

On specification tests for composite likelihood inference

BY JING HUANG

*Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania,
Philadelphia, Pennsylvania 19104, U.S.A.*

jing14@upenn.edu

5

YANG NING

Department of Statistics and Data Science, Cornell University, Ithaca, New York 14853, U.S.A.

yn265@cornell.edu

NANCY REID

Department of Statistical Sciences, University of Toronto, Toronto M5S 3G5, Canada

reid@utstat.utoronto.ca

10

AND YONG CHEN

*Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania,
Philadelphia, Pennsylvania 19104, U.S.A.*

ychen123@upenn.edu

15

SUMMARY

Composite likelihood functions are often used for inference in applications where the data have complex structure. While inference based on the composite likelihood can be more robust than inference based on the full likelihood, the inference is not valid if the associated conditional or marginal models are misspecified. In this paper, we propose a general class of specification tests for composite likelihood inference. The test statistics are motivated by the fact that the second Bartlett identity holds for each component of the composite likelihood function when these components are correctly specified. We construct the test statistics based on the discrepancy between the so-called composite information matrix and the sensitivity matrix. As an illustration, we study three important cases of the proposed tests and establish their limiting distributions under both null and local alternative hypotheses. Finally, we evaluate the finite sample performance of the proposed tests in several examples.

Some key words: Bartlett identity; information matrix; misspecification tests; model specification.

1. INTRODUCTION

The composite likelihood function (Besag, 1974; Lindsay, 1988; Cox & Reid, 2004) is an inference function constructed as the product of a set of conditional or marginal density functions, whether or not the events defined in each component are mutually independent. It has been used in longitudinal studies (Molenberghs & Verbeke, 2005; Chandler & Bate, 2007), for the analysis of panel data (Wellner & Zhang, 2007), with spatial modelling (Heagerty & Lele, 1998; Guan, 2006), with missing data (He & Yi, 2011), with graphical models (Xue et al., 2012; Chen

et al., 2015; Yang et al., 2015), with change point detection in multivariate time series (Ma & Yau, 2016), among many other settings. One of the main reasons for its widespread application is that only part of data generating mechanism needs to be specified, which can reduce the computational cost and provide some robustness to model misspecification. For a comprehensive discussion of composite likelihood and its applications, see the review paper by Varin et al. (2011).

Although inference based on the composite likelihood has been well developed (Lindsay, 1988; Kent, 1982; Pagui et al., 2014), a corresponding specification test does not seem to exist. Aside from technical difficulties, on which we will elaborate later, the lack of development in this area may be due to the perception that composite likelihood inference is robust to partial model misspecification (Xu & Reid, 2011). However, composite likelihood inference still requires the specification of a set of lower dimensional conditional or marginal models, and the corresponding results can be misleading, if these are misspecified. For instance, Ogden (2016) showed that the maximum composite likelihood estimator is inconsistent in a generalized linear mixed model with a misspecified random effect distribution, and the misspecification bias of the maximum composite likelihood estimator is significantly larger than the misspecification bias in the maximum likelihood estimator.

In this paper, we propose a class of specification tests for composite likelihood functions. A general test for model specification with ordinary likelihood functions is the information matrix test proposed by White (1982). This relies on the second Bartlett identity, which holds when the model is correctly specified: it compares the variability matrix to the sensitivity matrix, both defined in §2 below. A Wald type test statistic is constructed from the difference between these two matrices. Presnell & Boos (2004) proposed an “in-and-out sample” likelihood ratio test, which is asymptotically equivalent to a multiplicative contrast between the sensitivity matrix and the variability matrix. More recently, Zhou et al. (2012) developed an information ratio test which can effectively test the specifications of variance and covariance functions in generalized estimating equations (Liang & Zeger, 1986).

These specification tests rely on the fact that the second Bartlett identity holds under the correct model. For a composite likelihood, the second Bartlett identity does not hold, even when the component densities are correctly specified (Varin et al., 2011). Thus, direct application of the information matrix based tests is invalid. To circumvent this difficulty, we define the composite information matrix, and show that this can recover a counterpart of the second Bartlett identity. We propose a general class of specification tests for composite likelihood inference based on the discrepancy between the composite information matrix and the sensitivity matrix. We illustrate the usefulness of the proposed class of tests with three important special cases: the composite information matrix test, composite information ratio test and composite information max test. We establish their asymptotic distributions under the null hypothesis, and asymptotic power under local alternatives.

