Working Paper Series, N. 4, May 2004

A note on composite likelihood inference and model selection

Cristiano Varin
Department of Statistical Sciences
University of Padua
Italy

Paolo Vidoni
Department of Statistics
University of Udine
Italy

Abstract: A composite likelihood consists in a combination of valid likelihood objects, usually related to small subsets of data. The merit of composite likelihood is to reduce the computational complexity so that it is possible to deal with large datasets and very complex models, even when the use of standard likelihood or Bayesian methods is not feasible. In this paper, we aim to suggest an integrated, general approach to inference and model selection using composite likelihood methods. In particular, we introduce an information criterion for model selection based on composite likelihood. Applications to modelling time series of counts through dynamic generalized linear models and to the analysis of the well-known Old Faithful geyser dataset are also given.

Keywords: AIC; dynamic generalized linear models; hidden Markov model; Old Faithful geyser data; overdispersion; pairwise likelihood; pseudolikelihood; TIC; tripletwise likelihood.

Contents

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

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

Corresponding author:
Firstname Lastname
tel: +39 049 827 4192
sammy@stat.unipd.it

A note on composite likelihood inference and model selection

Cristiano Varin
Department of Statistical Sciences
University of Padua
Italy

Paolo Vidoni
Department of Statistics
University of Udine
Italy

Abstract: A composite likelihood consists in a combination of valid likelihood objects, usually related to small subsets of data. The merit of composite likelihood is to reduce the computational complexity so that it is possible to deal with large datasets and very complex models, even when the use of standard likelihood or Bayesian methods is not feasible. In this paper, we aim to suggest an integrated, general approach to inference and model selection using composite likelihood methods. In particular, we introduce an information criterion for model selection based on composite likelihood. Applications to modelling time series of counts through dynamic generalized linear models and to the analysis of the well-known Old Faithful geyser dataset are also given.

Keywords: AIC; dynamic generalized linear models; hidden Markov model; Old Faithful geyser data; overdispersion; pairwise likelihood; pseudolikelihood; TIC; tripletwise likelihood.

1 Introduction

In a number of applications, the presence of large correlated datasets or the specification of very complex models make unfeasible the use of the likelihood function, since too computationally demanding. One possibility is to avoid ordinary likelihood methods, or Bayesian strategies, and to adopt simpler pseudolikelihoods, like those belonging to the composite likelihood class (Lindsay 1988, Cox & Reid 2003). A composite likelihood consists in a combination of valid likelihood objects, usually related to small subsets of data. It has good theoretical properties and it behaves well in many complex applications, for example spatial statistics (Besag 1974, Vecchia 1988, Hjort & Omre 1994, Heagerty & Lele 1998, Nott & Rydén 1999), multivariate survival analysis (Parner 2001), generalized linear mixed models (Renard, Molenberghs & Geys 2004), frailty models (Henderson & Shimakura 2003), genetics (Fearnhead & Donnelly 2002).

In this paper, we aim to set and justify an integrated, general approach for inference and model selection, using composite likelihood methods. In particular, we focus on a new information criterion for model selection, which is the counterpart of the Takeuchi's information criterion (Takeuchi 1976, Shibata 1989, Burnham & Anderson 2002), abbreviated as TIC, based on composite likelihood. The paper is organized as follow. In Section 2, we restore the concept of composite likelihood and derive a first-order unbiased composite likelihood selection statistics. In Section 3, we test our methodology with regard to dynamic generalized linear models (West & Harrison 1997) for time series of counts and to the analysis of the well-known Old Faithfull geyser dataset (Azzalini & Bowman 1990).

2 Inference and model selection using composite likelihood

The term composite likelihood (Lindsay 1988) denotes a rich class of pseudolikelihoods based on likelihood-type objects. We start by restoring its definition.

Definition 1. Let $\big\{f(y;\theta), y \in \mathcal{Y}, \theta \in \Theta\big\}$ be a parametric statistical model, with $\mathcal{Y} \subseteq \mathbb{R}^n$, $\Theta \subseteq \mathbb{R}^d$, $n \geq 1$, $d \geq 1$. Consider a set of events $\{\mathcal{A}_i : \mathcal{A}_i \subseteq \mathcal{F}, i \in I\}$, where $I \subseteq \mathbb{N}$ and $\mathcal{F}$ is some sigma algebra on $\mathcal{Y}$. Then, a *composite likelihood* is a function of θ defined as

$$\mathrm{CL}_f(\theta; y) = \prod_{i \in I} f(y \in \mathcal{A}_i; \theta)^{w_i},$$

where $f(y \in \mathcal{A}_i;\theta) = f\big(\{y_j \in y : y_j \in \mathcal{A}_i\};\theta\big)$, with $y = (y_1, \dots, y_n)$, while $\{w_i, i \in I\}$ is a set of suitable weights. The associated *composite loglikelihood* is $\log \mathrm{CL}_f(\theta; y)$.

Example We present three important examples of composite loglikelihoods.

(i) The *"full" loglikelihood*, given by $\log \mathrm{L}(\theta; y) = \log f(y;\theta)$.

(ii) The *pairwise loglikelihood*, defined as $\log \mathrm{PL}(\theta; y) = \sum_{j<k} \log f(y_j, y_k; \theta) w_{(j,k)}$, where the summation is over all the pairs (y_j, y_k), $j, k = 1, \dots, n$, of observations. With a slight abuse of notation we denote with $w_{(j,k)}$ the weight associated to (y_j, y_k). Analogously, we may define the *tripletwise loglikelihood*, where triplets of observations are taken into account, and so on.

(iii) The *Besag's pseudologlikelihood*, defined as $\log \mathrm{BPL}(\theta; y) = \sum_{j=1}^{n} \log f(y_j | y_{(-j)}; \theta) w_j$, where the summation is over all the conditional events $\{y_j | y_{(-j)}\}$, with $y_{(-j)}$ the subset of the components of vector y without the j-th element y_j (Besag 1974).

◇

The usefulness of the composite likelihood ideas naturally arises in an estimating function framework (Heyde 1997). Indeed, given the set of realized events $\{\mathcal{A}_i : \mathcal{A}_i \subseteq \mathcal{F}, i \in I\}$, the *maximum composite likelihood estimator* is usually defined as a solution of the *composite likelihood equation*

$$\nabla \log \mathrm{CL}_f(\theta; y) = 0, \tag{1}$$

where $\nabla \log \mathrm{CL}_f(\theta; y) = \sum_{i\in I} \nabla \log f(y \in \mathcal{A}_i; \theta) w_i$ is the *composite score function.* Hereafter, we use the notation $\nabla h(\theta)$ for the column vector of the first partial derivatives of function $h(\theta)$, while $\nabla^2 h(\theta)$ is the symmetric matrix of second derivatives. Since the composite score function is a linear combination of unbiased estimating functions, then, under suitable regularity conditions, the maximum composite likelihood estimator $\widehat{\theta}_{\mathrm{MCL}}(Y)$ is consistent and asymptotically normal distributed, that is

$$\widehat{\theta}_{\mathrm{MCL}}(Y) \xrightarrow{d} \mathcal{N}\{\theta, H(\theta)^{-1} J(\theta) H(\theta)^{-\mathrm{T}}\},$$

where

$$J(\theta) = \mathrm{var}_{f(y)}\{\nabla \log \mathrm{CL}(\theta; Y)\} \quad \text{and} \quad H(\theta) = E_{f(y)}\{\nabla^2 \log \mathrm{CL}(\theta; Y)\}.$$

Although a deep study on efficiency issues is still lacking, some useful results may be found in Heyde (1997), Heagerty & Lele (1998) and Nott & Rydén (1999). A recent account on these key aspects of pseudolikelihood inferential procedures is given by Cox & Reid (2003).

