Inference from PseudoLikelihoods with Plug-In Estimates

DOI:10.1111/anzs.12121

INFERENCE FROM PSEUDO LIKELIHOODS WITH PLUG-IN ESTIMATES

Luigi Pace, Alessandra Salvan[1] and Nicola Sartori

University of Udine, University of Padova and University of Padova

Abstract

Effective implementation of likelihood inference in models for high-dimensional data often requires a simplified treatment of nuisance parameters, replaced by handy estimates. The likelihood function as well could be simplified by using a partial specification of the model, as in composite likelihood. Tests and confidence regions for the parameter of interest may then be constructed using Wald type and score type statistics, accounting for nuisance parameters estimation or partial specification of the likelihood. Here, a general analytical expression for the needed asymptotic covariance matrices is derived, together with suggestions for obtaining Monte Carlo approximations. The same matrices are involved in a rescaling adjustment of the log likelihood ratio type statistic we propose here. This adjustment recovers the usual chi-squared asymptotic distribution, generally failing after the simplifications considered. The practical implication is that, for a wide variety of likelihoods and nuisance parameter estimates, confidence regions for the parameter of interest are readily computable from the rescaled log likelihood ratio type statistic as well as from the Wald type and score type statistics. Two examples, a measurement error model with full likelihood and a spatial correlation model with pairwise likelihood, illustrate and compare the procedures. Wald type and score type statistics give rise to confidence regions that may have unsatisfactory shape in small and moderate samples. In addition to satisfactory shape, regions based on the rescaled log likelihood ratio type statistic show empirical coverage in reasonable agreement with nominal confidence levels.

Keywords: Composite likelihood; Estimating equation; Nuisance parameter; Pairwise likelihood; Profile likelihood.

1 Introduction

Consider a model for data y indexed by two sets of parameters: a parameter of interest θ and a nuisance parameter ϕ. Complex model structures may raise diffi-

[1]Corresponding author

culties in the elimination of nuisance parameters, both when using the full log likelihood $\ell(\theta, \phi)$ and when using a suitably simplified likelihood such as a composite likelihood (Lindsay, 1988; Varin et al., 2011). Profiling, i.e. replacing ϕ with $\hat{\phi}_\theta$, the maximiser of the full or simplified likelihood with respect to ϕ with θ fixed, may not be practically feasible or advisable, especially with high-dimensional parameters. Use of Gong & Samaniego (1981) suggestion, to plug-in a simple estimate $\tilde{\phi}$ of ϕ in the working log likelihood, is common in applications due to its computational convenience. The resulting function of the parameter of interest is referred to as a pseudo log likelihood and will be denoted by $\tilde{\ell}(\theta)$. Recent examples are in Guolo (2011), Ghosh et al. (2013), Wang et al. (2014) for the full likelihood, and in Pakel et al. (2011) and Varin et al. (2011, §§ 3 and 4) for composite likelihoods.

Tests and confidence regions for the parameter of interest may be constructed from $\tilde{\ell}(\theta)$ using Wald type and score type statistics, accounting for nuisance parameters estimation or partial specification of the likelihood. In Section 2.1 of this paper, a general analytical expression for the needed asymptotic covariance matrices is derived, together with widely feasible Monte Carlo approximations. The ensuing Wald-type and score-type regions, however, may suffer from theoretical and practical difficulties, such as lack of parameterization invariance, numerical instability (Molenberghs & Verbeke, 2005, § 9.3.2), leading to poor shape and unboundedness of the region, and possibily poor coverage accuracy in small samples (Hauck & Donner, 1977), as is shown by the examples in Section 3.

These inconveniences may be mitigated by using likelihood ratio type statistics. However, the latter have a nonstandard asymptotic distribution given by a weighted sum of independent chi-squared variables. See Kent (1982) for a general misspecified likelihood, Liang & Self (1996) and Chen & Liang (2010) for plug-in estimation in a full likelihood and Varin et al. (2011) for the composite likelihood. In principle, p-values can be computed from this distribution, while construction of confidence regions seems to be a daunting task because the weights in the sum of chi-squares depend on (θ, ϕ). As an alternative procedure, which is preferable for confidence regions, we suggest in Section 2.2 a rescaled pseudo likelihood ratio statistic, which recovers the asymptotic chi-squared distribution. Rescaling is based on the pseudo score, its asymptotic covariance and the pseudo likelihood Hessian.

The practical implication is that, for a wide variety of likelihoods and nuisance parameter estimates, confidence regions for the parameter of interest are readily computable from the rescaled log likelihood ratio type statistic as well as from the Wald type and score type statistics. The two examples below illustrate the simplification gained by using $\tilde{\ell}(\theta)$ as a basis for inference and will be continued in Section 3 with the construction of confidence regions.

1.1 Example 1: A measurement error model

As noted in Chen & Liang (2010, § 1.6), if a full log likelihood $\ell(\theta, \phi)$ decomposes as

$$\ell(\theta, \phi) = \ell_1(\phi) + \ell_2(\theta, \phi)$$

with both $\ell_1(\phi)$ and $\ell_2(\theta, \phi)$ genuine log likelihoods, a simplified estimate $\tilde{\phi}$ of ϕ is provided by the maximiser of $\ell_1(\phi)$. Inference about θ may then be based on the pseudo log likelihood

$$\tilde{\ell}(\theta) = \ell_2(\theta, \tilde{\phi}) ,$$

which is in general computationally more practical than the profile log likelihood $\ell_P(\theta) = \ell(\theta, \hat{\phi}_\theta)$, with $\hat{\phi}_\theta$ the maximizer of $\ell(\theta, \phi)$ with respect to ϕ with θ fixed. A major instance of the decomposition of $\ell(\theta, \phi)$ as $\ell_1(\phi) + \ell_2(\theta, \phi)$ is offered by measurement error models. They relate a response variable Y, via a parameter θ, to the true covariate X, which is observed indirectly through a surrogate W, whose distribution given X depends on an additional parameter ϕ.

As a simple illustration, let $(w_1, y_1), \ldots, (w_n, y_n)$ be independent observations from the bivariate random variable (W, Y), where $W = X + U$ and $Y = \beta_0 + \beta_1 X + \varepsilon$, with ε, X, U independent unobservable variables with $X \sim N(\mu_X, \sigma_X^2)$, $U \sim N(0, \sigma_U^2)$, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Above, $\mu_X, \beta_0, \beta_1 \in \mathrm{I\!R}$ and $\sigma_X^2, \sigma_U^2, \sigma_\varepsilon^2$ are positive. To allow later comparison with inference based on the profile log likelihood, we assume that $\sigma_X^2 = k^2 \sigma_U^2$ and $\sigma_\varepsilon^2 = h^2 \sigma_U^2$, with k^2 and h^2 known positive constants. Hence, data are a random sample of size n from a bivariate normal distribution (W, Y), with means $E(W) = \mu_X$, $E(Y) = \beta_0 + \beta_1 \mu_X$, variances $Var(W) = (1 + k^2)\sigma_U^2$, $Var(Y) = (\beta_1^2 k^2 + h^2)\sigma_U^2$ and covariance $\mathrm{Cov}(W, Y) = \beta_1 k^2 \sigma_U^2$.