2. SPECIFICATION TESTS FOR COMPOSITE LIKELIHOOD FUNCTIONS

2.1. Composite information matrix and the Bartlett identity

Let $f(y; \theta)$ be the joint probability density function of a multidimensional random vector Y , indexed by a p -dimensional parameter $\theta = (\theta_1, \dots, \theta_p)$ in the parameter space Ω . Let $\{\mathcal{A}_1(y), \dots, \mathcal{A}_K(y)\}$ denote a sequence of sets. The log probability of the event $\{Y \in \mathcal{A}_k(y)\}$ is $\ell_k(\theta; y) = \log \int_{u \in \mathcal{A}_k(y)} f(u; \theta) du$, where $k = 1, 2, \dots, K$ and K is the total number of events.

Assume that n independently and identically distributed random variables $Y_1, \dots, Y_n$ are observed from the model $f(y; \theta)$. A composite log likelihood function based on this sequence of events is

80

$$c\ell(\theta) = n^{-1} \sum_{i=1}^n \sum_{k=1}^K \omega_{ik} \ell_k(\theta; y_i),$$

where ω_{ik} is a nonnegative weight associated with the event $\{Y_i \in \mathcal{A}_k(y_i)\}$. The variability and sensitivity matrices associated with $c\ell(\theta)$ are

$$J(\theta) = E \left[\left\{ \sum_{k=1}^K \omega_{ik} \frac{\partial}{\partial \theta} \ell_k(\theta; y_i) \right\}^{\otimes 2} \right], \quad H(\theta) = -E \left\{ \sum_{k=1}^K \omega_{ik} \frac{\partial^2}{\partial \theta \partial \theta^T} \ell_k(\theta; y_i) \right\}.$$

Since defining a composite likelihood only requires models for the events $\{Y \in \mathcal{A}_k(y)\}$ for $1 \leq k \leq K$, specification tests should be tailored to the modeling of $\{Y \in \mathcal{A}_k(y)\}$. For instance, in the analysis of clustered data, the independence likelihood in Chandler & Bate (2007) only involves modeling the conditional mean function. Thus, the corresponding specification tests should be to verify the validity of the conditional mean model rather than, for example, the correlation structure of the full model.

85

Formally, the null hypothesis for a specification test of the composite likelihood can be formulated as

$$H_0 : \text{there exists a } \theta \in \Omega, \text{ such that } \text{pr}\{Y \in \mathcal{A}_k(y)\} \propto \exp\{\ell_k(\theta; y)\}, \quad \text{for all } k = 1, \dots, K.$$

As noted in §1, the second Bartlett identity, which is the basis for all existing information based specification tests, does not hold for the composite likelihood even when the assumed model is correct. This statement is illustrated by an example of multivariate normal model in the Supplementary Material. To develop specification tests, we construct a new identity by defining the composite information matrix $I_c(\theta)$, as follows

$$I_c(\theta) = E \left[\sum_{k=1}^K \omega_{ik} \left\{ \frac{\partial}{\partial \theta} \ell_k(\theta; y_i) \right\}^{\otimes 2} \right].$$

When each log likelihood function $\ell_k(\theta; y)$ is correctly specified, under mild regularity conditions, we have

90

$$I_c(\theta) - H(\theta) = 0. \quad (1)$$

Based on this identity, we propose a general class of specification tests for the composite likelihood.

2.2. A general class of specification tests for the composite likelihood

Define the maximum composite likelihood estimator as $\hat{\theta}_n = \text{argmax}_{\theta \in \Omega} c\ell(\theta)$, and the observed composite information matrix and sensitivity matrix as

$$\hat{I}_c(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \omega_{ik} \left\{ \frac{\partial}{\partial \theta} \ell_k(\hat{\theta}_n; y_i) \right\}^{\otimes 2}, \quad \hat{H}(\hat{\theta}_n) = -\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \omega_{ik} \frac{\partial^2}{\partial \theta \partial \theta^T} \ell_k(\hat{\theta}_n; y_i).$$

Our family of specification tests is constructed by comparing $\hat{I}_c(\hat{\theta}_n)$ and $\hat{H}(\hat{\theta}_n)$. Specifically, we consider a discrepancy function $d : \mathbb{R}^{p \times p} \times \mathbb{R}^{p \times p} \rightarrow [0, \infty)$, which satisfies $d(M_1, M_2) \geq 0$

95

and $d(M_1, M_1) = 0$ for any $M_1, M_2 \in \mathbb{R}^{p \times p}$. It is seen that d is a premetric, more general than a standard metric in the sense that we do not require the symmetry or subadditivity of d . As will be seen later, some special cases of d are indeed only a premetric. Given a discrepancy function d , we consider the corresponding test statistic

$$Q = d\{\hat{I}_c(\hat{\theta}_n), \hat{H}(\hat{\theta}_n)\}, \quad (2)$$

which measures the distance between the observed composite information matrix and sensitivity matrix. Under H_0 , equation (1) holds and therefore $d\{I_c(\theta), H(\theta)\} = 0$. Under mild regularity conditions, the composite information matrix and the sensitivity matrix $I_c(\theta)$ and $H(\theta)$ can be consistently estimated by $\hat{I}_c(\hat{\theta}_n)$ and $\hat{H}(\hat{\theta}_n)$ respectively. Thus, we expect Q to be small when the composite likelihood is correctly specified. On the other hand, a large value of Q implies the potential misspecification of the composite likelihood. In the following, we consider three special cases of d and study their asymptotic null distributions and local asymptotic powers.

