Article

Composite Likelihood Methods Based on Minimum Density Power Divergence Estimator

Elena Castilla [1]*, Nirian Martín [2], Leandro Pardo[1] and Konstantinos Zografos [3]

[1] Department of Statistics and O.R. I, Complutense University of Madrid, 28040 Madrid, Spain; lpardo@mat.ucm.es
[2] Department of Statistics and O.R. II, Complutense University of Madrid, 28003 Madrid, Spain; nirian@estad.ucm.es
[3] Department of Mathematics, University of Ioannina, 45110 Ioannina, Greece; kzograf@uoi.gr
* Correspondence: elecasti@mat.ucm.es

Abstract: In this paper a robust version of the Wald test statistic for composite likelihood is considered by using the composite minimum density power divergence estimator instead of the composite maximum likelihood estimator. This new family of test statistics will be called Wald-type test statistics. The problem of testing a simple and a composite null hypothesis is considered and the robustness is studied on the basis of a simulation study. Previously, the composite minimum density power divergence estimator is introduced and its asymptotic properties are studied.

Keywords: composite likelihood; maximum composite likelihood estimator; Wald test statistic; composite minimum density power divergence estimator; Wald-type test statistics.

---

1. Introduction

It is well-known that the likelihood function is one of the most important tools in the classical inference and the resultant estimator, the maximum likelihood estimator (MLE), has nice efficient properties although it has no so good robustness properties.

Tests based on MLE (likelihood ratio test, Wald test, Rao's test, etc.) have, usually, good efficient properties but in presence of outliers the behavior is not so good. To solve these situations many robust estimators have been introduced in the statistical literature, some of them based on distance measures or divergence measures. In particular, density power divergence measures introduced in [1] have given good robust estimators: minimum density power divergences estimators (MDPDE) and, based on them, some robust test statistics have been considered for testing simple and composite null hypotheses. Some of these tests are based on divergence measures (see [2] and [3]) and some other are used to extend the classical Wald test, see [4], [5], [6] and references therein.

The classical likelihood function requires exact specification of the probability density function but in most applications the true distribution is unknown. In some cases, where the data distribution is available in an analytic form, the likelihood function is still mathematically intractable due to the complexity of the probability density function. There are many alternatives to the classical likelihood function; in this paper we focus on the composite likelihood. Composite likelihood is an inference function derived by multiplying a collection of component likelihoods; the particular collection used is a conditional determined by the context. Therefore, the composite likelihood reduces the computational complexity so that it is possible to deal with large datasets and very complex models even when the use of standard likelihood methods is not feasible. Asymptotic normality of the composite maximum likelihood estimator (CMLE) still holds with Godambe information matrix to replace the expected information in the expression of the asymptotic variance-covariance matrix. This allows the construction of composite likelihood ratio test statistics, Wald-type test statistics as well as Score-type statistics. A review of composite likelihood methods is given in [7]. We have to mention at this point that CMLE, as well as the respective test statistics, are seriously affected by the presence of outliers in the set of available data.

The main purpose of the paper is to introduce a new robust family of estimators, namely, composite minimum density power divergence estimators (CMDPDE) as well as a new family of Wald-type test statistics based on the CMDPDE in order to get broad classes of robust estimators and test statistics.

In Section 2 we introduce the CMDPDE and we obtain the estimating system of equations to find it. The asymptotic distribution of the CMDPDE is obtained in Subsection 2.1. Subsection 2.2 is devoted to the definition of a family of Wald-type test statistics, based on CMDPDE, for testing simple and composite null hypotheses. The asymptotic distribution of these Wald-type test statistics is obtained as well as some asymptotic approximations to the power function. A numerical example, presented previously in [8], is studied in Section 3. A simulation study based on this example is also presented (Subsection 3.1), in order to study the robustness of the CMDPDE as well as the performance of the Wald-type test statistics based on CMDPDE. Proofs of results are presented in the Appendix A.