In this paper, we aim to emphasize the opportunity of using composite likelihood methods both for making inference and for model selection purposes. The extension of the pseudolikelihood framework to model selection is natural, and in some sense obvious, however, to our concern, it is not explicitly considered so far in the literature.

In the sequel, we shall introduce a predictive model selection procedure based on the following generalization of the Kullback-Leibler divergence.

Definition 2. Given two density functions $g(z)$ and $h(z)$ for a random variable Z, the associated *composite Kullback-Leibler information* is defined by the non-negative quantity

$$I_c(g, h) = E_{g(z)}\{\log(\mathrm{CL}_g(Z)/\mathrm{CL}_h(Z))\} = \sum_{i\in I} E_{g(z)}\big\{\log g(Z \in \mathcal{A}_i) - \log h(Z \in \mathcal{A}_i)\big\} w_i,$$

where the expectation is with respect to $g(z)$, $\log \mathrm{CL}_g(Z) = \sum_{i\in I} w_i \log g(Z \in \mathcal{A}_i)$ and $\log \mathrm{CL}_h(Z) = \sum_{i\in I} w_i \log h(Z \in \mathcal{A}_i)$.

Note that $I_c(g, h)$ is a linear combination of the ordinary Kullback-Leibler divergences, corresponding to each likelihood object forming the composite likelihood function. Moreover, $I_c(g, h)$ is a key quantity in order to asses the properties of composite likelihood inferential and model selection procedures, under an possibly misspecified statistical model.

In this context, we focus on a new information criterion, defined as a first order unbiased estimator for a target quantity related to the expected composite Kullback-Liebler information between the true unknown density of a potential future observation and the corresponding estimated density. Since this criterion is, in some sense, a composite likelihood generalization of the TIC, and its derivation follows a standard procedure, we shall present only the key steps. Let us consider the sample $Y = (Y_1, \dots, Y_n)$ and a parametric statistical model specified by the family of density functions $\big\{f(y; \theta), y \in \mathcal{Y}, \theta \in \theta\big\}$, with respect to a common dominating measure.

There could be several plausible statistical models for Y, which may or may not contain the true $g(y)$. We would like to choose the model which offers the most satisfactory predictive description for the observed data y. More precisely, let Z be a future random variable, defined as an independent copy of Y, we are interested in the choice of the "best" model for forecasting Z, given a realization of Y, using composite likelihood methods.

As for the TIC, model selection can be approached on the basis of the expected composite Kullback-Leibler information between the true density $g(z)$ and the estimated density $\widehat{f}(z) = f(z; \widehat{\theta}_{\mathrm{MCL}}(Y))$, under the assumed statistical model. Namely, we select the model which minimizes $E_{g(y)}\{I_c(g, \widehat{f})\}$ or, analogously, which maximizes the theoretical target quantity

$$\varphi(g, f) = \sum_{i \in I} E_{g(y)}\big[E_{g(z)}\{\log f(Z \in \mathcal{A}_i; \widehat{\theta}_{\mathrm{MCL}}(Y))\}\big] w_i, \tag{2}$$

called the *expected predictive composite loglikelihood*

Since the computation of (2) requires the knowledge of the true density $g(z)$, and then it is in fact unfeasible, model selection may be approached by maximizing a selection statistic defined as a suitable estimator for $\varphi(g, f)$. As proved in the Appendix, under standard regularity conditions, the following information criterion is based on a selection statistic, which is a first order unbiased estimator for $\varphi(g, f)$.

Definition 3. Let us consider a random sample Y, as previously defined. The *composite likelihood information criterion* (CLIC) selects the model maximizing

$$\Psi^c(Y; f) = \Psi(Y; f) + \mathrm{tr}\{\widehat{J}(Y)\widehat{H}(Y)^{-1}\}, \tag{3}$$

where

$$\Psi(Y; f) = \log \mathrm{CL}_f(\widehat{\theta}_{\mathrm{MCL}}(Y); Y) = \sum_{i \in I} \log f(Y \in \mathcal{A}_i; \widehat{\theta}_{\mathrm{MCL}}(Y)) w_i,$$

and $\widehat{J}(Y)$ and $\widehat{H}(Y)$ are suitable consistent, first order unbiased, estimators for $J(\theta)$ and $H(\theta)$, respectively, based on Y.

Note that the selection statistics (3) corresponds to a modification of the estimated composite loglikelihood $\Psi(Y; f)$, obtained by adding penalty term as for TIC in the ordinary likelihood framework. Moreover, it easy to see that, in the particular case when the composite likelihood is in fact the likelihood function, the composite Kullback-Leibler divergence $I_c(g, \widehat{f})$ equals the usual Kullback-Leibler divergence. Thus, as obtained in the following example, part (i), the CLIC equals the TIC and, when the model is true or it is a good approximation to the "truth", it coincide with the AIC, namely the Akaike's information criterion (Akaike 1973). In the following section, we shall consider the CLIC based on the selection statistic $-\Psi^c(Y; f)$, which is in accordance with the usual representation chosen for the AIC and the TIC.

The CLIC is a generalized information criterion, which is useful whenever the computation of the likelihood, and of the penalty term given by the TIC, is too computationally demanding, and then not convenient or even possible. Furthermore, it is worthwhile to emphasize that, when the model is the true one, the CLIC does not usually coincide with a sort of *composite*-AIC, that is $\Psi(Y; f) - d$, unless we consider the "full" likelihood itself. In general, since the composite likelihood differs from the ordinary likelihood function, a composite likelihood analogous of the well-known information identity does not usually hold and the penalty term $\mathrm{tr}\{J(Y)H(Y)^{-1}\}$ is expected to differ from d.