2.3. Special case 1: Composite information matrix test

In this subsection, we take d as a scaled L_2 metric, which leads to a composite information matrix test. Since $\hat{I}_c(\hat{\theta}_n)$ and $\hat{H}(\hat{\theta}_n)$ are symmetric matrices, we only need to consider their upper triangular entries when constructing Q . That is

$$d(M_1, M_2) = (m_1 - m_2)^T W (m_1 - m_2), \quad (3)$$

where m_1 and m_2 are $p(p+1)/2$ dimensional vectors obtained by vectorizing the upper triangular elements of the matrices M_1 and M_2 , and W is a positive definite matrix used to standardize the test statistic. In this case, d is indeed a metric. When $\ell(\theta)$ corresponds to the log likelihood of the full model, the test statistic Q in (2) with the metric d defined in (3) reduces to the information matrix test in White (1982).

Formally, consider the $p(p+1)/2$ upper triangular elements of $\hat{I}_c(\hat{\theta}_n) - \hat{H}(\hat{\theta}_n)$, and denote the contribution of the i th observation to the (j, t) element ($1 \leq j \leq t \leq p$) of the matrix $\hat{I}_c(\theta) - \hat{H}(\theta)$ by

$$e_{jt}(y_i, \theta) = \sum_{k=1}^K \omega_{ik} \left\{ \frac{\partial}{\partial \theta_j} \ell_k(\theta; y_i) \frac{\partial}{\partial \theta_t} \ell_k(\theta; y_i) + \frac{\partial^2}{\partial \theta_j \partial \theta_t} \ell_k(\theta; y_i) \right\},$$

where θ_j and θ_t are the j -th and t -th elements of θ . With $q = p(p+1)/2$, we stack the q elements $e_{jt}(y_i, \theta)$ for $1 \leq j \leq t \leq p$ to obtain a $q \times 1$ vector $e(y_i, \theta)$. Define $T_M(\theta) = n^{-1} \sum_{i=1}^n e(y_i, \theta)$ and

$$V_M(\theta) = E \left[e(y_i, \theta) + E \left\{ \frac{\partial}{\partial \theta} e(y_i, \theta) \right\} H^{-1}(\theta) \sum_{k=1}^K \omega_{ik} \frac{\partial}{\partial \theta} \ell_k(\theta; y_i) \right]^{\otimes 2}.$$

By Taylor expansion and the properties of influence functions, $V_M(\theta)$ is the asymptotic variance matrix of $n^{1/2}T_M(\hat{\theta}_n)$ under H_0 , and it can be consistently estimated by

$$\hat{V}_M = n^{-1} \sum_{i=1}^n \left\{ e(y_i, \hat{\theta}_n) + \frac{\partial}{\partial \theta} T_M(\hat{\theta}_n) \hat{H}^{-1}(\hat{\theta}_n) \sum_{k=1}^K \omega_{ik} \frac{\partial}{\partial \theta} \ell_k(\hat{\theta}_n; y_i) \right\}^{\otimes 2}.$$

The proposed composite information matrix test Q^{matrix} is defined as

$$Q^{\text{matrix}} = n T_M^T(\hat{\theta}_n) \hat{V}_M^{-1} T_M(\hat{\theta}_n),$$

where we set $W = \hat{V}_M^{-1}$ in (3) to construct an asymptotic χ^2 test of Wald type. Throughout the paper, we consider the scenario that the sample size n increases with the total number of events K fixed. The following theorem states the asymptotic distribution of Q^{matrix} under the null hypothesis. 125

THEOREM 1. *Given regularity conditions R1–R7 in the Supplementary Material, as n increases the composite information matrix test statistic Q^{matrix} converges in distribution to χ_q^2 under H_0 .* 130

The above theorem suggests the use of $(1 - \alpha)$ quantile of the χ_q^2 distribution as the cutoff value for the test Q^{matrix} . Since the degrees of freedom q increase as a quadratic function of the number of parameters in the composite likelihood and the composite information matrix test aims to pick up a wide range of deviations from the null, this test may have low power when p is relatively large. This phenomenon is also confirmed in our simulation studies. In the following subsections, we propose alternative choices of the discrepancy function d , which attempts to address this issue. 135