2. Composite Minimum Density Power Divergence Estimator

We adopt here the notation by [9], regarding composite likelihood function and the respective CMLE. In this regard, let $\{f(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p, p \geq 1\}$ be a parametric identifiable family of distributions for an observation $\boldsymbol{y}$, a realization of a random m-vector $\boldsymbol{Y}$. In this setting, the composite density based on K different marginal or conditional distributions has the form

$$\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}) = \prod_{k=1}^{K} f_{A_k}^{w_k}(y_j, j \in A_k; \boldsymbol{\theta})$$

and the corresponding composite log-density has the form

$$c\ell(\boldsymbol{\theta},\boldsymbol{y}\ ) = \sum_{k=1}^{K} w_k \ell_{A_k}(\boldsymbol{\theta},\boldsymbol{y}),$$

with

$$\ell_{A_k}(\boldsymbol{\theta},\boldsymbol{y}) = \log f_{A_k}(y_j, j \in A_k; \boldsymbol{\theta}),$$

where $\{A_k\}_{k=1}^K$ is a family of random variables associated either with marginal or conditional distributions involving some y_j, $j \in \{1,...,m\}$ and w_k, $k = 1,...,K$ are non-negative and known weights. If the weights are all equal, then they can be ignored. In this case all the statistical procedures produce equivalent results.

Let also $\boldsymbol{y}_1,...,\boldsymbol{y}_n$ be independent and identically distributed replications of $\boldsymbol{y}$. We denote by

$$c\ell(\boldsymbol{\theta},\boldsymbol{y}_1,...,\boldsymbol{y}_n) = \sum_{i=1}^{n} c\ell(\boldsymbol{\theta},\boldsymbol{y}_i)$$

the composite log-likelihood function for the whole sample. In complete accordance with the classic MLE, the CMLE, $\widehat{\boldsymbol{\theta}}_c$, is defined by

$$\widehat{\boldsymbol{\theta}}_c = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n} c\ell(\boldsymbol{\theta},\boldsymbol{y}_i) = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n} \sum_{k=1}^{K} w_k \ell_{A_k}(\boldsymbol{\theta},\boldsymbol{y}_i). \tag{1}$$

It can be also obtained by the solution of the equations

$$\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}_1,...,\boldsymbol{y}_n) = \boldsymbol{0}_p,$$

where

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1

Peer-reviewed version available at Entropy 2018, 20, 18; doi:10.3390/e20010018

$$u(\boldsymbol{\theta},\boldsymbol{y}_1,...,\boldsymbol{y}_n) = \frac{\partial c\ell(\boldsymbol{\theta},\boldsymbol{y}_1,...,\boldsymbol{y}_n)}{\partial\boldsymbol{\theta}} = \sum_{i=1}^{n}\sum_{k=1}^{K} w_k \frac{\partial \ell_{A_k}(\boldsymbol{\theta},\boldsymbol{y}_i)}{\partial\boldsymbol{\theta}}.$$

We are going to see how it is possible to get the CMLE, $\widehat{\boldsymbol{\theta}}_c$, on the basis of the Kullback-Leibler divergence measure. We shall denote by $g(\boldsymbol{y})$ the density generating the data with respective distribution function denoted by G. The Kullback-Leibler divergence between the density function $g(\boldsymbol{y})$ and the composite density function $\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})$ is given by

$$\begin{aligned} d_{KL}(g(.),\mathcal{CL}(\boldsymbol{\theta},.)) &= \int_{\mathbb{R}^m} g(\boldsymbol{y}) \log \frac{g(\boldsymbol{y})}{\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})} d\boldsymbol{y} \\ &= \int_{\mathbb{R}^m} g(\boldsymbol{y}) \log g(\boldsymbol{y})\, d\boldsymbol{y} - \int_{\mathbb{R}^m} g(\boldsymbol{y}) \log \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}) d\boldsymbol{y}. \end{aligned}$$

The term

$$\int_{\mathbb{R}^m} g(\boldsymbol{y}) \log g(\boldsymbol{y}) d\boldsymbol{y}$$

can be removed because it does not depend on $\boldsymbol{\theta}$; hence, we can define the following estimator of $\boldsymbol{\theta}$, based on the Kullback-Leibler divergence

$$\widehat{\boldsymbol{\theta}}_{KL} = \arg\min_{\boldsymbol{\theta}} d_{KL}(g(.),\, \mathcal{CL}(\boldsymbol{\theta},.))$$

or equivalently

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{KL} &= \arg\min_{\boldsymbol{\theta}} \left(-\int_{\mathbb{R}^m} g(\boldsymbol{y}) \log \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}) d\boldsymbol{y} \right) \\ &= \arg\min_{\boldsymbol{\theta}} \left(-\int_{\mathbb{R}^m} \log \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}) dG(\boldsymbol{y}) \right). \end{aligned} \tag{2}$$