Example (continued)

(i) The CLIC for the "full" likelihood is based on

$$\Psi^c(Y; f) = \log f(Y; \widehat{\theta}_{\mathrm{ML}}(Y)) + \mathrm{tr}\{\widehat{J}(Y)\widehat{H}(Y)^{-1}\}, \tag{4}$$

where $\widehat{J}(Y)$ and $\widehat{H}(Y)$ are convenient estimators for $J(\theta) = \mathrm{var}_{g(y)}\{\nabla \log f(Y; \theta)\}$ and $H(\theta) = E_{g(y)}\{\nabla^2 \log f(Y; \theta)\}$, respectively; $\widehat{\theta}_{\mathrm{ML}}(Y)$ is the maximum likelihood estimator. In this case the CLIC corresponds to the TIC. Moreover, if we (optimistically) assume that the model is correctly specified for Y, then $J(\theta) = -H(\theta)$ and (4) simplifies to the familiar AIC

$$\Psi^c(Y; f) = \log f(Y; \widehat{\theta}_{\mathrm{ML}}(Y)) - d.$$

(ii) The CLIC for the pairwise likelihood is based on

$$\Psi^c(Y; f) = \sum_{j<k} \log f(Y_j, Y_k; \widehat{\theta}_{\mathrm{MPL}}(Y)) w_{(j,k)} + \mathrm{tr}\{\widehat{J}(Y)\widehat{H}(Y)^{-1}\},$$

where $\widehat{\theta}_{\mathrm{MPL}}(Y)$ is the maximum pairwise likelihood estimator and $\widehat{J}(Y)$, $\widehat{H}(Y)$ estimate, respectively,

$$J(\theta) = \sum_{j<k} \sum_{l<m} \mathrm{cov}_{g(y)}\{\nabla \log f(Y_j, Y_k; \theta), \nabla \log f(Y_l, Y_m; \theta)\} w_{(j,k)} w_{(l,m)},$$

$$H(\theta) = \sum_{j<k} E_{g(y)}\{\nabla^2 \log f(Y_j, Y_k; \theta)\} w_{(j,k)}.$$

(iii) The CLIC for the Besag's pseudolikelihood is based on

$$\Psi^c(Y; f) = \sum_{j=1}^{n} \log f(Y_j | Y_{(-j)}; \widehat{\theta}_{\mathrm{MBPL}}(Y)) w_j + \mathrm{tr}\{\widehat{J}(Y)\widehat{H}(Y)^{-1}\},$$

where $\widehat{\theta}_{\mathrm{MBPL}}(Y)$ is the maximum Besag's pseudolikelihood estimator and $\widehat{J}(Y)$, $\widehat{H}(Y)$ estimate, respectively,

$$J(\theta) = \sum_{j=1}^{n} \sum_{k=1}^{n} \mathrm{cov}_{g(y)}\{\nabla \log f(Y_j | Y_{(-j)}; \theta), \nabla \log f(Y_k | Y_{(-k)}; \theta)\} w_j w_k,$$

$$H(\theta) = \sum_{j=1}^{n} E_{g(y)}\{\nabla^2 \log f(Y_j | Y_{(-j)}; \theta)\} w_j.$$

◇

Finally, we briefly mention two further important points. The first one regards the choice of the weights in the composite likelihood. Typically, with regard to the pairwise likelihood, the weights are chosen in order to cut off the pairs of not-neighboring observations, which should be less informative. The simpler weighting strategy is then to estimate the correlation range and put equal to zero all the pairs at a distance larger than such a range. A more accurate approach consists in choosing the pairs under some optimality criterion (see, for example, Nott & Rydén (1999)).

The second one concerns the efficient estimation of $J(\theta)$ and $H(\theta)$. The latter does not pose difficulties and, under standard regularity conditions, a consistent estimator is $\widehat{H}(\widehat{\theta}_{\mathrm{MCL}}(Y)) = \nabla^2 \log \mathrm{CL}(\widehat{\theta}_{\mathrm{MCL}}(Y); Y)$. Much more complex is to estimate $J(\theta)$, since its intuitive empirical estimate $\widehat{J}(\theta; y) = \nabla \log \mathrm{PL}(\theta; y) \nabla \log \mathrm{PL}(\theta; y)^T$ vanishes when evaluated at the maximum composite likelihood estimate. In practice, the estimation of $J(\theta)$ is performed by means of different strategies, depending on the selection problem taken into account and on the particular composite likelihood which is considered.

In Section 3.1, we shall analyze discrete values time series data using the pairwise likelihood and $J(\theta)$ is suitably estimated by means of a *reuse sampling* procedure, called *window subsampling* (Carlstein 1986, Hall & Jing 1996, Garcia-Soidan & Hall 1997, Heagerty & Lele 1998, Lumley & Heagerty 1999). We briefly restore the basic procedure behind window subsampling strategies, in the case of time series data $y = (y_1, \ldots, y_n)$. More precisely, the method is based on the following steps: (i) consider a set of overlapping subseries $y_i^{i+m} = (y_i, \ldots, y_{i+m-1})$ of dimension m; (ii) compute the empirical estimators $\widehat{J}(\widehat{\theta}_{\mathrm{MPL}}; y_i^{i+m})$ related to each subseries; (iii) take their average, with a suitable scale transformation accounting for the subseries dimension, giving the window subsampling estimate

$$\widehat{J}^{(m)}(\widehat{\theta}_{\mathrm{MPL}}; y) = \frac{1}{(n-m+1)} \sum_{i=1}^{n-m+1} \widehat{J}(\widehat{\theta}_{\mathrm{MPL}}; y_i^{i+m}) \frac{m}{n}.$$

Heagerty & Lumley (2000) show that the optimal dimension m for the subwindows is $Cn^{1/3}$, where C is a constant, with suggested values between 4 and 8, which depends on the strength of the dependence within the data.

3 Applications

In this section, we shall present two applications of composite likelihood methods, with the aim of emphasizing some interesting features of these pseudolikelihood-based inferential procedures. The first application involves simulations related to time series of counts, modelled by dynamic generalized linear models, in the presence of overdispersion. The second one concerns the analysis of the well-known the Old Faithful geyser dataset (Azzalini & Bowman 1990).

3.1 Testing overdispersion in dynamic generalized linear models

When modelling counts data, usually by means of the Poisson distribution, a common problem to be faced is that of accounting for potential overdispersion (McCullagh

& Nelder 1989). In this context a standard overdispersion model is that one obtained by replacing the Poisson distribution with the negative binomial one and a simple procedure for testing the presence of overdispersion is to compare these two alternative nested models, using for example the AIC (Lindsey 1999). Here, by means of a simulation study, we discuss the use of composite likelihood methods for testing the presence of overdispersion with regard to time series count data. Since, in this case, the ordinary likelihood-based methods are not feasible, we use the CLIC to compare two dynamic generalized linear models (West & Harrison 1997) based, respectively, on the Poisson and on the negative binomial distribution.