Let us consider $\theta = (\beta_0, \beta_1)$ and $\phi = (\mu_X, \sigma_U^2)$. We have $\ell(\theta, \phi; w, y) = \ell_1(\phi; w) + \ell_2(\theta, \phi; w, y)$, where, neglecting constants,

$$\begin{aligned}
\ell_1(\phi; w) &= -\frac{n}{2} \log \sigma_U^2 - \frac{1}{2(1+k^2)\sigma_U^2} \sum_{i=1}^{n} (w_i - \mu_X)^2 , \\
\ell_2(\theta, \phi; w, y) &= -\frac{n}{2} \log v(\beta_1, \sigma_U^2) - \frac{1}{2\, v(\beta_1, \sigma_U^2)} \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 \frac{\mu_X + w_i k^2}{1 + k^2} \right)^2 ,
\end{aligned}$$

with $v(\beta_1, \sigma_U^2) = \left\{ h^2 + \beta_1^2 k^2/(1+k^2) \right\} \sigma_U^2$. Replacing ϕ with an estimate based on $\ell_1(\phi; w)$ is simpler than calculating $\ell_P(\theta)$. The maximiser of $\ell_1(\phi; w)$ is $\tilde{\phi} = (\tilde{\mu}_X, \tilde{\sigma}_U^2)$, where

$$\tilde{\mu}_X = \bar{w}_n = \frac{1}{n} \sum_{i=1}^{n} w_i , \quad \tilde{\sigma}_U^2 = \frac{1}{(1+k^2)n} \sum_{i=1i}^{n} (w_i - \bar{w}_n)^2 .$$

The resulting pseudo log likelihood for θ is, neglecting constants,

$$\begin{aligned}\tilde{\ell}(\theta) &= \ell_2(\theta, \tilde{\phi}; w, y) \\ &= -\frac{n}{2}\log v(\beta_1, \tilde{\sigma}_U^2) - \frac{1}{2\,v(\beta_1, \tilde{\sigma}_U^2)}\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1\frac{\tilde{\mu}_X + w_i k^2}{1+k^2}\right)^2 .\end{aligned}$$

We will be interested in the construction of confidence regions for θ based on $\tilde{\ell}(\theta)$.

1.2 Example 2: Data from spatial Gaussian random fields

Gaussian random fields play a central role in the construction of models for the analysis of geostatistical data (Cressie, 1993). For these models, exact computation of the log likelihood becomes unfeasible as the number of sample points gets large. Composite likelihood methods (Lindsay, 1988; Varin et al., 2011) offer one appealing computational simplification, while preserving part of the properties of the full likelihood. For an up to date review of composite likelihood for spatial Gaussian random fields, we refer to Bevilacqua & Gaetan (2014).

Consider a vector $y = (y_1, \ldots, y_m)^\top$ of measurements of the phenomenon of interest at m monitoring stations, observed together with a vector of k explanatory variables at each station. We model y as a realization of Y, having an m-variate normal distribution with mean $\mu = X\beta$, where X is a full rank $m \times k$ fixed matrix, $\beta \in \mathbb{R}^k$, and covariance $\sigma^2 R(\theta)$, where $\sigma^2 > 0$ is the common variance of the components of Y and $R(\theta)$ has generic element $\rho_{rs} = \rho_{rs}(\theta)$. The parameters of interest in models for data from spatial random fields are typically those in the autocorrelation matrix. With $\phi = (\beta, \sigma^2)$, the full log likelihood for (θ, ϕ) is

$$\ell(\theta, \phi) = -\frac{m}{2}\log\sigma^2 - \frac{1}{2}\log|R(\theta)| - \frac{1}{2\sigma^2}(y - X\beta)^\top[R(\theta)]^{-1}(y - X\beta)\,,$$

where evaluation of $|R(\theta)|$ and $[R(\theta)]^{-1}$ becomes demanding as m gets large, since their computational cost is, with the most widely used algorithms, of order $O(m^3)$. A simple composite log likelihood is the pairwise log likelihood (see for instance Bevilacqua & Gaetan, 2014, § 3), which only requires calculation of the Gaussian log likelihood for pairs (y_r, y_s), $r, s = 1, \ldots, m$, $r \neq s$. Denoting by w_{rs} some known weights, the pairwise log likelihood is

$$p\ell(\theta, \phi) = \sum_{\substack{r,s=1 \\ r \neq s}}^{m} w_{rs}\log p_{Y_r Y_s}(y_r, y_s; \theta, \phi)\,. \tag{1}$$

The computational cost of $p\ell(\theta, \phi)$ is at most of order $O(m^2)$, and it can be smaller if many weights are null.

We consider a specification of $R(\theta)$ of exponential form (Diggle & Ribeiro, 2007, § 3.4),

$$\rho_{rs} = \exp(-d_{rs}/\theta)\,,$$

where d_{rs} is a distance between spatial locations giving rise to Y_r and Y_s and $\theta > 0$. The quantities below are, however, general for normal models with parameters of interest in a structured correlation matrix $R(\theta)$. We will use (1) with weights $w_{rs} = 1$ if d_{rs} is smaller than a fixed threshold d_0 and zero otherwise. The choice of d_0, and more in general of the weights w_{rs}, may affect the efficiency of inference based on the pairwise likelihood. This issue is beyond the scope of the present paper and will not be discussed here. A recent account is given by Bevilacqua & Gaetan (2014).

The generic pair (Y_r, Y_s), $r \neq s$, has a bivariate normal distribution with mean (μ_r, μ_s), common variance σ^2, and correlation ρ_{rs}. Therefore, (1) is

$$p\ell(\theta,\phi) = \sum_{r=1}^{m-1}\sum_{s=r+1}^{m} w_{rs}\left\{-\log\sigma^2 - \frac{1}{2}\log(1-\rho_{rs}^2) - \frac{A_{rs}}{2\sigma^2(1-\rho_{rs}^2)}\right\}\,,$$

with $A_{rs} = (y_r-\mu_r)^2 + (y_s-\mu_s)^2 - 2\rho_{rs}(y_r-\mu_r)(y_s-\mu_s)\,.$

A further simplification is obtained by replacing ϕ in $p\ell(\theta,\phi)$ with the usual least squares estimates under independence $\tilde\phi = (\tilde\beta, \tilde\sigma^2)$, where

$$\tilde\beta = (X^\top X)^{-1}X^\top y\,, \qquad \tilde\sigma^2 = \frac{1}{m-k}(y - X\tilde\beta)^\top(y - X\tilde\beta).$$

This gives the pseudo pairwise log likelihood $\tilde\ell(\theta) = p\ell(\theta,\tilde\phi)$ from which confidence regions for θ are to be constructed.

2 Approximate pivots from pseudo likelihoods

Consider inference about a p-dimensional parameter of interest θ, in the presence of a q-dimensional nuisance parameter ϕ, based on data $y = (y_1,\ldots,y_n)$. Let $y_1,\ldots,y_n$ be observations of m-dimensional independent random variables $Y_1,\ldots,Y_n$. More generally, we assume that the information about (θ,ϕ) is of order $O(n)$. Let $\ell(\theta,\phi)$ be a full or composite log likelihood for (θ,ϕ).

We consider for inference about θ the pseudo log likelihood

$$\tilde\ell(\theta) = \ell(\theta,\tilde\phi)\,.$$

The estimate $\tilde\phi$ is a simple estimate. It is assumed to be the solution of an estimating equation $g(\phi;y) = \sum_{i=1}^n g_i(\phi;y_i) = 0$, such that $E_{\theta,\phi}\left(g_i(\phi;Y_i)\right) = 0$, for $i =$

$1, \dots, n$, and for every θ and ϕ, or, more generally, $E_{\theta,\phi}\left(g(\phi; Y)\right) = O(1)$, with $Y = (Y_1, \dots, Y_n)$.

The following notation is used in the rest of the paper. Let $\tilde{\theta}$ be the maximizer of $\tilde{\ell}(\theta)$. We assume that $\tilde{\theta}$ is the unique solution of $\tilde{U}(\theta) = 0$, where $\tilde{U}(\theta) = (\partial/\partial\theta)\tilde{\ell}(\theta)$ is the pseudo score. We have that $\tilde{U}(\theta) = U_\theta(\theta, \tilde{\phi})$, with $U_\theta(\theta,\phi) = (\partial/\partial\theta)\ell(\theta,\phi)$.