If we replace in (2) the distribution function G by the empirical distribution function G_n we have

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{KL} &= \arg\min_{\boldsymbol{\theta}} \left(-\int_{\mathbb{R}^m} \log \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}) dG_n(\boldsymbol{y}) \right) \\ &= \arg\min_{\boldsymbol{\theta}} \left(-\frac{1}{n} \sum_{i=1}^{n} c\ell(\boldsymbol{\theta},\boldsymbol{y}_i) \right) \end{aligned}$$

and this expression is equivalent to the expression (1). Therefore, the estimator $\widehat{\boldsymbol{\theta}}_{KL}$ coincides with the CMLE. Based on the previous idea we are going to introduce, in a natural way, the composite minimum density power divergence estimator (CMDPDE).

The CMLE, $\widehat{\boldsymbol{\theta}}_c$, obeys asymptotic normality, see [9], and in particular

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{G}_*^{-1}(\boldsymbol{\theta})\right),$$

where $\boldsymbol{G}_*(\boldsymbol{\theta})$ denotes Godambe information matrix, defined by

$$\boldsymbol{G}_*(\boldsymbol{\theta}) = \boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{J}^{-1}(\boldsymbol{\theta})\boldsymbol{H}(\boldsymbol{\theta}),$$

with $\boldsymbol{H}(\boldsymbol{\theta})$ being the sensitivity or Hessian matrix and $\boldsymbol{J}(\boldsymbol{\theta})$ being the variability matrix, defined, respectively, by

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1

Peer-reviewed version available at Entropy 2018, 20, 18; doi:10.3390/e20010018

$$\boldsymbol{H}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[-\tfrac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{u}^T(\boldsymbol{\theta},\boldsymbol{Y})],$$
$$\boldsymbol{J}(\boldsymbol{\theta}) = Var_{\boldsymbol{\theta}}[\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{Y})] = E_{\boldsymbol{\theta}}[\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{Y})\boldsymbol{u}^T(\boldsymbol{\theta},\boldsymbol{Y})],$$

where the superscript T denotes the transpose of a vector or a matrix.

The matrices $\boldsymbol{H}(\boldsymbol{\theta})$ and $\boldsymbol{J}(\boldsymbol{\theta})$ are, by definition, nonegative definite matrices but throughout this paper both, $\boldsymbol{H}(\boldsymbol{\theta})$ and $\boldsymbol{J}(\boldsymbol{\theta})$, are assumed to be positive definite matrices. Since the component score functions can be correlated, we have $\boldsymbol{H}(\boldsymbol{\theta}) \neq \boldsymbol{J}(\boldsymbol{\theta})$. If $c\ell(\boldsymbol{\theta},\boldsymbol{y})$ is a true log-likelihood function then $\boldsymbol{H}(\boldsymbol{\theta}) = \boldsymbol{J}(\boldsymbol{\theta}) = \boldsymbol{I}_F(\boldsymbol{\theta})$, being $\boldsymbol{I}_F(\boldsymbol{\theta})$ the Fisher information matrix of the model. Using multivariate version of the Cauchy-Schwarz inequality we have that the matrix $\boldsymbol{G}_*(\boldsymbol{\theta}) - \boldsymbol{I}_F(\boldsymbol{\theta})$ is non-negative definite, i.e., the full likelihood function is more efficient than any other composite likelihood function (cf. [10], Lemma 4A).

We are going now to proceed to the definition of the CMDPDE which is based on the density power divergence measure, defined as follows. For two densities p and q associated with two m-dimensional random variables respectively, density power divergence (DPD) between p and q was defined in [1] by

$$d_\beta(p,q) = \int_{\mathbb{R}^m} \left\{ q(\boldsymbol{y})^{1+\beta} - \left(1+\frac{1}{\beta}\right) q(\boldsymbol{y})^\beta p(\boldsymbol{y}) + \frac{1}{\beta}\, p(\boldsymbol{y})^{1+\beta} \right\} d\boldsymbol{y},$$

for $\beta > 0$, while for $\beta = 0$ it is defined by

$$\lim_{\beta\to 0} d_\beta(p,q) = d_{KL}(p,q).$$

For more details about this family of divergence measures we refer to [11].