2.4. Special case 2: Composite information ratio test

Instead of using the L_2 metric in (3), we consider a discrepancy function which takes a ratio contrast between two matrices, defined as 140

$$d(M_1, M_2) = \{\text{tr}(M_1 M_2^{-1}) - p\}^2 / W, \quad (4)$$

where W is a positive number used to standardize the test statistic. It can be shown that d in (4) is a premetric rather than a metric. Motivated by (4), we can construct a parsimonious summary statistic by comparing the trace of $\hat{I}_c(\hat{\theta}_n) \hat{H}^{-1}(\hat{\theta}_n)$ to p . This type of test was proposed by Zhou et al. (2012) for the quasi-likelihood function with application to the selection of the covariance structure in generalized estimating equations. Unlike quasi-likelihood where the second Bartlett identity holds, the construction of this test is built on the identity in (1). Inspired by Zhou et al. (2012), we call this the composite information ratio test. 145

The test statistic is

$$Q^{\text{ratio}} = \left[n^{1/2} \left\{ T_R(\hat{\theta}_n) - p \right\} \right]^2 / \hat{V}_R,$$

where $T_R(\hat{\theta}_n) = \text{tr} \left\{ \hat{I}_c(\hat{\theta}_n) \hat{H}^{-1}(\hat{\theta}_n) \right\}$, and $\hat{V}_R$ is the estimated asymptotic variance of $T_R(\hat{\theta}_n)$. In the Supplementary Material, we derive the estimator 150

$$\begin{aligned} \hat{V}_R = n^{-1} \sum_{i=1}^n & \left[\sum_{k=1}^K \omega_{ik} \left\{ \frac{\partial}{\partial \theta} \ell_k(\hat{\theta}_n; y_i) \right\}^T \hat{H}^{-1}(\hat{\theta}_n) \frac{\partial}{\partial \theta} \ell_k(\hat{\theta}_n; y_i) \right. \\ & + \frac{\partial}{\partial \theta} T_R(\hat{\theta}_n) \hat{H}^{-1}(\hat{\theta}_n) \sum_{k=1}^K \omega_{ik} \frac{\partial}{\partial \theta} \ell_k(\hat{\theta}_n; y_i) \\ & \left. + \text{tr} \left\{ \sum_{k=1}^K \omega_{ik} \frac{\partial^2}{\partial \theta \partial \theta^T} \ell_k(\hat{\theta}_n; y_i) \hat{H}^{-1}(\hat{\theta}_n) \right\} \right]^2. \end{aligned}$$

THEOREM 2. *Given regularity conditions R1–R7 in the Supplementary Material, the composite information ratio test statistic Q^{ratio} converges in distribution to χ_1^2 under H_0 .* 155

In the simulation studies, the composite information ratio test tends to yield more accurate Type I errors than the composite information matrix test.

2.5. *Special case 3: Composite information max test*

In this subsection, we propose another test by considering the maximum discrepancy of the diagonal elements of the linear contrast between the observed composite information matrix and sensitivity matrix. We call this test the composite information max test. This corresponds to the general test statistic Q in (2) with

$$d(M_1, M_2) = \|W(m_1^* - m_2^*)\|_{\max}, \quad (5)$$

where W is a $p \times p$ positive definite matrix used to standardize the test statistic, and $m_1^* = \{(M_1)_{11}, \dots, (M_1)_{pp}\}^T$ is a $p \times 1$ vector.

Denote the diagonal elements of the linear contrast, contributed by the i th observation, by $e_{jj}(y_i, \theta)$, $j = 1, \dots, p$, and let $e^*(y_i, \theta) = \{e_{11}(y_i, \theta), \dots, e_{pp}(y_i, \theta)\}^T$. Define $T_M^*(\hat{\theta}_n) = n^{-1} \sum_{i=1}^n e^*(y_i, \hat{\theta}_n)$. Denote by $V_M^*(\theta)$ the asymptotic variance matrix of $n^{1/2}T_M^*(\hat{\theta}_n)$ and by $\hat{V}_M^*$ its empirical counterpart. Define $S = n^{1/2}\hat{U}_M^*T_M^*(\hat{\theta}_n)$, where S_j is the j th element of S , $j = 1, \dots, p$, and $\hat{U}_M^*$ is a $p \times p$ matrix such that $\hat{U}_M^{*T}\hat{U}_M^* = \hat{V}_M^{*-1}$. Our asymptotic results hold for any choice of $\hat{U}_M^*$. The composite information max test is given by

$$Q^{\max} = \max_{1 \leq j \leq p} |S_j|.$$

The following theorem establishes the asymptotic distribution of composite information max test under the null hypothesis.