2.1 Pseudo Wald and score type statistics from $\tilde{\ell}(\theta)$

Using standard arguments, we may easily recover a general analytical expression for the asymptotic covariance matrices of $\tilde{U}(\theta)$ and of $\tilde{\theta}$. Hereafter, the symbol $\dot{\sim}$ is a shorthand for 'is approximately distributed as', with the approximation having error of order $O_p(n^{-1/2})$. Under regularity conditions stated for instance in Molenberghs & Verbeke (2005, Section 9.2), we have

$$\left(U_\theta(\theta,\phi)^\intercal/\sqrt{n},\, \sqrt{n}(\tilde{\phi}-\phi)^\intercal\right)^\intercal \dot{\sim} N_{p+q}(0, V)\,,$$

where

$$V = \begin{pmatrix} J_{\theta\theta}/n & \Omega_{\theta\phi}(Q^{-1})^\top \\ Q^{-1}\Omega_{\theta\phi}^\top & n\Sigma \end{pmatrix}\,,$$

with

$$\begin{aligned} J_{\theta\theta} &= E_{\theta,\phi}\{U_\theta(\theta,\phi)U_\theta(\theta,\phi)^\top\}\,, \\ \Omega_{\theta\phi} &= Cov_{\theta,\phi}(U_\theta(\theta,\phi), g(\phi;Y)) = E_{\theta,\phi}(U_\theta(\theta,\phi)\; g(\phi;Y)^\top)\,, \\ Q &= E_{\theta,\phi}\left(-\partial/\partial\phi^\top g(\phi;Y)\right)\,, \quad \Sigma = Q^{-1}S(Q^{-1})^\top\,. \end{aligned}$$

Above, $S = Var_{\theta,\phi}\left(g(\phi;Y)\right) = E_{\theta,\phi}\left(g(\phi;Y)g(\phi;Y)^\top\right)$, where the last equality holds with error $O(1)$ if $E_{\theta,\phi}\left(g(\phi;Y)\right) = O(1)$. From the expansion

$$U_\theta(\theta,\tilde{\phi}) = U_\theta(\theta,\phi) - H_{\theta\phi}(\tilde{\phi}-\phi) + O_p(1)\,,$$

with $H_{\theta\phi} = E_{\theta,\phi}\{-\partial/\partial\phi^\top U_\theta(\theta,\phi)\}$, we get

$$\tilde{U}(\theta) \,\dot{\sim}\, N_p(0, K)\,,$$

where

$$K = J_{\theta\theta} + H_{\theta\phi}\Sigma H_{\theta\phi}^\top - \Omega_{\theta\phi}(Q^{-1})^\top H_{\theta\phi}^\top - H_{\theta\phi}Q^{-1}\Omega_{\theta\phi}^\top\,. \tag{2}$$

Formula (2) accounts both for the use of a simplified likelihood in place of the full likelihood and for plug-in estimation of nuisance parameters. It unifies various special cases in the literature ensuing from Gong & Samaniego (1981).

Remark 1. Formula (2) reduces to $K = J_{\theta\theta} + H_{\theta\phi}\Sigma(H_{\theta\phi})^\top$ when $\Omega_{\theta\phi} = 0$. This happens in particular if $\tilde{\phi}$ is evaluated on independent data, as when using past studies.

Remark 2. When $\ell(\theta, \phi)$ is a full likelihood, $J_{\theta\theta} = i_{\theta\theta}$ and $H_{\theta\phi} = i_{\theta\phi}$, where $i_{\theta\theta}$ and $i_{\theta\phi}$ are blocks of the Fisher information matrix. Moreover, $\Omega_{\theta\phi} = 0$ as a consequence of the efficiency of $U_\theta(\theta, \phi)$ as an estimating function for θ at the true ϕ (Pierce, 1982; Chen & Liang, 2010, Theorem 1). Hence, $K = i_{\theta\theta} + i_{\theta\phi}\Sigma i_{\theta\phi}^\top$.

Remark 3. When $U_\theta(\theta, \phi)$ is nuisance parameter-insensitive, that is when $H_{\theta\phi} = 0$ (Jørgensen & Knudsen, 2004), we get $K = J_{\theta\theta}$.

Let now

$$H_{\theta\theta} = E_{\theta,\phi}(-\partial/\partial\theta^\top U_\theta(\theta, \phi)) = E_{\theta,\phi}(-\partial/\partial\theta^\top \tilde{U}(\theta)) + o(n)\,.$$

From $\tilde{\theta} - \theta = H_{\theta\theta}^{-1}\tilde{U}(\theta) + o_p(n^{-1/2})$, we get

$$\tilde{\theta} - \theta \,\dot{\sim}\, N_p(0, H_{\theta\theta}^{-1} K H_{\theta\theta}^{-1})\,. \tag{3}$$

It follows that the Wald-type statistic

$$\tilde{w}^e(\theta) = (\tilde{\theta} - \theta)^\top H_{\theta\theta} K^{-1} H_{\theta\theta}(\tilde{\theta} - \theta)\,, \tag{4}$$

and the score-type statistic

$$\tilde{w}^u(\theta) = \tilde{U}(\theta)^\top K^{-1} \tilde{U}(\theta)\,,$$

have the usual asymptotic χ^2_p distribution at the true θ. When $\ell(\theta, \phi)$ is a full log likelihood $H_{\theta\theta} = i_{\theta\theta}$.

Matrices K and $H_{\theta\theta}$ will in general depend on θ and ϕ. The latter will be replaced by $\tilde{\phi}$. In Wald-type statistics, it is customary to evaluate K and $H_{\theta\theta}$ at $(\tilde{\theta}, \tilde{\phi})$. The quantity $H_{\theta\theta}$ depends on the curvature of $\tilde{\ell}(\theta)$ and, in general, is easy to calculate, or to estimate. Approximation of K by means of Monte Carlo simulation of the variance of $\tilde{U}(\theta)$ is feasible in wide generality, adapting to $\tilde{\ell}(\theta)$ the ideas proposed in Varin et al. (2011, § 5.1) and in Cattelan & Sartori (2014) for composite likelihoods. Use of this Monte Carlo approximation is illustrated in the examples of Section 3. Cattelan & Sartori (2014, § 4.1) perform a simulation study for values of M ranging from 250 to 1000, confirming that a moderate number of replications such as $M = 250$ is enough for satisfactory accuracy of empirical coverage probabilities. Moreover, for the pairwise likelihood of Section 3.2, the analytical version (2) requires the quantity $J_{\theta\theta}$, which has a computational cost of order $O(m^4)$. This is higher than the order of the computation cost of the likelihood. Therefore, the simulated version, which has computation cost of order $O(Mm^2)$ could be the only viable solution when m is large. Some details are given in Section 3.2.

2.2 Rescaled pseudo likelihood ratio

Inference based on $\tilde{w}^e(\theta)$ and $\tilde{w}^u(\theta)$ may suffer from some drawbacks, such as lack of invariance under reparameterizations, numerical instability, unboundedness and poor shape of confidence regions. The pseudo likelihood ratio statistic

$$\tilde{w}(\theta) = 2\left\{\tilde{\ell}(\tilde{\theta}) - \tilde{\ell}(\theta)\right\}$$

could be a more appealing basis for inference, but it is more difficult to calibrate, because, at the true value of (θ, ϕ),

$$\tilde{w}(\theta) \dot{\sim} \sum_{j=1}^{p} \nu_j Z_j^2\,,$$

where $Z_1, \ldots, Z_p$ are i.i.d. $N(0,1)$ and the $\nu_1, \ldots, \nu_p$ are the eigenvalues of $KH_{\theta\theta}^{-1}$ (cf. Theorem 8.5 in Severini, 2005), which in general depend on (θ, ϕ).

The asymptotic null distribution of $\tilde{w}(\theta)$ is chi-square on p degrees of freedom if and only if $KH_{\theta\theta}^{-1} = I_p + o(1)$, that is if and only if the pseudo log likelihood $\tilde{\ell}(\theta)$ satisfies the information identity $Var_{\theta,\phi}(\tilde{U}(\theta)) = E_{\theta,\phi}\{-(\partial/\partial\theta^\top)\tilde{U}(\theta)\}$ with error $o(n)$. Such identity generally fails for the profiled composite log likelihood, as illustrated in Pace et al. (2011). It typically fails even for a pseudo log likelihood $\tilde{\ell}(\theta)$ based on a genuine likelihood. When this is the case, quantiles of the approximate distribution of $\tilde{w}(\theta)$ depend on θ and contruction of confidence regions based on inversion of acceptance regions is far from being straightforward.