A dynamic generalized linear models may be viewed as a state space model, that is as double stochastic process $\{Y_i, X_i\}_{i\geq 1}$, where the observable random variables $\{Y_i\}_{i\geq 1}$ are assumed to be conditionally independent given a hidden Markov process $\{X_i\}_{i\geq 1}$, describing the latent unobserved evolution of the system. Hereafter, $f(x_1;\theta)$ and $f(x_i|x_{i-1};\theta)$, $i > 1$, are, respectively, the initial and the transition probability (density) functions of the latent Markov process, while $f(y_i|x_i;\theta)$, $i \geq 1$, is the conditional probability (density) function of Y_i given $X_i = x_i$, which does not depend on i; $\theta \in \Theta \subseteq \mathbb{R}^d$ indicates, as usual, the unknown parameter.

In this framework, a simple model for describing the time series count data $y = (y_1, \ldots, y_n)$ is the Poisson-AR(1) model, where the latent process is an AR(1) process and the observations are conditionally independent, following a Poisson distribution. More precisely,

$$\begin{cases} Y_i|X_i = x_i \sim \mathrm{Po}(\exp\{x_i\}),\ i = 1, \ldots, n, \\ X_i = \lambda X_{i-1} + \epsilon_i,\ i = 2, \ldots, n, \end{cases}$$

with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, $i = 1, \ldots, n$, independent random variables. We assume $|\lambda| < 1$, so that the latent AR(1) model is indeed stationary and we set $X_1 \sim \mathcal{N}(0, \sigma^2/(1 - \lambda^2))$.

An alternative model, useful for describing overdispersion, is obtained by substituting the Poisson distribution with a negative binomial one, that is by assuming that $Y_i|X_i = x_i$, $i = 1, \ldots, n$, follows a negative binomial distribution with mean $\mu_i = \exp\{x_i\}$ and an additional size parameter $\kappa > 0$. Thus, adopting a suitable parameterization, we state

$$f(y_i|x_i;\kappa) = \frac{\Gamma(\kappa^{-1} + y_i)}{\Gamma(\kappa^{-1})y_i!}\left(\frac{\kappa\mu_i}{1+\kappa\mu_i}\right)^{y_i}\left(\frac{1}{1+\kappa\mu_i}\right)^{1/\kappa},\ y_i = 0, 1, \ldots,$$

for $i = 1, \ldots, n$, where $\Gamma(\cdot)$ is the gamma function. Note that these two models are nested, since the second one tends to the first one as $\kappa \to 0$. The comparison between these two models could be of practical interest, since the presence of the latent AR(1) process, in the Poisson-based model, may not give a complete description of data overdispersion.

Although the above models are appealing, their analysis using standard likelihood inference and model selection procedures could be problematic, since the computation of the likelihood function needs the solution of the following n-dimensional intractable integral

$$\mathrm{L}(\theta; y) = \int \prod_{i=1}^{n} f(y_i|x_i;\theta) f(x_i|x_{i-1};\theta) \mathrm{d}x_1 \cdots \mathrm{d}x_n,$$

where $f(x_1|x_0;\theta) = f(x_1;\theta)$. A known solution to this problem involves computational intensive, simulation-based, approximations (see, for example, Durbin & Koopman (2001) and references therein). In this example, we perform a different strategy based on a suitable composite likelihood and, in particular, we use the CLIC for deciding which of the two alternative models is more suited for the data under examination. We consider the pairwise likelihood formed by taking the $n-1$ subsequent pairs of observations obtained from data $y = (y_1, \ldots, y_n)$

$$\mathrm{PL}(\theta; y) = \prod_{i=2}^{n} \iint f(y_i|x_i;\theta) f(y_{i-1}|x_{i-1};\theta) f(x_i|x_{i-1};\theta) f(x_{i-1};\theta) \mathrm{d}x_i \mathrm{d}x_{i-1}. \quad (5)$$

Note that the evaluation of the $n-1$ two-dimensional intractable integrals, defining $\mathrm{PL}(\theta; y)$, is much simpler than the computation of the full likelihood.

In the next lines, we present a simulation study in order to assess the performance of the maximum pairwise likelihood estimators and to test the usefulness of model selection based on the CLIC. The $n-1$ integrals involved in the pairwise likelihood (5) are computed by means of standard numerical approximations via deterministic quadrature rules, such as the Gauss-Hermite procedure. In the first simulation, we generate 500 datasets with $n = 300$ observations from the Poisson-AR(1) model with $\lambda = 0.35$ and $\sigma = 1$. The 299 two-dimensional integrals forming the pairwise likelihood (5) are approximated by means of a double Gauss-Hermite quadrature with 10 nodes for each dimension, that is a total amount of 100 nodes. Using the adaptive quadrature gives similar results. The sample means of the simulated pairwise likelihood estimators for λ and σ are, respectively, 0.34996 and 0.9820, with sample standard deviations 0.1128 and 0.0831. We repeat the simulation with $\lambda = -0.5$ and $\sigma = 1.2$. We obtain that the sample means of the simulated pairwise likelihood estimators for λ and σ are, respectively, -0.4922 and 1.1397, with sample standard deviations 0.0671 and 0.0897. In both cases the maximum pairwise likelihood estimators show a slight tendency to underestimate σ.

The second simulation study deals with model selection. We generate 100 datasets with $n = 300$ observations from the negative binomial-AR(1) model with $\lambda = 0.35$, $\sigma = 0.5$ and alternative values for the size parameter κ, namely $1, 1/2, 1/4$, $1/8, 0$. Here, $\kappa = 0$ means that the simulations are in fact from the Poisson-AR(1) model. We compare the negative binomial-AR(1) and the Poisson-AR(1) models using the CLIC. The matrix $J(\theta)$ is estimated with a suitable window subsampling procedure involving 251 overlapped subseries of observations of dimension $m = 50$. Table 1 (a) gives the frequencies of correct model selection over the 100 simulated datasets, for different values for κ. We do not report the results for $\kappa > 1$, since in this case the CLIC indicates almost always the true model, that is the negative binomial one. These results give a preliminary justification for using composite likelihood methods for model selection purposes.