In this paper we are going to consider DPD measures between the density function $g(\boldsymbol{y})$ and the composite density function $\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})$, i.e.,

$$d_\beta(g(.),\mathcal{CL}(\boldsymbol{\theta},.)) = \int_{\mathbb{R}^m} \left\{ \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta} - \left(1+\frac{1}{\beta}\right) \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^\beta g(\boldsymbol{y}) + \frac{1}{\beta}\, g(\boldsymbol{y})^{1+\beta} \right\} d\boldsymbol{y} \quad (3)$$

for $\beta > 0$, while for $\beta = 0$ we have,

$$\lim_{\beta\to 0} d_\beta(g(.),\mathcal{CL}(\boldsymbol{\theta},.)) = d_{KL}(g(.),\mathcal{CL}(\boldsymbol{\theta},.)).$$

The CMDPDE, $\widehat{\boldsymbol{\theta}}_c^\beta$, is defined by

$$\widehat{\boldsymbol{\theta}}_c^\beta = \arg\min_{\boldsymbol{\theta}\in\Theta} d_\beta(g(.),\mathcal{CL}(\boldsymbol{\theta},.)).$$

The term

$$\int_{\mathbb{R}^m} g(\boldsymbol{y})^{1+\beta} d\boldsymbol{y}$$

does not depend on $\boldsymbol{\theta}$ and consequently the minimization of (3) with respect to $\boldsymbol{\theta}$ is equivalent to minimize

$$\int_{\mathbb{R}^m} \left(\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta} - \left(1+\frac{1}{\beta}\right) \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^\beta g(\boldsymbol{y}) \right) d\boldsymbol{y}$$

or

$$\int_{\mathbb{R}^m} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta} d\boldsymbol{y} - \left(1+\frac{1}{\beta}\right) \int_{\mathbb{R}^m} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta} dG(\boldsymbol{y}).$$

Now, we replace the distribution function G by the empirical distribution function G_n and we get

$$\int_{\mathbb{R}^m} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta} d\boldsymbol{y} - \left(1+\frac{1}{\beta}\right) \frac{1}{n}\sum_{i=1}^{n} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta}. \tag{4}$$

In consequence, for a fixed value of β, the CMDPDE of $\boldsymbol{\theta}$ can be obtained by minimizing the expression given in (4). Or equivalently by maximizing the expression

$$\frac{1}{n\beta}\sum_{i=1}^{n} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta} - \frac{1}{1+\beta}\int_{\mathbb{R}^m} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta} d\boldsymbol{y}. \tag{5}$$

Under differentiability of the model the maximization of the function in equation (5) leads to an estimating system of equations of the form

$$\frac{1}{n}\sum_{i=1}^{n} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta} \frac{\partial c\ell(\boldsymbol{\theta},\boldsymbol{y}_i)}{\partial\boldsymbol{\theta}} - \int_{\mathbb{R}^m} \frac{\partial c\ell(\boldsymbol{\theta},\boldsymbol{y})}{\partial\boldsymbol{\theta}} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta} d\boldsymbol{y} = \mathbf{0}. \tag{6}$$

The system of equations (6) can be written as

$$\frac{1}{n}\sum_{i=1}^{n} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta} \boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}_i) - \int_{\mathbb{R}^m} \boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}) \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta} d\boldsymbol{y} = \mathbf{0}. \tag{7}$$

and the CMDPDE $\widehat{\boldsymbol{\theta}}_c^{\beta}$ of $\boldsymbol{\theta}$ is obtained by the solution of (7).

2.1. Asymptotic Distribution of the Composite Minimum Density Power Divergence Estimator

Equation (7) can be written as follows

$$\frac{1}{n}\sum_{i=1}^{n} \Psi_{\beta}\left(\boldsymbol{y}_i, \boldsymbol{\theta}\right) = \mathbf{0}$$

with

$$\Psi_{\beta}\left(\boldsymbol{y}_i, \boldsymbol{\theta}\right) = \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta} \boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}_i) - \int_{\mathbb{R}^m} \boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}) \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta} d\boldsymbol{y}.$$

Therefore the CMDPDE, $\widehat{\boldsymbol{\theta}}_c^{\beta}$, is an M-estimator. In this case it is well-known (cf.[12]) that the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_c^{\beta}$ is given by

