1-\rho}{\sigma^2} \left\{ \sum_{r=2}^q (y_r - \mu) + \sum_{r=2}^q (y_{r-1} - \mu) \right\}, \\ pU_{\rho}(\theta) &= -\frac{(q-1)\rho}{1-\rho^2} + \frac{1}{\sigma^2} \sum_{r=2}^q (y_r - \mu)(y_{r-1} - \mu), \\ pU_{\sigma^2}(\theta) &= \frac{1}{2(\sigma^2)^2} \left\{ \sum_{r=2}^q (y_r - \mu)^2 + \sum_{r=2}^q (y_{r-1} - \mu)^2 - 2\rho \sum_{r=2}^q (y_r - \mu)(y_{r-1} - \mu) \right\} \\ &\quad - \frac{q-1}{\sigma^2}. \end{aligned}$$

The matrix  $H(\theta)$  has elements

$$H_{\mu\mu}(\theta) = \frac{2(q-1)(1-\rho)}{\sigma^2}, \quad H_{\rho\rho} = \frac{(q-1)(1+\rho^2)}{(1-\rho^2)^2},$$

$$H_{\sigma^2\sigma^2}(\theta) = \frac{q-1}{(\sigma^2)^2}, \quad H_{\mu\rho}(\theta) = H_{\mu\sigma^2}(\theta) = 0, \quad H_{\rho\sigma^2} = \frac{(q-1)\rho}{\sigma^2(1-\rho^2)},$$

while the matrix  $J(\theta)$  has

$$J_{\mu\mu}(\theta) = \frac{2(1-\rho)^2}{\sigma^2(1-\rho^2)} \sum_{r=2}^q \sum_{s=2}^q \left\{ \rho^{|r-s|} + \rho^{|r-s+1|} \right\}, \quad J_{\mu\rho}(\theta) = J_{\mu\sigma^2}(\theta) = 0,$$

$$J_{\rho\rho}(\theta) = \frac{1}{(1-\rho^2)^2} \sum_{r=2}^q \sum_{s=2}^q \left\{ \rho^{2|r-s|} + \rho^{|r-s+1|+|r-s-1|} \right\},$$

$$J_{\rho\sigma^2}(\theta) = \frac{1}{\sigma^2(1-\rho^2)^2} \sum_{r=2}^q \sum_{s=2}^q \left\{ \rho^{|r-s|+|r-s-1|} + \rho^{|r-s|+|r-s+1|} - \rho^{2|r-s|+1} - \rho^{|r-s-1|+|r-s+1|+1} \right\},$$

$$J_{\sigma^2\sigma^2}(\theta) = \frac{1}{(\sigma^2)^2(1-\rho^2)^2} \sum_{r=2}^q \sum_{s=2}^q \left\{ (1+\rho^2)\rho^{2|r-s|} + \rho^{|r-s+1|+|r-s-1|+2} + \rho^{2|r-s+1|} - 2\rho^{|r-s|+|r-s+1|+1} - 2\rho^{|r-s|+|r-s-1|+1} \right\}.$$

## References

- Chandler, R. E. and Bate, S. (2007). Inference for clustered data using the independence log-likelihood. *Biometrika* **94**, 167-183.
- Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91**, 729-737.
- Davison, A. C. (2003). *Statistical Models*. Cambridge Univ. Press, Cambridge.
- Foutz, R. V. and Srivastava, R. C. (1977). The performance of the likelihood ratio test when the model is incorrect. *Ann. Statist.* **5**, 1183-1194.
- Foutz, R. V. and Srivastava, R. C. (1978). The asymptotic distribution of the likelihood ratio when the model is incorrect. *Canad. J. Statist.* **6**, 273-279.
- Geys, H. and Molenberghs, G. (2002). Pseudo-likelihood inference. In *Topics in Modelling of Clustered Data* (Edited by M. Aerts et al.), 115-125. Chapman and Hall, London.
- Kent, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika* **69**, 19-27.
- Lindsay, B. G. (1988). Composite likelihood methods. *Contemp. Math.* **80**, 221-239.
- Lindsay, B. G., Pilla, R. S. and Basak, P. (2000). Moment-based approximations of distributions using mixtures: theory and applications. *Ann. Inst. Statist. Math.* **52**, 215-230.
- Mardia, K. V., Kent, J. T., Hughes, G. and Taylor, C. C. (2009). Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika* **96**, 975-982.

- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer, New York.
- Rotnitzky, A. and Jewell, N. P. (1990). Hypothesis testing of regression parameters in semi-parametric generalized linear models for cluster correlated data. *Biometrika* **77**, 485-497.
- Royall, R. and Tsou, T.-S. (2003). Interpreting statistical evidence by using imperfect models: robust adjusted likelihood functions. *J. Roy. Statist. Soc. Ser. B* **65**, 391-404.
- Satterthwaite, F. E. (1946). Approximate distribution of estimates of variance components. *Biometrics Bull.* **2**, 110-114.
- Stafford, J. E. (1996). A robust adjustment of the profile likelihood. *Ann. Statist.* **24**, 336-352.
- Varin, C. (2008). On composite marginal likelihoods. *Adv. Stat. Anal.* **95**, 1-28.
- Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21**, 5-42.
- Wood, A. T. A. (1989). An F approximation to the distribution of a linear combination of chi-squared variables. *Commun. Statist. Simula.* **18**, 1439-1456.

Department of Statistics, University of Udine, Via Treppo of 18, 33100 Udine, Italy.

E-mail: pace@dss.uniud.it

Department of Statistics, University of Padova, Via C. Battisti 241, 35121 Padova, Italy.

E-mail: alessandra.salvan@unipd.it

Department of Statistics, University of Padova, Via C. Battisti 241, 35121 Padova, Italy.

E-mail: nicola.sartori@unipd.it

(Received October 2009; accepted February 2010)