Note that, as expected, whenever κ approaches to zero, the CLIC tends to choose more often the wrong model, since there is a potential slight overdispersion in the observations and then the two models provide an almost equivalent data description, as detected by the composite Kullback-Liebler divergence. When κ is less than $1/4$, the observed values of the selection statistics for the two alternative models are

(a)

κ	1	1/2	1/4	1/8	0
	100	90	64	59	55

(b)

κ	1	1/2	1/4	1/8	0
	99	82	64	45	58

Table 1: Frequencies of correct model selection over the 100 simulated datasets with $\kappa = 1, 1/2, 1/4, 1/8, 0$ and (a) $\lambda = 0.35$, $\sigma = 0.5$ (b) $\lambda = -0.6$, $\sigma = 0.7$; the case $\kappa = 0$ indicates a true Poisson-AR(1) model.

usually very closed. This is confirmed by the boxplots of Figure 1 (a), where we summarize the observed differences between the CLIC selection statistics for the two alternative models. In particular, for $\kappa = 1/8, 0$, the differences are usually

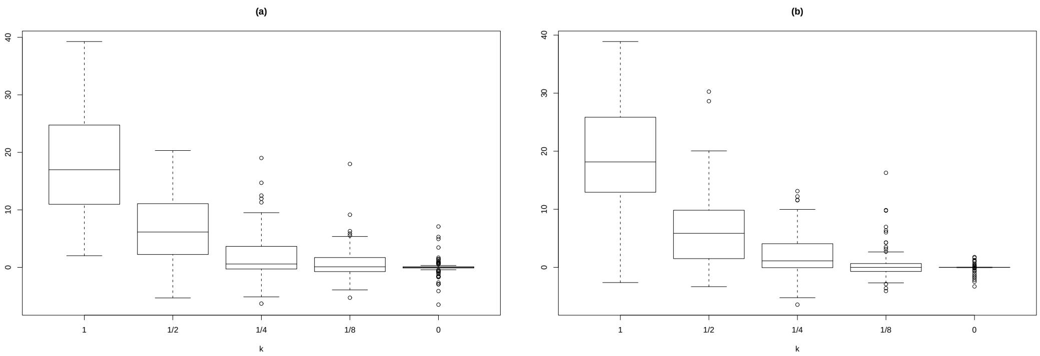

Figure 1: Boxplots of the observed differences between the CLIC selection statistics for the two alternative models, based on the 100 simulated datasets with $\kappa = 1, 1/2, 1/4, 1/8, 0$ and (a) $\lambda = 0.35$, $\sigma = 0.5$ (b) $\lambda = -0.6$, $\sigma = 0.7$.

negligible, since the values of the selection statistics are around 900.

We repeat the simulation study with $\lambda = -0.6$, $\sigma = 0.7$ and $\kappa = 1, 1/2, 1/4, 1/8, 0$. In this case, as presented in Table 1 (b) and Figure 1 (b), we obtain similar results.

3.2 The Old Faithful geyser data

Here we present an application to the Old Faithful geyser dataset discussed in Azzalini & Bowman (1990), with the aim of comparing binary Markov and hidden Markov models. In this context, we shall show that composite likelihood methods, based on the triplewise likelihood, lead to good inferential conclusions. However, for model selection purposes, the associated CLIC fails to distinguish between the two alternative models.

The data consists in the time series of the duration of the successive eruptions at the Old Faithful geyser in the Yellowstone National Park in the period from 1 to 15 August 1985. Azzalini & Bowman (1990) and MacDonald & Zucchini (1997, §4.2) consider a binary version of this data, described as short or long eruptions, obtained by thresholding the time series at 3 minutes. Let us label the short and the long eruptions with the states 0 and 1, respectively. The random variables N_r, $r = 0, 1$, indicate the corresponding number of observed eruptions. We have that $N_0 = 105$ and $N_1 = 194$; moreover, the one-step observed transition matrix is

$$\begin{pmatrix} N_{00} & N_{01} \\ N_{10} & N_{11} \end{pmatrix} = \begin{pmatrix} 0 & 104 \\ 105 & 89 \end{pmatrix},$$

where N_{rs}, $r, s = 0, 1$, in the number of one-step transition from state r to state s. Note that no transition from state 0 to itself is occurred. For the models discussed in the sequel, it is also relevant to consider the two-steps transitions. Since $N_{00} = 0$, only five triplets were observed. Being N_{rst}, $r, s, t = 0, 1$, the number of two-step transitions from state r to state s and then to state t, the non-null observations are: $N_{010} = 69$, $N_{110} = 35$, $N_{011} = 35$, $N_{101} = 104$ and $N_{111} = 54$.

In Azzalini & Bowman (1990), the time series is first analyzed by a first-order Markov chain model. Then, since this model does not fit very well the autocorrelation function, they move to a second-order Markov chain model, which seems more plausible. The same data are also analyzed by MacDonald & Zucchini (1997, §4.2). They consider some hidden Markow models based on the binomial distribution and compare them with the Markov chain models of Azzalini & Bowman (1990), using the AIC and the BIC, namely the Bayesian information criterion (Schwarz 1978). They conclude, see Table 2 (a), that both the AIC and the BIC indicate that the model for Old Faithful geyser data is the second-order Markov chain, even if the two-state binomial hidden Markov model is quite similar in performance. In our opinion, this conclusion is questionable, since the observed value of the AIC and of the BIC selection statistics are very closed for the two alternative models. Indeed, by removing the last observation, we obtain that both the AIC and the BIC prefer the two-state hidden Markov model, as reported in Table 2 (b).

(a)

Model	AIC	BIC
MC2	**262.24**	**277.04**
HMM	262.62	277.42

(b)

Model	AIC	BIC
MC2	260.37	275.17
HMM	**260.22**	**275.02**

Table 2: Old Faithful geyser dataset. Values for the AIC and the BIC for the second-order Markov chain (MC2) and the two-state hidden Markov model (HMM) based (a) on all the 299 observations and (b) on the observations after removing the last one.

Let us start by recalling that an hidden Markov model indicates a state space model $\{Y_i, X_i\}_{i \geq 1}$, with a discrete-valued latent process. More precisely, $\{X_i\}_{i \geq 1}$ is an hidden Markov chain. We shall assume that this latent Markov chain is stationary and irreducible, with $w \in \mathbb{N}^+$ states. Thus, the bivariate process $\{Y_i, X_i\}_{i \geq 1}$ is itself stationary. The likelihood function for the unknown parameter θ, based on the available observations $y = (y_1, \ldots, y_n)$, is

$$\mathrm{L}(\theta; y) = \sum_{x_1} \ldots \sum_{x_n} f(x_1) f(y_1 | x_1; \theta) \prod_{i=2}^{n} f(x_i | x_{i-1}; \theta) f(y_i | x_i; \theta), \tag{6}$$

where the summations are over the w states and the initial probability function $f(x_1)$ is not necessarily that related to the stationary distribution of the chain (Leroux 1992). Since the evaluation of (6) requires $O(w^n)$ computations, MacDonald & Zucchini (1997) rearrange the terms in (6) in order to reduce significantly the computational burden. However, this rearrangement does not seem useful, if one desires to compute likelihood quantities such as the derivatives of function $\log \mathrm{L}(\theta; y)$, with respect to θ, and their expectations with respect to the true unknown distribution, which are required for the ordinary likelihood-based TIC.