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_c^{\beta} - \boldsymbol{\theta}) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{H}_{\beta}^{-1}(\boldsymbol{\theta}) \boldsymbol{J}_{\beta}(\boldsymbol{\theta}) \boldsymbol{H}_{\beta}^{-1}(\boldsymbol{\theta})\right),$$

being

$$\boldsymbol{H}_{\beta}(\boldsymbol{\theta}) = \boldsymbol{E}_{\boldsymbol{\theta}}\left[-\frac{\partial \Psi_{\beta}\left(\boldsymbol{Y},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^T}\right]$$

and

$$\boldsymbol{J}_{\beta}(\boldsymbol{\theta}) = \boldsymbol{E}_{\theta}\left[\Psi_{\beta}\left(\boldsymbol{Y},\boldsymbol{\theta}\right) \Psi_{\beta}\left(\boldsymbol{Y},\boldsymbol{\theta}\right)^T\right].$$

We are going to establish the expressions of $\boldsymbol{H}_{\beta}(\boldsymbol{\theta})$ and $\boldsymbol{J}_{\beta}(\boldsymbol{\theta})$. In relation to $\boldsymbol{H}_{\beta}(\boldsymbol{\theta})$ we have

$$\begin{aligned}\frac{\partial \Psi_\beta(\boldsymbol{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} &= \beta \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta-1}\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}) + \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta}\frac{\partial \boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T}{\partial \boldsymbol{\theta}} \\ &\quad - \int_{\mathbb{R}^m} \frac{\partial \boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T}{\partial \boldsymbol{\theta}}\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta}d\boldsymbol{y} - (1+\beta)\int_{\mathbb{R}^m}\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta}\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})d\boldsymbol{y}\end{aligned}$$

and

$$\boldsymbol{H}_\beta(\boldsymbol{\theta}) = \boldsymbol{E}_{\boldsymbol{\theta}}\left[-\frac{\partial \Psi_\beta(\boldsymbol{Y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right] = \int_{\mathbb{R}^m}\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})d\boldsymbol{y}. \tag{8}$$

In relation to $\boldsymbol{J}_\beta(\boldsymbol{\theta})$ we have,

$$\begin{aligned}\Psi_\beta(\boldsymbol{Y},\boldsymbol{\theta})\Psi_\beta(\boldsymbol{Y},\boldsymbol{\theta})^T &= \left(\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}) - \int_{\mathbb{R}^m}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta}d\boldsymbol{y}\right) \\ &\quad \left(\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T - \int_{\mathbb{R}^m}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta}d\boldsymbol{y}\right) \\ &= \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{2\beta}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T - \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})\int_{\mathbb{R}^m}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta}d\boldsymbol{y} \\ &\quad -\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T\int_{\mathbb{R}^m}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta}d\boldsymbol{y} \\ &\quad + \left(\int_{\mathbb{R}^m}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta}d\boldsymbol{y}\right)\left(\int_{\mathbb{R}^m}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta}d\boldsymbol{y}\right).\end{aligned}$$

Then

$$\begin{aligned}\boldsymbol{J}_\beta(\boldsymbol{\theta}) &= \boldsymbol{E}_\theta\left[\Psi_\beta(\boldsymbol{Y},\boldsymbol{\theta})\Psi_\beta(\boldsymbol{Y},\boldsymbol{\theta})^T\right] = \int_{\mathbb{R}^m}\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{2\beta+1}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T d\boldsymbol{y} && (9) \\ &\quad - \int_{\mathbb{R}^m}\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})d\boldsymbol{y}\int_{\mathbb{R}^m}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta}d\boldsymbol{y}. && (10)\end{aligned}$$

Based on the previous results we have the following Theorem.

Theorem 1. *Under some regularity conditions (cf. [13], pp.58 or [14], pp.144) we have*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_c^{\beta} - \boldsymbol{\theta}) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta})\boldsymbol{J}_\beta(\boldsymbol{\theta})\boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta})\right),$$

where the matrices $\boldsymbol{H}_\beta(\boldsymbol{\theta})$ and $\boldsymbol{J}_\beta(\boldsymbol{\theta})$ were defined in (8) and (9), respectively.

Remark 1. If we apply the previous theorem for $\beta = 0$ then we get the CMLE and the asymptotic variance covariance matrix coincides with Godambe information matrix because

$$\boldsymbol{H}_\beta(\boldsymbol{\theta}) = \boldsymbol{H}(\boldsymbol{\theta}) \text{ and } \boldsymbol{J}_\beta(\boldsymbol{\theta}) = \boldsymbol{J}(\boldsymbol{\theta}),$$

for $\beta = 0$.