In order to recover a χ_p^2 asymptotic distribution for the pseudo likelihood ratio, we propose an adjustment of $\tilde{w}(\theta)$. The adjustment applies to multidimensional parameters of interest and extends to a general framework the proposal of Pace et al. (2011) referred to composite likelihoods and elimination of nuisance parameters by profiling. The rescaled likelihood ratio statistic has the form

$$\tilde{w}^*(\theta) = \frac{\tilde{U}(\theta)^\top K^{-1}\tilde{U}(\theta)}{\tilde{U}(\theta)^\top H_{\theta\theta}^{-1}\tilde{U}(\theta)}\,\tilde{w}(\theta)\,. \tag{5}$$

The asymptotic χ^2_p distribution of $\tilde{w}^*(\theta)$ follows from the expansion $\tilde{w}(\theta) = \tilde{U}(\theta)^\top H_{\theta\theta}^{-1}\tilde{U}(\theta) + o_p(1)$, so that $\tilde{w}^*(\theta) = \tilde{U}(\theta)^\top K^{-1}\tilde{U}(\theta) + o_p(1) = \tilde{w}^u(\theta) + o_p(1)$ and the three forms $\tilde{w}^*(\theta)$, $\tilde{w}^e(\theta)$ and $\tilde{w}^u(\theta)$ are asymptotically equivalent. With scalar θ, (5) simplifies to $\tilde{w}^*(\theta) = (H_{\theta\theta}/K)\,\tilde{w}(\theta)$. This agrees with several proposals in the literature, see e.g. Kent (1982), Stafford (1996), Geys et al. (1999).

Confidence regions with asymptotic level $1-\alpha$ based on $\tilde{w}^*(\theta)$ have the form $\{\theta : \tilde{w}^*(\theta) \le \chi^2_{p,1-\alpha}\}$. On the other hand, regions of the form $\{\theta : \tilde{w}(\theta) \le \chi^2_{p,1-\alpha}\}$, ignoring the cost of estimating nuisance parameters, are typically undercovering, being wrongly calibrated. Moreover, in the examples of Section 3, regions based on $\tilde{w}^*(\theta)$ appear to be close in shape and coverage to ideal regions based on profiling the full likelihood, of the form $\{\theta : w_{_P}(\theta) \le \chi^2_{p,1-\alpha}\}$, where $w_{_P}(\theta)$ is the profile likelihood ratio from the full likelihood. Regions based on the score type statistic $\tilde{w}^u(\theta)$, i.e. of the form $\{\theta : \tilde{w}^u(\theta) \le \chi^2_{p,1-\alpha}\}$ have asymptotic level $1-\alpha$, but may be unbounded in small samples. In turn, regions based on the Wald type statistic $\{\theta : \tilde{w}^e(\theta) \le \chi^2_{p,1-\alpha}\}$ might have asymptotic correct coverage, but have forced ellyptical shapes and lack parameterization invariance.

An appealing feature of $\tilde{w}^*(\theta)$ is that inference about a p_0-dimensional component ψ of $\theta = (\psi, \lambda)$ can be based on

$$\tilde{w}^*_{_P}(\psi) = \min_{\lambda} \tilde{w}^*(\psi, \lambda)\,, \tag{6}$$

with a limiting $\chi^2_{p_0}$ null distribution. The latter result can be easily shown using standard likelihood asymptotics from the asymptotic equivalence between $\tilde{w}^*(\theta)$ and $\tilde{w}^e(\theta)$ given by (4). So, in the end, some nuisance parameters may be dealt with through separate estimation plus adjusting, while other ones afterward through profiling.

3 Numerical illustration

We provide here numerical evidence of the accuracy of the methods in Section 2 applied to the examples introduced in Sections 1.1 and 1.2.

3.1 Measurement error model

For the measurement error model of Section 1.1, the covariance matrix of $\tilde{\phi}$ is $\Sigma = n^{-1}\mathrm{diag}\left((1+k^2)\sigma^2_{_U}, 2\sigma^4_{_U}\right)$. Adjustment (5) has $K = i_{\theta\theta} + i_{\theta\phi}\Sigma i_{\theta\phi}^\top$, with elements of $i_{\theta\theta}$ and $i_{\theta\phi}$ given in the Appendix, and $H_{\theta\theta} = i_{\theta\theta}$.

As an illustration, we consider a simulated dataset with $n = 20$, $\theta = (1, 2)$, $\mu_{_X} = 2$, $\sigma^2_{_U} = 1.1$ and $h = k = 1$. Confidence regions for θ with nominal level 0.95 are displayed in Figure 1. The region based on $\tilde{w}(\theta)$, wrongly calibrated on a χ^2_2 scale, is too narrow. The score-type confidence region has an unusual shape and departs remarkably from that based on the profile log likelihood ratio statistic, $w_P(\theta)$, considered as the gold standard. At least in the present parameterization, the Wald-type region is closer to the target. The closest agreement in shape and location with the region based on $w_P(\theta)$ is obtained using $\tilde{w}^*(\theta)$. However, the region based on $\tilde{w}^*(\theta)$ is somehow larger than the region based on $w_P(\theta)$, as is to be expected.

A small simulation study was performed in order to compare the coverage of the confidence regions for θ based on the different likelihood ratio statistics. Estimated coverage probabilities are summarized in Table 1. The proposed adjustments show a reasonable performance, substantially improving on the unadjusted pseudo likelihood ratio statistic. Table 1 also includes results for the versions of $\tilde{w}^u(\theta)$ and $\tilde{w}^*(\theta)$ with a simulated estimate of K, based on 500 replications, which are denoted respectively by $\tilde{w}^u_{sim}(\theta)$ and $\tilde{w}^*_{sim}(\theta)$. These versions for small to moderate sample sizes tend to give regions with larger empirical coverages compared with those obtained from $\tilde{w}^u(\theta)$ and $\tilde{w}^*(\theta)$. The confidence region based on $\tilde{w}^u(\theta)$ has an unpleasant shape, see the bottom left panel of Figure 1, which is not improved by $\tilde{w}^u_{sim}(\theta)$, not shown in the figure.

Table 1: Measurement error model. Empirical coverage (%) of confidence regions of nominal level 90%, 95%, 99% for θ based on different statistics in three simulations, each with $10,000$ replications, $\theta = (1, 2)$, $\mu_X = 2$, $\sigma^2_U = 1.1$, $h = k = 1$ and $n = 10, 20, 100$.