An alternative to the full likelihood may be found within the composite likelihood family. Again, the simpler useful composite likelihood is the pairwise likelihood, based on the pairs of subsequent observations,

$$\mathrm{PL}(\theta; y) = \prod_{i=2}^{n} \sum_{x_{i-1}, x_i} f(x_{i-1}; \theta) f(x_i | x_{i-1}; \theta) f(y_{i-1} | x_{i-1}; \theta) f(y_i | x_i; \theta),$$

where the summation is over the pairs of subsequent latent observations. However, this is not a good candidate for our inferential and model selection problem, since, for second-order Markov chains, the composite likelihood equation has an infinity number of solutions. Then, we have to move to the, slightly more complex, tripletwise likelihood. For a hidden Markow model, the tripletwise likelihood, based on the triplets of subsequent observations, is given by

$$\mathrm{TL}(\theta; y) = \prod_{i=3}^{n} \sum_{x_{i-2}, x_{i-1}, x_i} f(x_{i-2}, x_{i-1}, x_i; \theta) f(y_{i-2} | x_{i-2}; \theta) f(y_{i-1} | x_{i-1}; \theta) f(y_i | x_i; \theta),$$

where $f(x_{i-2}, x_{i-1}, x_i; \theta)$ is the joint probability function of (X_{i-2}, X_{i-1}, X_i) and the summation is over the triplets of subsequent latent observations. When dealing with binary data, as with the dataset under discussion, if we assume stationarity, the tripletwise likelihood looks like

$$\mathrm{TL}(\theta; y) = \prod_{r,\, s,\, t \in \{0,1\}} \mathrm{pr}(Y_{i-2} = r, Y_{i-1} = s, Y_i = t)^{N_{rst}},$$

where N_{rst}, $r, s, t = 0, 1$, defined above, is the number of realized events $(Y_{i-2} = r, Y_{i-1} = s, Y_i = t)$, $i > 2$. Note that, in this case, $\mathrm{TL}(\theta; y)$ consists in eight terms. However, since in the Old Faithful geyser dataset there are no transitions from state 0 to itself, in fact only five terms enter in the function.

In the following, we shall present the two models under competition and we compute the corresponding tripletwise likelihoods. We implicitly assume that the assumptions required for asymptotic unbiasedness of the CLIC, see the Appendix, are fulfilled. In fact, the consistency and asymptotic normality of the tripletwise likelihood can be proved using the framework suggested in Renard (2002, §3.2), whose assumptions are here satisfied because the Markov and the hidden Markov models are supposed to be stationary.

The first model is a second-order two-states Markov chain. In order to compute the probabilities associated with the triplets (Y_{i-2}, Y_{i-1}, Y_i), $i > 2$, is convenient to consider the transition probability matrix whose entries are $\Delta_{(sr)(ts)} = \text{pr}(Y_{i-1} = s, Y_i = r | Y_{i-2} = t, Y_{i-1} = s) = \text{pr}(Y_i = r | Y_{i-1} = s, Y_{i-2} = t)$, $r, s, t = 0, 1$, $i > 2$,

$$\Delta^{MC2} = \begin{pmatrix} 1-k & k & 0 & 0 \\ 0 & 0 & b & 1-b \\ 0 & 1 & 0 & 0 \\ 0 & 0 & c & 1-c \end{pmatrix},$$

with $b, c \in (0, 1)$ unknown parameters and k is any real number in (0,1). The presence of this arbitrary value k is not relevant since, as noted by MacDonald & Zucchini (1997, §4.2), the pair $(0, 0)$ can be disregarded without loss of information.

The associated bivariate stationary distribution is

$$\pi^{MC2} = \frac{1}{2c + (1-b)} \begin{pmatrix} 0 & c \\ c & (1-b) \end{pmatrix}$$

and then the joint probabilities for the five relevant triplets are, for $i > 2$,

$$\begin{aligned} \text{pr}(Y_{i-2} = 0, Y_{i-1} = 1, Y_i = 0) &= \frac{cb}{2c + (1-b)}, \\ \text{pr}(Y_{i-2} = 0, Y_{i-1} = 1, Y_i = 1) &= \frac{(1-b)c}{2c + (1-b)}, \\ \text{pr}(Y_{i-2} = 1, Y_{i-1} = 0, Y_i = 1) &= \frac{c}{2c + (1-b)}, \\ \text{pr}(Y_{i-2} = 1, Y_{i-1} = 1, Y_i = 1) &= \frac{(1-b)(1-c)}{2c + (1-b)} \end{aligned}$$

and $\text{pr}(Y_{i-2} = 1, Y_{i-1} = 1, Y_i = 0) = \text{pr}(Y_{i-2} = 0, Y_{i-1} = 1, Y_i = 1)$. Since $\theta = (b, c)$, the tripletwise loglikelihood is

$$\begin{aligned} \log \text{TL}(b, c; y) = &-(N-2)\log(2c + 1 - b) + N_{010}\log(cb) + (N_{011} + N_{110})\log(c(1-b)) \\ &+ N_{101}\log c + N_{111}\log((1-b)(1-c)). \end{aligned}$$

Here the maximum tripletwise likelihood estimates are found to be $\widehat{b}_{MTL} = 0.6634$, $\widehat{c}_{MTL} = 0.3932$, which equal the maximum likelihood estimates. Graphical inspection from Figure 2 shows, as expected, that the ordinary loglikelihood has a more peaked form; indeed, the contour plots show different gradient directions. The maximal value of the tripletwise loglikelihood is $\log \text{TL}(\widehat{b}_{MTL}, \widehat{c}_{MTL}; y) = -451.5889$. It is worthwhile to emphasize that the maximum tripletwise likelihood estimates allow for the equality between the estimated and the observed frequencies for the five triplets of potential observations. Namely, we have that

$$\widehat{\text{pr}}(Y_{i-2} = 0, Y_{i-1} = 1, Y_i = 0) = \frac{\widehat{c}\widehat{b}}{2\widehat{c} + (1 - \widehat{b})} = \frac{N_{010}}{n-2}$$

and similarly for the remaining triplets. Then, we can say that this model reaches a sort of "best" possible fitting, as detected by the tripletwise likelihood.

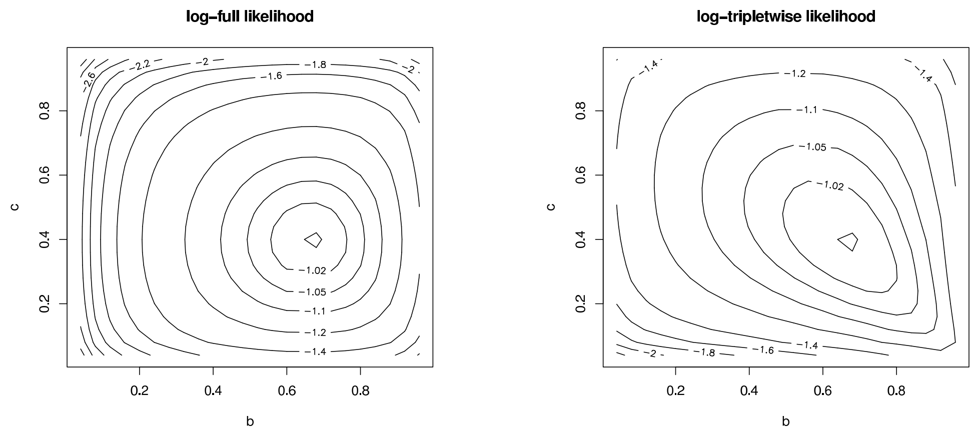

Figure 2: Contour plots of the ordinary loglikelihood (left) and of the tripletwise loglikelihood (right) for the second order Markov chains model fitted to the Old Faithful geyser data.