2.2. Wald-Type Tests Statistics Based on Composite Minimum Power Divergence Estimator

Wald-type test statistics based on MDPDE have been considered with excellent results in relation to the robustness in different statistical problems, see for instance [4], [5] and [6].

Motivated by those works, we focus in this section on the definition and the study of Wald-type test statistics which are defined by means of CMDPDE estimators instead of MDPDE estimators. In this context, if we are interested in testing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1

Peer-reviewed version available at *Entropy* **2018**, *20*, 18; doi:10.3390/e20010018

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \text{ against } H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \tag{11}$$

we can consider the family of Wald-type test statistics

$$W_{n,\beta}^0 = n(\widehat{\boldsymbol{\theta}}_c^\beta - \boldsymbol{\theta}_0)^T \left(\boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta(\boldsymbol{\theta}_0) \boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0) \right)^{-1} (\widehat{\boldsymbol{\theta}}_c^\beta - \boldsymbol{\theta}_0). \tag{12}$$

For $\beta = 0$ we get the classical Wald type test statistic considered in the composite likelihood methods (see, for instance, [7]).

In the following Theorem we present the asymptotic null distribution of the family of the Wald-type test statistics $W_{n,\beta}^0$.

Theorem 2. *The asymptotic distribution of the Wald-type test statistics given in (12) is a chi-square distribution with p degrees of freedom.*

The proof of this Theorem 2 is given in the Appendix A.1.

Theorem 3. *Let $\boldsymbol{\theta}^*$ be the true value of the parameter $\boldsymbol{\theta}$, with $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$. Then it holds*

$$\sqrt{n} \left(l\left(\widehat{\boldsymbol{\theta}}_c^\beta \right) - l\left(\boldsymbol{\theta}^* \right) \right) \xrightarrow[n\to\infty]{\mathcal{L}} N(\boldsymbol{0}, \sigma^2_{W_\beta^0}\left(\boldsymbol{\theta}^* \right)),$$

being

$$l\left(\boldsymbol{\theta} \right) = \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0 \right)^T \left(\boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta(\boldsymbol{\theta}_0) \boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0) \right)^{-1} \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0 \right)$$

and

$$\sigma^2_{W_\beta^0}\left(\boldsymbol{\theta}^* \right) = 4\left(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0 \right)^T \left(\boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0) \boldsymbol{J}_\beta(\boldsymbol{\theta}_0) \boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0) \right)^{-1} \left(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0 \right). \tag{13}$$

The proof of the Theorem is outlined in the Appendix A.2.

Remark 2. Based on the previous result we can approximate the power, $\beta_{W_n^0}$, of the Wald-type test statistics in $\boldsymbol{\theta}^*$, by

$$\begin{aligned}
\beta_{W_{n,\beta}^0}\left(\boldsymbol{\theta}^* \right) &= \Pr\left(W_{n,\beta}^0 > \chi^2_{p,\alpha} / \boldsymbol{\theta} = \boldsymbol{\theta}^* \right) \\
&= \Pr\left(l\left(\widehat{\boldsymbol{\theta}}_c^\beta \right) - l\left(\boldsymbol{\theta}^* \right) > \frac{\chi^2_{p,\alpha}}{n} - l\left(\boldsymbol{\theta}^* \right) \middle| \boldsymbol{\theta} = \boldsymbol{\theta}^* \right) \\
&= \Pr\left(\sqrt{n} \left(l\left(\widehat{\boldsymbol{\theta}}_c^\beta \right) - l\left(\boldsymbol{\theta}^* \right) \right) > \sqrt{n} \left(\frac{\chi^2_{p,\alpha}}{n} - l\left(\boldsymbol{\theta}^* \right) \right) \middle| \boldsymbol{\theta} = \boldsymbol{\theta}^* \right) \\
&= \Pr\left(\sqrt{n} \frac{\left(l\left(\widehat{\boldsymbol{\theta}}_c^\beta \right) - l\left(\boldsymbol{\theta}^* \right) \right)}{\sigma_{W_{n,\beta}^0}\left(\boldsymbol{\theta}^* \right)} > \frac{\sqrt{n}}{\sigma_{W_{n,\beta}^0}\left(\boldsymbol{\theta}^* \right)} \left(\frac{\chi^2_{p,\alpha}}{n} - l\left(\boldsymbol{\theta}^* \right) \right) \middle| \boldsymbol{\theta} = \boldsymbol{\theta}^* \right) \\
&= 1 - \Phi_n \left(\frac{\sqrt{n}}{\sigma_{W_{n,\beta}^0}\left(\boldsymbol{\theta}^* \right)} \left(\frac{\chi^2_{p,\alpha}}{n} - l\left(\boldsymbol{\theta}^* \right) \right) \right),
\end{aligned}$$