	$n = 10$			$n = 20$			$n = 100$		
nominal	90	95	99	90	95	99	90	95	99
$w_P(\theta)$	85·3	91·6	97·9	88·0	93·7	98·5	89·9	95·0	98·9
$\tilde{w}(\theta)$	69·1	76·2	86·3	72·5	80·7	90·6	75·4	83·8	93·5
$\tilde{w}^u(\theta)$	78·7	82·5	87·6	83·9	88·1	93·3	89·0	93·6	97·8
$\tilde{w}^u_{sim}(\theta)$	93·5	95·2	97·3	91·2	94·0	97·3	90·9	94·8	98·3
$\tilde{w}^e(\theta)$	80·7	86·7	93·5	86·0	91·3	96·6	89·3	94·5	98·7
$\tilde{w}^*(\theta)$	81·7	87·3	93·7	86·5	91·6	96·7	89·5	94·5	98·6
$\tilde{w}^*_{sim}(\theta)$	97·9	98·8	99·6	94·8	97·2	99·3	91·6	96·0	99·1

Figure 2 shows confidence regions from $w_P(\theta)$ and $\tilde{w}^*(\theta)$ of nominal level 0.95 for $\theta = (\beta_0, \beta_1)$, together with confidence intervals for the two components, β_0 and β_1, obtained by profiling $\tilde{w}^*(\theta)$, as in formula (6). The confidence intervals from the lower-right panel in Figure 2 are $(-0.672, 3.044)$ ($\tilde{w}^*_P(\beta_0)$) and $(-0.498, 2.879)$ ($w_P(\beta_0)$). For comparison, the Wald-type confidence interval is

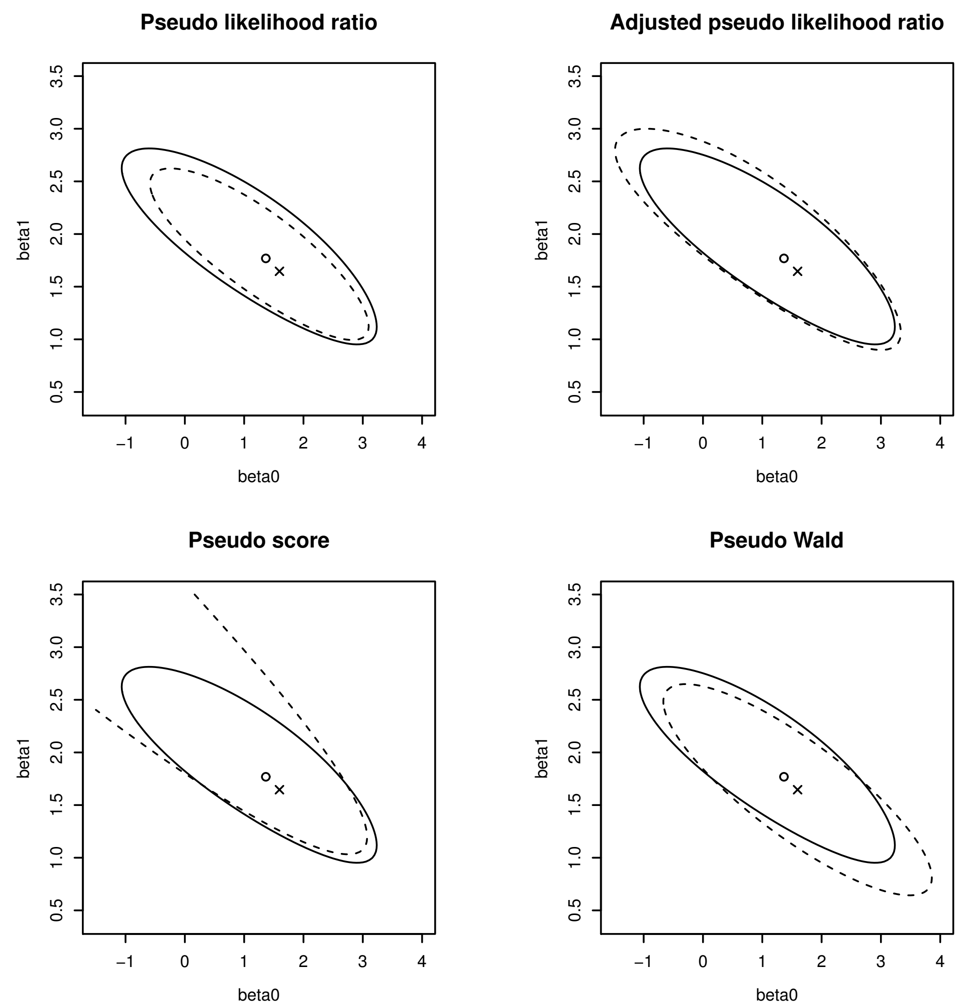

Figure 1: Measurement error model. Confidence regions for regression parameters for simulated data of Section 3.1. In all panels, the solid lines represent regions with nominal level 0.95 based on $w_{P}(\theta)$, while the circle and the cross represent $\hat{\theta}$ and $\tilde{\theta}$, respectively. Moving clockwise from the top left panel, the dashed lines represent the confidence region based on $\tilde{w}(\theta)$, $\tilde{w}^{*}(\theta)$, $\tilde{w}^{e}(\theta)$ and $\tilde{w}^{u}(\theta)$.

$(-0.218, 3.408)$. The corresponding intervals for β_1, from the upper-left panel are, respectively, $(1.022, 2.643)$, $(1.111, 2.569)$ and $(0.842, 2.449)$. In both cases, intervals based on the profile adjusted pseudo likelihood ratio are closer to full profile likelihood intervals than Wald-type ones.

3.2 Spatial Gaussian random fields

As an illustration of the example in Section 1.2, we consider the Wolfcamp acquifer data (Cressie, 1993, pp. 212-214) consisting of $m = 85$ irregularly spatially located measurements of piezometric head, available in the `R` package `geoR`. The strong trend in the northeast-southwest direction is removed by linear regression on spatial coordinates, so that $\phi = (\beta_0, \beta_1, \beta_2, \sigma^2)$. The pairwise log likelihood for (θ, ϕ) has been calculated with threshold $d_0 = 100$ kilometers for the weights. Full maximum likelihood estimates are $(\hat{\theta}, \hat{\phi}) = (18.93, 616.45, -1.29, -1.24, 4344.00)$, while $(\tilde{\theta}, \tilde{\phi}) = (21.56, 607.77, -1.28, -1.14, 3880.49)$. The pseudo pairwise log likelihood ratio $\tilde{w}(\theta)$ and adjusted versions $\tilde{w}^*(\theta)$ and $\tilde{w}^*_{sim}(\theta)$ are compared with the profile log likelihood ratio in Figure 3. The needed quantities for all statistics are given in the Appendix for the more general case of a random sample of size n from the model in Section 1.2. The confidence interval with nominal level 0.95 from the profile log likelihood is $(9.77, 37.17)$. The analogous intervals are $(15.36, 27.76)$ from $\tilde{w}(\theta)$, $(10.23, 74.15)$ from $\tilde{w}^*(\theta)$ and $(11.15, 52.75)$ from $\tilde{w}^*_{sim}(\theta)$. The figure also displays score-type and Wald-type statistics $\tilde{w}^u(\theta)$ and $\tilde{w}^e(\theta)$ as well as the version with simulated K, $\tilde{w}^u_{sim}(\theta)$ (K simulated with 200 replications). As in the previous example, the score-type statistic has an irregular shape that reflects on the costruction of confidence intervals. For this reason, these are not reported, while the Wald confidence interval is $(4.98, 38.27)$. Qualitatively, the behaviour of the profile likelihood ratio is best recovered by the adjusted versions of the pseudo pairwise likelihood ratio. This is also supported by empirical coverages in the simulation studies discussed below.

An alternative solution could be to use confidence intervals based on the bootstrap distribution of $\tilde{\theta}$. This can be obtained using samples simulated from the observed $(\tilde{\theta}, \tilde{\phi})$. With Wolfcamp data, the 0.95 percentile bootstrap confidence interval with 1000 bootstrap samples is $(7.68, 30.11)$. This is quite different from the profile gold standard and empirical coverage, estimated by simulation in the setting described below with $n = 1$, is 0.494, which is rather poor. Although for this particular data set the computational cost is fairly reasonable, this might not be the case for more complex settings since it requires optimization of the pseudo log likelihood for each bootstrap sample. Also a nonparametric bootstrap could be considered. However, in the present model, like with time series models, the resampling method needs to account for the data dependence structure, as in window