THEOREM 3. *Given regularity conditions R1–R7 in the Supplementary Material, we have*

$$Q^{\max} \rightarrow \max_{1 \leq j \leq p} |N_j|,$$

in distribution under H_0 , where $N_1, \dots, N_p$ are independent and identically distributed standard normal variables. In other words, for any $t > 0$, $\lim_{n \rightarrow \infty} |\text{pr}(Q^{\max} \leq t) - F(t)^p| = 0$, where $F(\cdot)$ is the cumulative distribution function of the folded standard normal distribution.

By Theorem 3, we reject the null hypothesis at level α if $Q^{\max} > F^{-1}\{(1 - \alpha)^{1/p}\}$, where $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$ defined in Theorem 3, as N_j and N_k are independent standard normal random variables for any $j \neq k$. In the construction of S , we transform $T_M^*(\hat{\theta}_n)$ by the matrix U_M^* so that the elements of S are de-correlated.

The main difference among $Q^{\max}$, Q^{matrix} and Q^{ratio} lies in how the individual statistics constructed from some components of the matrix equation (1) are combined. Specifically, the test statistic $Q^{\max}$ combines individual statistics by searching for their maximum across the diagonal, whereas Q^{matrix} and Q^{ratio} exploit different quadratic forms of individual statistics. In the Supplementary Material, we investigate the asymptotic local power of the three composite information tests under a sequence of locally misspecified models (Copas & Eguchi, 2005).

3. EXAMPLE

3.1. *Extended Mantel-Haenszel method for stratified case-control studies*

The Mantel-Haenszel estimator is widely used to estimate the common odds ratio in a series of 2×2 tables in epidemiological studies. Liang (1987) extended the Mantel-Haenszel approach to logistic regression models for stratified case-control studies, to allow simultaneous estimation of effect sizes of multivariate risk factors.

Consider a stratified case-control study, where $x_{i1}, \dots, x_{id_i}$ denote the $p \times 1$ vectors of potential risk factors of d_i cases, and let $x_{id_i+1}, \dots, x_{ih_i}$ denote the potential risk factors of m_i controls

in the i th stratum with $m_i = h_i - d_i$, and $i = 1, \dots, n$. A logistic regression model allowing for stratum-specific effects is

$$\text{logit pr}(y_{ij} = 1|x_{ij}) = \alpha_i + \beta^T x_{ij}, \quad i = 1, \dots, n, \quad (6)$$

where the coefficient β quantifies the effects of risk factors x_{ij} on the disease status y_{ij} .

190

Liang (1987) proposed a composite likelihood method where the nuisance parameters α_i 's are eliminated by conditioning. Specifically, for the (j, l) case-control pair of subjects in the i th stratum ($j = 1, \dots, d_i; l = d_i + 1, \dots, h_i$), the conditional probability that x_{ij} is from the case given that x_{ij} and x_{il} are observed is

$$\text{pr}(y_{ij} = 1, y_{il} = 0 | y_{ij} + y_{il} = 1, x_{ij}, x_{il}; \alpha_i, \beta) = \frac{e^{\beta^T x_{ij}}}{e^{\beta^T x_{ij}} + e^{\beta^T x_{il}}}.$$

Thus a composite likelihood can be formulated by considering all $d_i m_i$ pairs in the i th stratum

$$L_i(\beta) = \prod_{j=1}^{d_i} \prod_{l=d_i+1}^{h_i} \frac{e^{\beta^T x_{ij}}}{e^{\beta^T x_{ij}} + e^{\beta^T x_{il}}}.$$

A weighted composite log likelihood function for β using the combining data from all n strata for β is then constructed as

$$c\ell(\beta) = n^{-1} \sum_{i=1}^n \omega_i \log \{L_i(\beta)\} = n^{-1} \sum_{i=1}^n \omega_i \sum_{j=1}^{d_i} \sum_{l=d_i+1}^{h_i} \log \left(\frac{e^{\beta^T x_{ij}}}{e^{\beta^T x_{ij}} + e^{\beta^T x_{il}}} \right).$$

We consider the weights $\omega_i = h_i^{-1}$; with this choice the maximum composite likelihood estimator reduces to the Mantel-Haenszel estimator when only a binary covariate is considered (Liang, 1987). The composite information matrix and the sensitivity matrix are

$$I_c(\beta) = E \left[h_i^{-1} \sum_{j=1}^{d_i} \sum_{l=d_i+1}^{h_i} \left\{ x_{ij} - \frac{x_{ij} \exp(\beta^T x_{ij}) + x_{il} \exp(\beta^T x_{il})}{\exp(\beta^T x_{ij}) + \exp(\beta^T x_{il})} \right\}^{\otimes 2} \right],$$

$$H(\beta) = E \left[h_i^{-1} \sum_{j=1}^{d_i} \sum_{l=d_i+1}^{h_i} - \frac{\{x_{ij} \exp(\beta^T x_{ij}) + x_{il} \exp(\beta^T x_{il})\}^{\otimes 2}}{\{\exp(\beta^T x_{ij}) + \exp(\beta^T x_{il})\}^2} \right. \\ \left. + \frac{\{x_{ij} x_{ij}^T \exp(\beta^T x_{ij}) + x_{il} x_{il}^T \exp(\beta^T x_{il})\}}{\exp(\beta^T x_{ij}) + \exp(\beta^T x_{il})} \right].$$

The proposed composite information tests can be constructed to test the specification of the model. The empirical performance of these tests is evaluated in §4.