where Φ_n is a sequence of distributions functions tending uniformly to the standard normal distribution function $\Phi(x)$.

It is clear that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1

Peer-reviewed version available at Entropy 2018, 20, 18; doi:10.3390/e20010018

$$\lim_{n\to\infty}\beta_{W^0_{n,\beta}}\left(\boldsymbol{\theta}^*\right)=1$$

for all $\alpha\in(0,1)$. Therefore the Wald-type test statistics are consistent in the sense of Fraser.

In many practical hypothesis testing problems, the restricted parameter space $\Theta_0\subset\Theta$ is defined by a set of r restrictions of the form

$$\boldsymbol{g}(\boldsymbol{\theta})=\mathbf{0}_r \tag{14}$$

on Θ, where $\boldsymbol{g}:\mathbb{R}^p\to\mathbb{R}^r$ is a vector-valued function such that the $p\times r$ matrix

$$\boldsymbol{G}\left(\boldsymbol{\theta}\right)=\frac{\partial\boldsymbol{g}^T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \tag{15}$$

exists and is continuous in $\boldsymbol{\theta}$ and $\text{rank}(\boldsymbol{G}\left(\boldsymbol{\theta}\right))=r$; where $\mathbf{0}_r$ denotes the null vector of dimension r.

Now we are going to consider composite null hypotheses, $\Theta_0\subset\Theta$, in the way considered in (14) and our interest is in testing

$$H_0:\boldsymbol{\theta}\in\Theta_0 \text{ against } H_1:\boldsymbol{\theta}\notin\Theta_0 \tag{16}$$

on the basis of a random simple of size n, $X_1,....X_n$.

Definition 4. The family of Wald-type test statistics for testing (16) is given by

$$W_{n,\beta}=n\boldsymbol{g}\left(\widehat{\boldsymbol{\theta}}_c^\beta\right)^T\left[\boldsymbol{G}^T(\widehat{\boldsymbol{\theta}}_c^\beta)\boldsymbol{H}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_c^\beta)\boldsymbol{J}_\beta(\widehat{\boldsymbol{\theta}}_c^\beta)\boldsymbol{H}_\beta^{-1}(\widehat{\boldsymbol{\theta}}_c^\beta)\boldsymbol{G}(\widehat{\boldsymbol{\theta}}_c^\beta)\right]^{-1}\boldsymbol{g}\left(\widehat{\boldsymbol{\theta}}_c^\beta\right), \tag{17}$$

where the matrices $\boldsymbol{G}(\boldsymbol{\theta}),\boldsymbol{H}_\beta\left(\boldsymbol{\theta}\right)$ and $\boldsymbol{J}_\beta\left(\boldsymbol{\theta}\right)$ were defined in (15), (8) and (9), respectively and the function $\boldsymbol{g}$ in (14).

If we consider $\beta=0$ then $\widehat{\boldsymbol{\theta}}_\beta$ coincides with the MLE, $\widehat{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}$ and $\boldsymbol{H}_\beta^{-1}(\widehat{\boldsymbol{\theta}})\boldsymbol{J}_\beta(\widehat{\boldsymbol{\theta}})\boldsymbol{H}_\beta^{-1}(\widehat{\boldsymbol{\theta}})$ with the inverse of the Fisher information matrix and then we get the classical Wald test statistic considered in the composite likelihood methods.

In the next theorem we present the asymptotic distribution of $W_{n,\beta}$.

Theorem 5. *The asymptotic distribution of the Wald-type test statistics, given in (17), is a chi-square distribution with r degrees of freedom.*

The proof of this Theorem is presented in the Appendix A.3.