The second model is a two-states hidden Markov model. The hidden process $\{X_i\}_{i\geq 1}$ is a Markov chain with one-step transition probabilities

$$\Gamma = \begin{pmatrix} 0 & 1 \\ a & 1-a \end{pmatrix},$$

with $a \in (0,1)$ unknown. The conditional probabilities for the observations given the latent variables are

$$\text{pr}(Y_i = y|X_i = 0) = \rho^y(1-\rho)^{1-y}, \quad y = 0,1,$$
$$\text{pr}(Y_1 = 1|X_i = 1) = 1,$$

for $i \geq 1$, with $\rho \in (0,1)$ an unknown parameter. The relevant triplet probabilities are, for $i > 2$,

$$\text{pr}(Y_{i-2} = 0, Y_{i-1} = 1, Y_i = 0) = \frac{(1-\rho)^2 a^2}{1+a},$$
$$\text{pr}(Y_{i-2} = 0, Y_{i-1} = 1, Y_i = 1) = \frac{\rho(1-\rho)a^2 + (1-\rho)(1-a)a}{1+a},$$
$$\text{pr}(Y_{i-2} = 1, Y_{i-1} = 0, Y_i = 1) = \frac{(1-\rho)a}{1+a},$$
$$\text{pr}(Y_{i-2} = 1, Y_{i-1} = 1, Y_i = 1) = \frac{\rho^2 a^2 + 2\rho(1-a)a + \rho a + (1-a)^2}{1+a}$$

and $\text{pr}(Y_{i-2} = 1, Y_{i-1} = 1, Y_i = 0) = \text{pr}(Y_{i-2} = 0, Y_{i-1} = 1, Y_i = 1)$. Here $\theta = (a, \rho)$ and the maximum tripletwise likelihood estimates are $\widehat{a}_{MTL} = 0.8948, \widehat{\rho}_{MTL} = 0.2584$, while the maximum likelihood estimates, thought not the same, are very closed; that is, $\widehat{a}_{ML} = 0.827, \widehat{\rho}_{ML} = 0.225$. Graphical inspection, from Figure 3, of the contour plots of the ordinary loglikelihood and of the tripletwise loglikelihood leads to conclusions similar to those emphasized in the previous case.

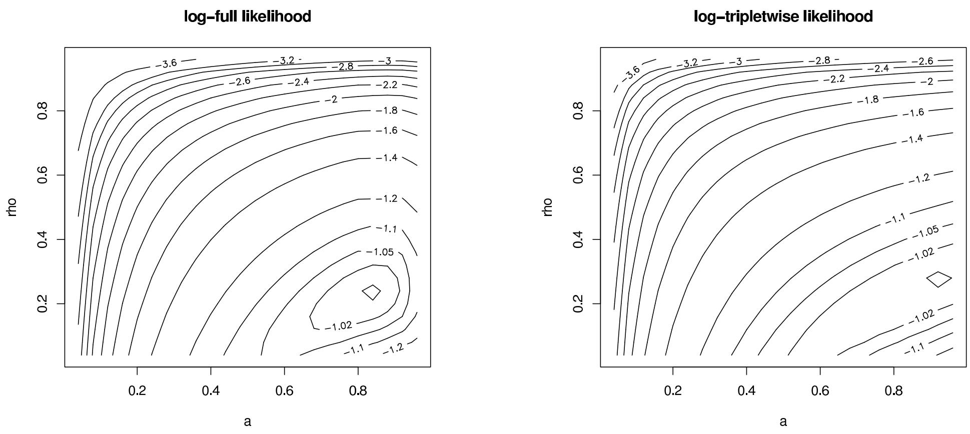

Figure 3: Contour plots of the ordinary loglikelihood (left) and of the tripletwise loglikelihood (right) for the hidden Markov model fitted to the Old Faithful geyser data.

Moreover, we find that the maximal value of the tripletwise loglikelihood, which corresponds to $\log \mathrm{TL}(\widehat{a}_{MTL}, \widehat{\rho}_{MTL}; y) = -451.5889$, coincides with that one obtained for the second-order Markov chain. This is a consequence of the perfect match between the estimated and the observed frequencies for the five triplets of potential observations, which holds for the two-states hidden Markov model as well. Indeed, also the computation, using Monte Carlo simulation, of the bias correction term $\mathrm{tr}\{\widehat{J}(Y)\widehat{H}(Y)^{-1}\}$ specifying the CLIC, gives the same approximated value 4.65 for the two models. Thus, in this case, the tripletwise likelihood, though useful for inferential purposes, does not discriminate between these two models. The potential drawback of this composite likelihood is related to the fact that the triplet-based likelihood-type objects involved in the computation do not detect the high-order structural differences between the estimated second-order Markov chain and the estimated two-states hidden Markov model. This conclusion emphasizes that a careful choice of the composite likelihood is necessary, both for inference and model selection, with the aim of balancing the improved computational facilities and the reduced descriptive ability, characterizing pseudolikelihood procedures.

A Appendix

We aim to justify the first order unbiasedness of the CLIC. Since the proof is very simple, and in fact similar to that considered for the TIC based on ordinary likelihood, we shall emphasize only some key issues, useful for generalizing the standard approach to the composite likelihood framework.

Recalling the notation and the definitions introduced in the previous section, we set the following assumptions.

A. The parametric space Θ is a compact subset of $\mathbb{R}^d$, $d \geq 1$, and, for every fixed $y \in \mathcal{Y}$, the composite likelihood function is two times differentiable with continuity, with respect to θ.

B. The maximum composite likelihood estimator $\widehat{\theta}_{\text{MCL}}(Y)$ is defined as a solution to the composite likelihood equation (1) and there exist a vector $\theta_* \in int(\Theta)$ such that, exactly or with an error term negligible as $n \to +\infty$,

$$E_{g(y)}\{\nabla \log \text{CL}_f(\theta_*; Y)\} = 0.$$

C. The maximum composite likelihood estimator $\widehat{\theta}_{\text{MCL}}(Y)$ is consistent, that is $\widehat{\theta}_{\text{MCL}}(Y) = \theta_* + o_p(1)$, and asymptotically normal distributed, as $n \to +\infty$, with a suitable asymptotic covariance matrix, as indicated in Section 2.