195

3.2. Undirected graphical model

The undirected graphical model has been widely used to describe the dependence structures of multivariate variables. For simplicity, most literature only assumes the pairwise interaction in the joint distribution of the multivariate variables (Chen et al., 2015; Yang et al., 2015). Assume that there are p random variables $Y_1, \dots, Y_p$ represented as nodes of the graph $G = (V, E)$, with the vertex set $V = \{1, \dots, p\}$ and the edge set $E \subseteq V \times V$. An edge between vertexes i and j indicates that Y_i and Y_j are conditionally dependent given the remaining variables. The joint

200

distribution of a pairwise graphical model corresponding to the graph $G = (V, E)$ takes the form

$$\text{pr}(y) = \exp \left\{ \sum_{s \in V} f_s(y_s; \alpha_s) + 1/2 \sum_{(s,t) \in E} \theta_{st} y_s y_t - \Phi(\Theta, \alpha) \right\}, \quad (7)$$

where $y = (y_1, \dots, y_p)^T$, $\Theta = (\theta_{st})_{p \times p}$ is a symmetric square matrix of parameters associated with edges, with the diagonal elements equal zero, $\alpha = (\alpha_1, \dots, \alpha_p)$ is a matrix of parameters with the s -th column α_s involved in the node potential function $f_s(y_s; \alpha_s)$, and $\Phi(\Theta, \alpha)$ is the log partition function. In many examples of exponential family graphical models, the log partition function is intractable. To circumvent this challenge, a composite likelihood function constructed by combining conditional likelihoods has been proposed by Xue et al. (2012), Chen et al. (2015), Yang et al. (2015), among others. Specifically, denote $y_{-s} = (y_1, \dots, y_{s-1}, y_{s+1}, \dots, y_p)^T$. The conditional distribution is $\text{pr}(y_s | y_{-s}) = \exp \left\{ f_s(y_s; \alpha_s) + \sum_{t \neq s} \theta_{st} y_s y_t - D_s(\eta_s) \right\}$, where η_s is a function of α_s, y_{-s} , and Θ_s , Θ_s is the s -th column of Θ without the diagonal element, and $D_s(\cdot)$ is a function whose form depends on the conditional distribution.

As a concrete example, consider the following Ising graphical model. Let Y be a vector of binary variables, taking values $\{-1, 1\}$. Assume that Y satisfies the joint distribution (7) with $f_s(y_s; \alpha_s) = \alpha_{1s} y_s$. Given n independent and identically distributed copies $Y^{(i)}$ of Y , the composite log likelihood function is $\text{cl}(\Theta, \alpha) = n^{-1} \sum_{i=1}^n \sum_{s=1}^p \ell_s^{(i)}(\Theta, \alpha)$, where

$$\ell_s^{(i)}(\Theta, \alpha) = \alpha_{1s} y_s^{(i)} + \sum_{t \neq s} \theta_{st} y_s^{(i)} y_t^{(i)} - D_s \left(\alpha_{1s} + \sum_{t \neq s} \theta_{st} y_t^{(i)} \right), \quad (8)$$

with $D_s(\eta) = \log\{\exp(\eta) + \exp(-\eta)\}$. The composite information and sensitivity matrices are $I_c(\Theta, \alpha) = \sum_{s=1}^p \text{E} \left\{ \partial \ell_s^{(i)}(\Theta, \alpha) / \partial(\Theta, \alpha) \right\}^{\otimes 2}$ and $H(\Theta, \alpha) = \sum_{s=1}^p \text{E} \left\{ -\partial^2 \ell_s^{(i)}(\Theta, \alpha) / \partial(\Theta, \alpha)^2 \right\}$. The proposed composite information tests can be used to test the specification of the composite likelihood for the Ising model.

4. SIMULATION

We consider the example in §3.1 with two scenarios: when the specified composite likelihood model correctly includes all the predictors, and when a predictor is falsely excluded.