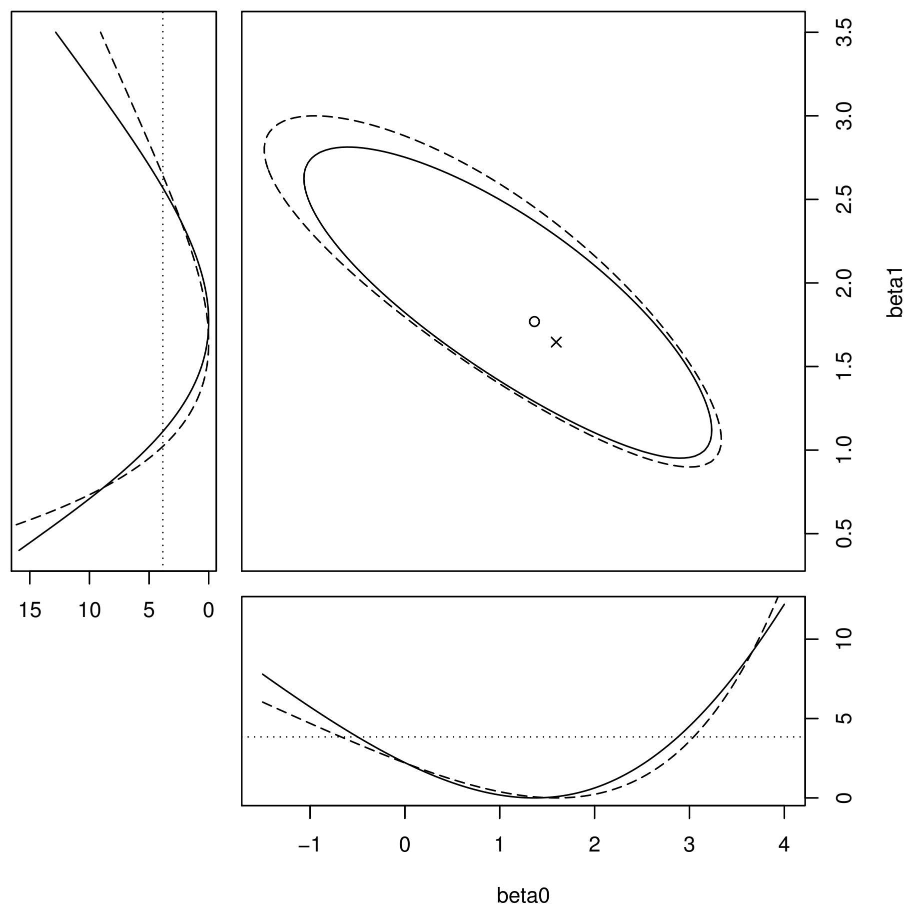

Figure 2: Measurement error model. Confidence regions for $\theta = (\beta_0, \beta_1)$, jointly and separately, for the simulated data of Section 3.1. In all panels, the solid lines represent regions with nominal level 0.95 based on the profile likelihood, while the dashed lines represent the same based on $\tilde{w}^*(\theta)$ (upper-right panel) and its profiling for β_1 (upper-left panel) and for β_0 (lower-right panel). The circle and the cross represent $\hat{\theta}$ and $\tilde{\theta}$, respectively.

subsampling schemes.

Coverage probabilities of the various confidence intervals have been checked through a simulation study for $n = 1, 5, 30$ and parameter values equal to the maximum likelihood estimates from the Wolfcamp data. Due to poor behaviour with $n = 1$ and prohibitive computational cost with $n = 5, 30$, bootstrap confidence intervals have not been considered. The statistics $\tilde{w}^*_{sim}(\theta)$ and $\tilde{w}^u_{sim}(\theta)$ have been computed with $M = 200$. The results are reported in Table 2. In the setting with $n = 1$ coverage probabilities for $\tilde{w}^*(\theta)$ are the closest to the nominal levels, even substantially improving on the profile. With additional independent observations at each spatial location ($n = 5$), coverage properties for $w_P(\theta)$ improve and $\tilde{w}^*(\theta)$ and $\tilde{w}^*_{sim}(\theta)$ are still competitive. When $n = 30$ nominal levels are attained by all statistics except of course by $\tilde{w}(\theta)$. Simulation results, not reported here, show stability of the results for $\tilde{w}^*_{sim}(\theta)$ and $\tilde{w}^u_{sim}(\theta)$ computed with $M = 500$, in particular for $n = 5, 30$.

In the above setting, with only $m = 85$ spatial locations, the differences in computational costs of $\tilde{w}^*(\theta)$, $\tilde{w}^*_{sim}(\theta)$ and $w_P(\theta)$ are barely noticeable. However, as m increases, such differences become more remarkable. As an example, we generated a sample of size $m = 500$ by resampling and perturbing the original spatial locations and using $(\hat{\theta}, \hat{\phi})$ as parameter value. As noted at the end of Section 2.1, in this setting $\tilde{w}^*_{sim}(\theta)$ is the most convenient choice, while the analytical version $\tilde{w}^*(\theta)$ is computationally more demanding even than $w_P(\theta)$. Indeed, the computation of $\tilde{w}^*_{sim}(\theta)$ took about 8 seconds, while those of $\tilde{w}^*(\theta)$ and $w_P(\theta)$ took about 940 and 110 seconds, respectively. Figure 4 shows $\tilde{w}(\theta)$ and $\tilde{w}^*_{sim}(\theta)$ for this dataset.

Finally, we mention an alternative formulation of inference for these models, that considers as nuisance parameter $\phi = \beta$, while the variance σ^2 and the correlation parameter are included in the parameter of interest θ. This choice would have the advantage of nuisance parameter-insensitivity, giving the simplification $K = J_{\theta\theta}$. However, numerical evidence (not reported here) indicates a poor behaviour of both adjusted pairwise and pseudo pairwise likelihoods. This seems to suggest that, as a general rule, it is convenient to treat as nuisance parameters, to be estimated by simple estimates, all parameters identifiable in univariate marginals.

4 Discussion

Simplified treatment of nuisance parameters with plug-in estimation needs to be supplemented with computationally affordable and reliable methods for constructing confidence regions. The rescaling adjustment for nuisance parameter estimation proposed here allows ready calculation of the usual likelihood-based regions

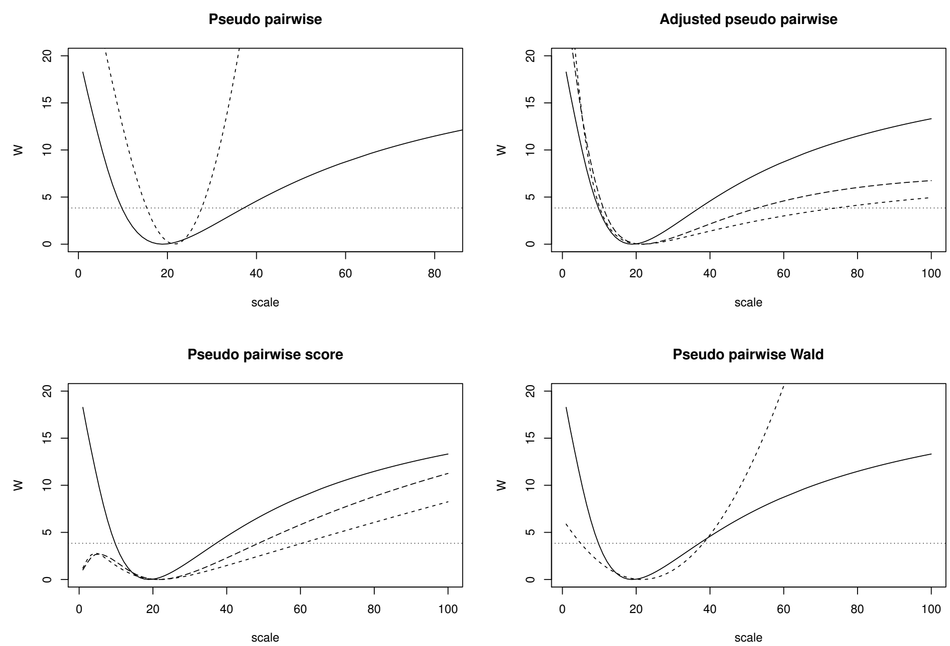

Figure 3: Wolfcamp data. In each panel, the statistic in the title is the dashed line, the solid line is $w_P(\theta)$, while the longdashed line is the statistic in the title with simulated K. The horizontal dotted line gives the approximate confidence interval of nominal level 95%.