Note that the first two assumptions correspond to the basic regularity conditions for the asymptotic properties of maximum likelihood, and in general maximum composite likelihood, estimators under a model which could be misspecified for Y (White 1994). The vector θ_* is a *pseudo-true parameter value*, defined as a value in $int(\Theta)$ such that the composite Kullback-Leibler divergence between $g(y)$ and $f(y;\theta)$ is minimal. If the true distribution belong to the working family of distributions, the model is correctly specified for Y, namely $g(y) = f(y;\theta_0)$, for some $\theta_0 \in int(\Theta)$. In this particular case, θ_0 is the true parameter value. With regard to the third assumption, in order to prove the asymptotic normality of $\widehat{\theta}_{\text{MCL}}(Y)$, we usually require that

$$\nabla^2 \log \text{CL}(\theta_*; Y) = E_{g(y)}\{\nabla^2 \log \text{CL}(\theta_*; Y)\} + o_p(n).$$

Then, the following results hold.

Lemma 1. *Under the assumptions A-C, we have that*

$$\varphi(g, f) = E_{g(y)}\{\log \text{CL}(\theta_*; Y)\} + \frac{1}{2}\,\text{tr}\{J(\theta_*)H(\theta_*)^{-1}\} + o(1).$$

Lemma 2. *Under the assumptions A-C, we have that*

$$E_{g(y)}\{\Psi(Y; f)\} = E_{g(y)}\{\log \text{CL}(\theta_*; Y)\} - \frac{1}{2}\,\text{tr}\{J(\theta_*)H(\theta_*)^{-1}\} + o(1).$$

The proof is omitted, since, under the above assumptions, it is a natural generalization to composite likelihood of that one considered for the ordinary likelihood setting (Burnham & Anderson 2002, §7.2). Using the previous lemmas, we can easily prove that the selection statistic (3), defining the CLIC, is a first order unbiased estimator for the expected predictive composite loglikelihood (2).

References

Akaike, H. (1973), Information theory and extension of the maximum likelihood principle, *in* N. Petron & F. Caski, eds, 'Second Symposium on Information Theory', Budapest: Akademiai Kiado, pp. 267–281.

Azzalini, A. & Bowman, A. W. (1990), 'A look at some data on the old faithful geyser', *Applied Statistics* **39**, 357–365.

Besag, J. E. (1974), 'Spatial interaction and the statistical analysis of lattice systems (with discussion)', *Journal of the Royal Statistical Society, Series B* **36**, 192–236.

Burnham, K. P. & Anderson, D. R. (2002), *Model selection and multimodel inference: a practical information-theoretic approach*, Springer-Verlag Inc.

Carlstein, E. (1986), 'The use of subseries values for estimating the variance of a general statistic from a stationary sequence', *The Annals of Statistics* **14**(3), 1171–1179.

Cox, D. R. & Reid, N. (2003), A note on pseudolikelihood constructed from marginal densities, Technical Report available at the URL `http://www.utstat.toronto.edu/reid/research.html`.

Durbin, J. & Koopman, S. J. (2001), *Time series analysis by state space methods*, Oxford University Press.

Fearnhead, P. & Donnelly, P. (2002), 'Approximate likelihood methods for estimating local recombination rates', *Journal of the Royal Statistical Society, Series B (Methodological)* **64**(4), 657–680.

Garcia-Soidan, P. H. & Hall, P. (1997), 'On sample reuse methods for spatial data', *Biometrics* **53**(1), 273–281.

Hall, P. & Jing, B. (1996), 'On sample reuse methods for dependent data', *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **58**(4), 727–737.

Heagerty, P. J. & Lele, S. R. (1998), 'A composite likelihood approach to binary spatial data', *Journal of the American Statistical Association* **93**, 1099–1111.

Heagerty, P. J. & Lumley, T. (2000), 'Window subsampling of estimating functions with application to regression models', *Journal of the American Statistical Association* **95**, 197–211.

Henderson, R. & Shimakura, S. (2003), 'A serially correlated gamma frailty model for longitudinal count data', *Biometrika* **90**(2), 355–366.

Heyde, C. C. (1997), *Quasi-likelihood and its Application*, Springer Verlag, New York.

Hjort, N. L. & Omre, H. (1994), 'Topics in spatial statistics', *The Scandinavian Journal of Statistics* **21**, 289–357.

Leroux, B. G. (1992), 'Maximum-likelihood estimation for hidden Markov models', *Stochastic Processes and their Applications* **40**, 127–143.

Lindsay, B. (1988), Composite likelihood methods, *in* N. U. Prabhu, ed., 'Statistical Inference from Stochastic Processes', Providence RI: American Mathematical Society.

Lindsey, J. K. (1999), 'On the use of corrections for overdispersion', *Applied Statistics* **48**, 553–561.

Lumley, T. & Heagerty, P. (1999), 'Weighted empirical adaptive variance estimators for correlated data regression', *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **61**(2), 459–477.

MacDonald, I. L. & Zucchini, W. (1997), *Hidden Markov and Other Models for Discrete-valued Time Series*, Chapman &*amp*; Hall.

McCullagh, P. & Nelder, J. A. (1989), *Generalized Linear Models*, 2nd edn, Chapman and Hall, London.

Nott, D. J. & Rydén, T. (1999), 'Pairwise likelihood methods for inference in image models', *Biometrika* **86**(3), 661–676.

Parner, E. T. (2001), 'A composite likelihood approach to multivariate survival data', *The Scandinavian Journal of Statistics* **28**, 295–302.

Renard, D. (2002), Topics in Modelling Multilevel and Longitudinal Data, PhD thesis, Limburgs Universitair Centrum, Holland.

Renard, D., Molenberghs, G. & Geys, H. (2004), 'A pairwise likelihood approach to estimation in multilevel probit models', *Journal of Computtational Statistics & Data Analysis* **44**, 649–667.

Schwarz, G. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**, 461–464.

Shibata, R. (1989), Statistical aspects of model selection, *in* J. Willems, ed., 'From Data to Model', New York: Springer-Verlag, pp. 215–240.

Takeuchi, K. (1976), 'Distribution of information statistics and criteria for adequacy of models (in japanese)', *Mathematical Science* **153**, 12–18.

Vecchia, A. V. (1988), 'Estimation and model identification for continuos spatial processes', *Journal of the Royal Statistical Society, Series B (Methodological)* **50**(2), 297–312.

West, M. & Harrison, J. (1997), *Bayesian forecasting and dynamic models*, Springer-Verlag Inc.

White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press, New York.

Acknowledgements

The authors would like to thank Professor A. Azzalini for helpful comments and Professor M. Chiogna for suggesting the application on testing the overdispersion in dynamic generalised linear models. This research is partially supported by MIUR (Italy) grant 2001132337: "Model selection: frequentist criteria based on predictive densities and the likelihood function" and by MIUR (Italy) grant 2002134337: "Statistics as an aid for environmental decisions: identification, monitoring and evaluation".

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it