We simulate data using equation (6), with three continuous predictors, x_1, x_2 , and x_3 . The first predictor x_1 is generated from uniform distribution from 0 to 5. The second and third predictors, x_2 and x_3 , are the quadratic and cubic terms of x_1 , i.e., $x_2 = x_1^2$ and $x_3 = x_1^3$, respectively. We set β_1 and β_2 at 2 and -0.5 , respectively, and let the magnitude of β_3 decrease from 0 to -0.5 to evaluate the size and power of the proposed tests. The stratum-specific intercepts α_i are generated from a uniform distribution from -2.5 to -2 . We calculate the individual probability of $Y_{ij} = 1$ using equation (6) and generate the binary response from the corresponding Bernoulli distribution. We randomly sample 5 cases ($y = 1$) and 5 controls ($y = 0$) from each stratum. Assume that only two predictors x_1 and x_2 are available and included in the composite likelihood method described in §3.1. We apply the proposed tests of model specification for the composite likelihood with only x_1 and x_2 . In practice, we cannot use the classical composite likelihood based tests for $\beta_3 = 0$, since we assume the value of x_3 is unobserved. For each of the scenarios, we consider $n = 200, 400$ and 600 strata.

Table 1 summarizes the empirical Type I error and power of the proposed tests at nominal levels of 5% and 10% based on 5,000 replicates. The composite information ratio test controls Type I error reasonably well at both nominal levels. The composite information max test is slightly liberal at the 5% nominal level when n is small but the Type I error tends to be more accurate as n increases. The composite information matrix test is relatively conservative which agrees with the literature on standard information matrix test.

As the magnitude of β_3 decreases from 0 to -0.5 , all three tests show increasing power. The composite information max test has largest power among the three tests. This gain in power may be due to the fact that it detects model misspecification by searching for the maximum discrepancy between the two matrices, and is more sensitive than the other two tests, which average the discrepancies. In our simulations, the composite information ratio test has relatively less power than the other two tests.

In the Supplementary Material, we conduct additional simulation studies in four different scenarios, including the Ising model described in §3.2, the independence likelihood for correlated data and the pairwise likelihood for multivariate outcomes. We find that the empirical Type I error rates of the proposed composite information tests are effectively controlled below the nominal levels and all three tests have good power for detecting model misspecifications. The matrix and max tests perform similarly in all examples, and their empirical Type I error rates converge to the nominal level more slowly than the ratio test, especially in the Ising model example. We find that the power of the three tests depends on the underlying data generating process. The information ratio test has relatively higher statistical power in two of the four scenarios. We refer to the Supplementary Material for further discussion on the choice of the tests.

Table 1. Empirical rejection rates (%) in 5,000 simulations of three proposed tests for H_0 , under varying numbers of strata $n = 200, 400, 600$, and log odds ratios $\beta_1 = 2$, $\beta_2 = -0.5$, and β_3 varying from 0 to -0.5 .

n	β_3	Rejection (%) at $\alpha = 0.05$			Rejection (%) at $\alpha = 0.10$		
		Matrix test	Ratio test	Max test	Matrix test	Ratio test	Max test
200	0.00	3.6	4.9	7.8	5.8	9.7	11.4
	-0.10	19.2	16.6	27.9	25.6	24.1	35.5
	-0.25	29.2	24.2	39.8	37.3	32.0	47.6
	-0.50	33.6	27.1	43.8	41.7	35.9	51.4
400	0.00	3.4	4.6	6.9	5.3	9.5	10.3
	-0.10	29.8	21.5	41.3	38.1	30.3	50.0
	-0.25	47.5	32.3	58.9	56.2	41.1	67.0
	-0.50	53.9	37.5	63.8	61.9	46.9	71.2
600	0.00	4.1	5.2	6.4	6.5	10.2	9.9
	-0.10	43.4	28.6	55.4	52.2	37.7	63.8
	-0.25	62.9	40.5	73.6	71.3	50.5	80.9
	-0.50	69.5	47.2	78.9	77.0	57.1	84.8

5. DISCUSSION

In applications with time series, data are no longer independent and our results do not apply. Davis & Yau (2011) proposed a pairwise likelihood approach for a broad class of linear time series models. A future research direction is to generalize the proposed tests to evaluate the specification of a composite likelihood with dependent data.

We considered in this paper that the total number of events K was fixed and the sample size n increased. Although the constructions of the new composite information matrix and the new

identity in (1) do not require such an assumption, derivations of the limiting distributions of the three proposed tests do assume that the maximum composite likelihood estimator $\hat{\theta}_n$ is a consistent estimator of θ with $n^{1/2}$ convergence rate under the null hypothesis. Such an assumption may not hold in some scenarios that the number of components also increases as the number of independent replicates grows. For example, Cox & Reid (2004) discussed the situation where a small number of individually large sequences is available for pairwise likelihood, the maximum composite likelihood estimator may not be consistent or the convergence rate may be very slow as the number of components increases, depending on the internal correlation among the components. In such scenarios, the limiting distribution of the proposed test needs to be derived differently.