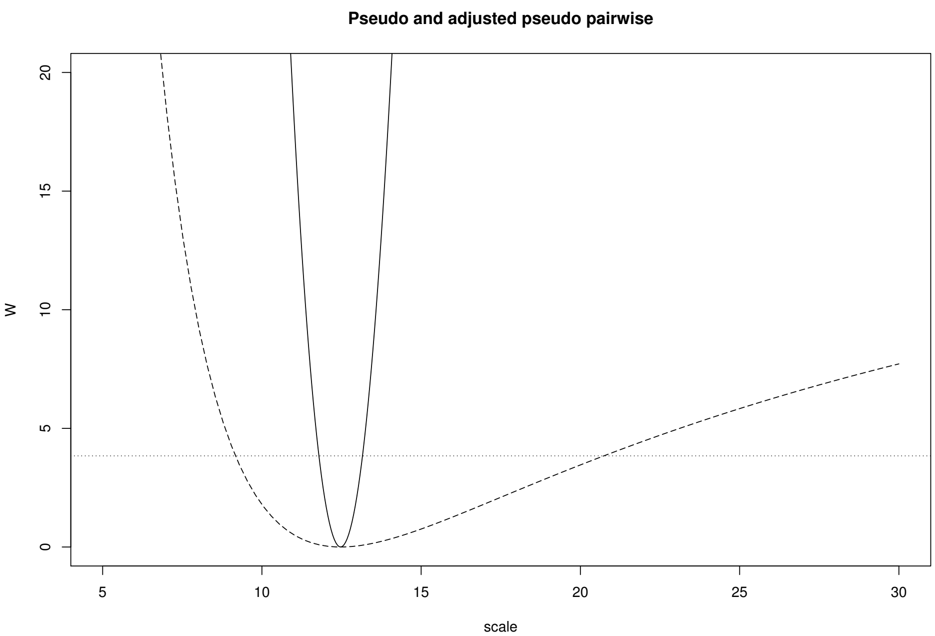

Figure 4: Simulated data with $m = 500$. The solid line is $\tilde{w}(\theta)$, while the dashed line is $\tilde{w}^*_{sim}(\theta)$. The horizontal dotted line gives the approximate confidence interval of nominal level 95%.

Table 2: Empirical coverage (%) of confidence regions of nominal level 90%, 95%, 99% based on different statistics in three simulations, $n = 1, 5, 30$, with $10,000$ replications from an exponential correlation model with covariates as in the Wolfcamp data and (θ, ϕ) equal to the maximum likelihood estimate from the dataset.

	$n=1$			$n=5$			$n=30$		
nominal	90	95	99	90	95	99	90	95	99
$w_P(\theta)$	80·0	86·0	91·7	88·8	94·4	98·7	90·0	95·0	99·0
$\tilde{w}(\theta)$	32·2	39·1	53·4	41·4	48·5	61·3	44·1	51·4	63·5
$\tilde{w}^u(\theta)$	85·3	93·2	99·2	88·1	94·1	98·9	90·2	95·1	98·9
$\tilde{w}^u_{sim}(\theta)$	73·8	84·7	96·0	90·2	95·5	99·3	90·1	95·0	98·9
$\tilde{w}^e(\theta)$	66·3	74·4	85·9	84·0	89·7	96·0	89·6	94·3	98·6
$\tilde{w}^*(\theta)$	90·5	97·1	99·9	89·0	95·0	99·3	90·4	95·2	98·9
$\tilde{w}^*_{sim}(\theta)$	79·3	90·0	98·7	91·1	96·4	99·6	90·3	95·2	98·9

with acceptable computational cost for a variety of models and likelihoods. Alternative solutions based on bootstrap methods, although viable in wide generality, seem less appealing in this context. On one side, complex models are typically related to complex dependencies, that need to be accounted for in the resampling scheme of a nonparametric bootstrap. On the other hand, parametric boostrap of the estimator with no prepivoting may lead to poor results as indicated in Section 3.2. Another possibility could be to use parametric bootstrap of $\tilde{w}(\theta)$ for obtaining a confidence region for θ. However, this solution turns out to be computationally very demanding: at each θ value, at least 1000 bootstrap samples are needed for reasonable accuracy, with optimization of the pseudo log likelihood required for each bootstrap sample. A numerical example with the Wolfcamp data (not reported here) shows that the bootstrap version of $\tilde{w}(\theta)$, that is the statistic obtained by inverting bootstrap p-values, obtained with 1000 simulated samples from $(\theta, \tilde{\phi})$, has a rather poor behaviour for large θ leading to an unbounded interval. Indeed, there is no guarantee to obtain accurate results since $\tilde{w}(\theta)$ is not a pivotal quantity (see, for instance, Young & Smith, 2005, § 11.2).

Acknowledgments

The authors are very grateful to the Associate Editor and to the Reviewers for valuable and constructive comments on earlier versions of this paper. This work was supported in part by grant no. CDPA131553 from the University of Padova.

Appendix

Measurement error model

For the model in Sections 1.1 and 3.1, with $\theta = (\beta_1, \beta_2)$ and $\phi = (\mu_X, \sigma_U^2)$, the adjustment in equation (5) has $K = i_{\theta\theta} + i_{\theta\phi}\Sigma i_{\theta\phi}^\top$, where $i_{\theta\theta}$ and $i_{\theta\phi}$ have elements

$$
\begin{aligned}
i_{\mu_X\beta_0} &= \frac{n\beta_1}{\sigma_U^2\left\{h^2(1+k^2)+\beta_1^2k^2\right\}}\,, \quad i_{\mu_X\beta_1} = \frac{n\beta_1\mu_X}{\sigma_U^2\left\{h^2(1+k^2)+\beta_1^2k^2\right\}}\,,\\
i_{\sigma_U^2\beta_0} &= 0\,, \qquad\qquad\qquad\qquad\qquad i_{\sigma_U^2\beta_1} = \frac{n\beta_1k^2}{\sigma_U^2\left\{h^2(1+k^2)+\beta_1^2k^2\right\}}\,,\\
i_{\beta_0\beta_0} &= \frac{n(1+k^2)}{\sigma_U^2\left\{h^2(1+k^2)+\beta_1^2k^2\right\}}\,, \quad i_{\beta_0\beta_1} = \frac{n\mu_X(1+k^2)}{\sigma_U^2\left\{h^2(1+k^2)+\beta_1^2k^2\right\}}\,,\\
i_{\beta_1\beta_1} &= \frac{2n\beta_1^2k^4}{(h^2(1+k^2)+\beta_1^2k^2)^2} + \frac{n}{(h^2(1+k^2)+\beta_1^2k^2)}\left\{k^4 + \frac{\mu_X^2(1+k^2)}{\sigma_U^2}\right\}.
\end{aligned}
$$

Spatial Gaussian random fields

In the following we give various quantities related to the pseudo pairwise log likelihood of Section 1.2, in the more general case in which we have n independent replications of the vector y, at the same m monitoring stations. In particular, the pseudo score is

$$
\tilde{U}(\theta) = \sum_{i=1}^{n}\sum_{s>r} w_{rs}\frac{\partial\rho_{rs}}{\partial\theta}\frac{1}{(1-\rho_{rs}^2)}\left\{\rho_{rs} - \frac{\rho_{rs}}{\tilde{\sigma}^2}\frac{\tilde{A}_{irs}}{(1-\rho_{rs}^2)} + \frac{(y_{ir}-\tilde{\mu}_{ir})(y_{is}-\tilde{\mu}_{is})}{\tilde{\sigma}^2}\right\},
$$

where $\tilde{\mu}_{ir}$ is the generic element of $\tilde{\mu}_i = X_i\tilde{\beta}$ and $\tilde{A}_{irs} = (y_{ir}-\tilde{\mu}_{ir})^2 + (y_{is} - \tilde{\mu}_{is})^2 - 2\rho_{rs}(y_{ir}-\tilde{\mu}_{ir})(y_{is}-\tilde{\mu}_{is})$.

The asymptotic covariance matrix of $\tilde{\phi}$ is $\Sigma = \mathrm{diag}\,(\Sigma_{11}, \Sigma_{22})$, with

$$
\begin{aligned}
\Sigma_{11} &= \sigma^2\left(\sum_{i=1}^{n}X_i^\top X_i\right)^{-1}\left(\sum_{i=1}^{n}X_i^\top R(\theta)X_i\right)\left(\sum_{i=1}^{n}X_i^\top X_i\right)^{-1}\\
\Sigma_{22} &= \frac{2n(\sigma^2)^2}{(nm+k)^2}\sum_{k=1}^{m}\gamma_k^2\,,
\end{aligned}
$$

where γ_k are the eigenvalues of $R(\theta)$.

We now give details on computation of the quantities needed for matrix K, given by (2). Using the relation $E_{\theta\phi}\{(Y_{ir}-\mu_{ir})(Y_{is}-\mu_{is})(Y_{it}-\mu_{it})(Y_{iu}-\mu_{iu})\} =$

$\sigma^4\rho_{rstu}$, where $\rho_{rstu} = \rho_{rs}\rho_{tu} + \rho_{rt}\rho_{su} + \rho_{ru}\rho_{st}$, we have

$$\begin{aligned}
J_{\theta\theta} &= n\sum_{s>r}\sum_{t>u} w_{rs}w_{tu}\frac{\partial\rho_{rs}}{\partial\theta}\frac{\partial\rho_{tu}}{\partial\theta^\top}\frac{1}{(1-\rho_{rs}^2)}\frac{1}{(1-\rho_{tu}^2)} \\
&\quad \{-\rho_{rs}\rho_{tu} + \rho_{rstu} - \frac{\rho_{rs}}{(1-\rho_{rs}^2)}(\rho_{rrtu} + \rho_{sstu} - 2\rho_{rs}\rho_{rstu}) \\
&\quad -\frac{\rho_{tu}}{(1-\rho_{tu}^2)}(\rho_{ttrs} + \rho_{uurs} - 2\rho_{tu}\rho_{rstu}) + \frac{\rho_{rs}}{(1-\rho_{rs}^2)}\frac{\rho_{tu}}{(1-\rho_{tu}^2)} \\
&\quad (\rho_{rrtt} + \rho_{rruu} + \rho_{sstt} + \rho_{ssuu} + 4\rho_{rs}\rho_{tu}\rho_{rstu} - 2\rho_{tu}\rho_{turr} \\
&\quad -2\rho_{tu}\rho_{tuss} - 2\rho_{rs}\rho_{rstt} - 2\rho_{rs}\rho_{rsuu})\} \\
H_{\theta\beta} &= 0 \\
H_{\theta\sigma^2} &= -\frac{n}{\sigma^2}\sum_{s>r} w_{rs}\frac{\partial\rho_{rs}}{\partial\theta}\frac{\rho_{rs}}{(1-\rho_{rs}^2)} \\
Q &= \mathrm{diag}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n} X_i^\top X_i, \frac{nm+k}{2\sigma^4}\right) \\
\Omega_{\theta\beta} &= 0 \\
\Omega_{\theta\sigma^2} &= \frac{n}{2\sigma^2}\sum_{t}\sum_{s>r} w_{rs}\frac{\partial\rho_{rs}}{\partial\theta}\frac{1}{(1-\rho_{rs}^2)} \\
&\quad \left\{\rho_{rs} - \frac{\rho_{rs}}{(1-\rho_{rs}^2)}\ (\rho_{rrtt} + \rho_{sstt} - 2\rho_{rs}\rho_{rstt}) + \rho_{rstt}\right\}\ .
\end{aligned}$$

Being $H_{\theta\beta} = \Omega_{\theta\beta} = 0$, we have

$$H_{\theta\phi}\Sigma H_{\theta\phi}^\intercal = \Sigma_{22}H_{\theta\sigma^2}H_{\theta\sigma^2}^\intercal\ , \quad \Omega_{\theta\phi}(Q^{-1})^\intercal H_{\theta\phi}^\intercal = \frac{2\sigma^4}{nm}\Omega_{\theta\sigma^2}H_{\theta\sigma^2}^\intercal\ .$$

The resulting K does not depend on β. Moreover,

$$H_{\theta\theta} = n\sum_{s>r} w_{rs}\frac{1+\rho_{rs}^2}{(1-\rho_{rs}^2)^2}\frac{\partial\rho_{rs}}{\partial\theta}\frac{\partial\rho_{rs}}{\partial\theta^\intercal}\ .$$

References

BEVILACQUA, M. & GAETAN, C. (2014). Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. *Statistics and Computing,* to appear.

CATTELAN, M. & SARTORI, N. (2014). Empirical and simulated adjustments of composite likelihood ratio statistics. Tech. Rep. 2/2014, Department of Statistical Sciences - University of Padova. ArXiv: 1403.7093v1.

CHEN, Y. & LIANG, K. (2010). On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. *Biometrika* **97**, 603–620.

CRESSIE, N. (1993). *Statistics for Spatial Data*. New York: Wiley.

DIGGLE, P. J. & RIBEIRO, P. J. (2007). *Model-based Geostatistics*. New York: Springer.

GEYS, H., MOLENBERGHS, G. & RYAN, L. (1999). Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *J. Amer. Statist. Assoc.* **94**, 734–745.

GHOSH, A., WRIGHT, F. A. & ZOU, F. (2013). Unified analysis of secondary traits in case–control association studies. *J. Amer. Statist. Assoc.* **108**, 566–576.

GONG, G. & SAMANIEGO, F. (1981). Pseudo maximum likelihood estimation: Theory and applications. *Ann. Statist.* **9**, 861–869.

GUOLO, A. (2011). Pseudo-likelihood inference for regression models with misclassified and mismeasured variables. *Statist. Sinica* **21**, 1639–1663.

HAUCK, W. & DONNER, A. (1977). Wald's test as applied to hypotheses in logit analysis. *J. Amer. Statist. Assoc.* **72**, 851–853.

JØRGENSEN, B. & KNUDSEN, S. J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scand. J. Statist.* **31**, 93–114.

KENT, J. (1982). Robust properties of likelihood ratio tests. *Biometrika* **69**, 19–27.

LIANG, K. & SELF, S. (1996). On the asymptotic behaviour of the pseudolikelihood ratio test statistic. *J. R. Statist. Soc.* B **58**, 785–796.

LINDSAY, B. G. (1988). Composite likelihood methods. *Contemp. Math.* **80**, 221–240.

MOLENBERGHS, G. & VERBEKE, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

PACE, L., SALVAN, A. & SARTORI, N. (2011). Adjusting composite likelihood ratio statistics. *Statist. Sinica* **21**, 129–148.

PAKEL, C., SHEPHARD, N. & SHEPPARD, K. (2011). Nuisance parameters, composite likelihood and a panel of GARCH models. *Statist. Sinica* **21**, 307–329.

PIERCE, D. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Ann. Statist.* **10**, 475–478.

SEVERINI, T. A. (2005). *Elements of Distribution Theory*. Cambridge: Cambridge University Press.

STAFFORD, J. (1996). A robust adjustment of the profile likelihood. *Ann. Statist.* **24**, 336–352.

VARIN, C., REID, N. & FIRTH, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21**, 5–42.

WANG, H., HU, X. J., MCBRIDE, M. L. & SPINELLI, J. J. (2014). Analysis of counts with two latent classes, with application to risk assessment based on physician-visit records of cancer survivors. *Biostatistics* **15**, 384–397.

YOUNG, G. A. & SMITH, R. L. (2005). *Essentials of Statistical Inference*. Cambridge: Cambridge University Press.