The validity of inference based on composite likelihoods usually requires both the correct specification of lower-dimensional conditional or marginal models and the existence of a joint model corresponding to the assumed composite likelihood. The latter is known as the compatibility of composite likelihoods; see Yi (2014) for further discussion. The proposed tests aim to investigate the specification of lower-dimensional models in composite likelihoods assuming the compatibility condition holds.

The information test (White, 1982) can be viewed as an omnibus test, whereas the proposed tests are tailored for the model specification required in the composite likelihood function. Thus, our tests can be more powerful than the existing information test based on the full likelihood, when the composite likelihood function is misspecified. If our tests indicate misspecification, it would be of interest to understand what modeling assumptions are violated. A future research direction is to design goodness-of-fit tests to detect specific departures from the assumed composite likelihood.

ACKNOWLEDGEMENT

The authors thank the referees, associate editor and editor for comments that substantially improved this work.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes regularity conditions, proof of Theorems 1-3, analysis of local asymptotic power and three additional simulation studies, including the Ising model described in §3.2, inference using independent likelihood for dependent data, and inference using pairwise likelihood for multivariate outcomes.

REFERENCES

- BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc. B* **36**, 192–239.
- CHANDLER, R. & BATE, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika* **94**, 167–183.
- CHEN, S., WITTEN, D. M. & ALI, S. (2015). Selection and estimation for mixed graphical models. *Biometrika* **102**, 47–64.
- COPAS, J. & EGUCHI, S. (2005). Local model uncertainty and incomplete-data bias (with discussion). *J. R. Statist. Soc. B* **67**, 459–513.
- COX, D. & REID, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91**, 729–737.
- DAVIS, R. A. & YAU, C. Y. (2011). Comments on pairwise likelihood in time series models. *Statist. Sinica* **21**, 255–277.

- GUAN, Y. (2006). A composite likelihood approach in fitting spatial point process models. *J. Am. Statist. Assoc.* **101**, 1502–1512.
- HE, W. & YI, G. (2011). A pairwise likelihood method for correlated binary data with/without missing observations under generalized partially linear single-index models. *Statist. Sinica* **21**, 207–229.
- HEAGERTY, P. & LELE, S. (1998). A composite likelihood approach to binary spatial data. *J. Am. Statist. Assoc.* **93**, 1099–1111. 315
- KENT, J. (1982). Robust properties of likelihood ratio tests. *Biometrika* **69**, 19–27.
- LIANG, K. (1987). Extended Mantel-Haenszel estimating procedure for multivariate logistic regression models. *Biometrics* **43**, 289–299.
- LIANG, K. & ZEGER, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22. 320
- LINDSAY, B. (1988). Composite likelihood methods. *Contemp. Math.* **80**, 221–39.
- MA, T. F. & YAU, C. Y. (2016). A pairwise likelihood-based approach for changepoint detection in multivariate time series models. *Biometrika* **103**, 409–421.
- MOLENBERGHS, G. & VERBEKE, G. (2005). *Models for discrete longitudinal data*. Springer.
- OGDEN, H. E. (2016). A caveat on the robustness of composite likelihood estimators: the case of a mis-specified random effect distribution. *Statist. Sinica* **26**, 639–651. 325
- PAGUI, K., CLOVIS, E., SALVAN, A. & SARTORI, N. (2014). Combined composite likelihood. *Can. J. Statist.* **42**, 525–543.
- PRESNELL, B. & BOOS, D. D. (2004). The IOS test for model misspecification. *J. Am. Statist. Assoc.* **99**, 216–227.
- VARIN, C., REID, N. & FIRTH, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21**, 5–42. 330
- WELLNER, J. & ZHANG, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Statist.* **35**, 2106–2142.
- WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–25.
- XU, X. & REID, N. (2011). On the robustness of maximum composite likelihood estimate. *J. Statist. Plann. Inference* **141**, 3047–3054. 335
- XUE, L., ZOU, H. & CAI, T. (2012). Nonconcave penalized composite conditional likelihood estimation of sparse Ising models. *Ann. Statist.* **40**, 1403–1429.
- YANG, E., RAVIKUMAR, P., ALLEN, G. I. & LIU, Z. (2015). Graphical models via univariate exponential family distributions. *J. Mach. Learn. Res.* **16**, 3813–3847.
- YI, G. Y. (2014). Composite likelihood/pseudolikelihood. *Wiley StatsRef: Statistics Reference Online*, 1–14. 340
- ZHOU, Q. M., SONG, P. X.-K. & THOMPSON, M. E. (2012). Information ratio test for model misspecification in quasi-likelihood inference. *J. Am. Statist. Assoc.* **107**, 205–213.

[Received ???? 2015. Revised ???? 2015]