Consider the null hypothesis $H_0:\boldsymbol{\theta}\in\Theta_0\subset\Theta$. By Theorem 5, the null hypothesis should be rejected if $W_{n,\beta}\geq\chi^2_{r,\alpha}$. The following theorem can be used to approximate the power function. Assume that $\boldsymbol{\theta}^*\notin\Theta_0$ is the true value of the parameter so that $\widehat{\boldsymbol{\theta}}_\beta\xrightarrow[n\to\infty]{a.s.}\boldsymbol{\theta}^*$.

Theorem 6. *Let $\boldsymbol{\theta}^*$ be the true value of the parameter, with $\boldsymbol{\theta}^*\neq\boldsymbol{\theta}_0$. Then it holds*

$$\sqrt{n}\left(l^*\left(\widehat{\boldsymbol{\theta}}_c^\beta\right)-l^*\left(\boldsymbol{\theta}^*\right)\right)\xrightarrow[n\to\infty]{L}N(0,\sigma^2_{W_\beta}\left(\boldsymbol{\theta}^*\right))$$

being

$$l^*\left(\boldsymbol{\theta}\right)=n\boldsymbol{g}\left(\boldsymbol{\theta}\right)^T\left[\boldsymbol{G}^T(\boldsymbol{\theta}_0)\boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{J}_\beta(\boldsymbol{\theta}_0)\boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}(\boldsymbol{\theta}_0)\right]^{-1}\boldsymbol{g}\left(\boldsymbol{\theta}\right)$$

and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1

Peer-reviewed version available at Entropy 2018, 20, 18; doi:10.3390/e20010018

$$\sigma^2_{W_\beta}(\boldsymbol{\theta}^*) = \left(\frac{\partial l^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{J}_\beta(\boldsymbol{\theta}_0)\boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0)\left(\frac{\partial l^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}. \tag{18}$$

3. Numerical Example

In this section we shall consider an example, studied previously by [8], in order to study the robustness of CMLE. The aim of this section is to clarify the different issues which are discussed in the previous sections.

Consider the random vector $\boldsymbol{Y} = (Y_1, Y_2, Y_3, Y_4)^T$ which follows a four dimensional normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4)^T$ and variance-covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho & 2\rho & 2\rho \\ \rho & 1 & 2\rho & 2\rho \\ 2\rho & 2\rho & 1 & \rho \\ 2\rho & 2\rho & \rho & 1 \end{pmatrix}, \tag{19}$$

i.e., we suppose that the correlation between Y_1 and Y_2 is the same as the correlation between Y_3 and Y_4. Taking into account that $\boldsymbol{\Sigma}$ should be semi- positive definite, the following condition is imposed, $-\frac{1}{5} \leq \rho \leq \frac{1}{3}$. In order to avoid several problems regarding the consistency of the CMLE of the parameter ρ (cf. [8]), we shall consider the composite likelihood function

$$\mathcal{CL}(\boldsymbol{\theta}, \boldsymbol{y}) = f_{A_1}(\boldsymbol{\theta}, \boldsymbol{y}) f_{A_2}(\boldsymbol{\theta}, \boldsymbol{y}),$$

where

$$f_{A_1}(\boldsymbol{\theta}, \boldsymbol{y}) = f_{12}(\mu_1, \mu_2, \rho, y_1, y_2),$$
$$f_{A_2}(\boldsymbol{\theta}, \boldsymbol{y}) = f_{34}(\mu_3, \mu_4, \rho, y_3, y_4),$$

where f_{12} and f_{34} are the densities of the marginals of $\boldsymbol{Y}$, i.e. bivariate normal distributions with mean vectors $(\mu_1, \mu_2)^T$ and $(\mu_3, \mu_4)^T$, respectively, and common variance-covariance matrix

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

with densities given by

$$f_{h,h+1}(\mu_h, \mu_{h+1}, \rho, y_h, y_{h+1}) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\tfrac{1}{2(1-\rho^2)} Q(y_h, y_{h+1})\right\}, \; h \in \{1, 3\},$$

being

$$Q(y_h, y_{h+1}) = (y_h - \mu_h)^2 - 2\rho(y_h - \mu_h)(y_{h+1} - \mu_{h+1}) + (y_{h+1} - \mu_{h+1})^2, \; h \in \{1, 3\}.$$

By $\boldsymbol{\theta}$ we are denoting the parameter vector of our model, i.e, $\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3, \mu_4, \rho)^T$. We are going to get the system of equations that it is necessary to solve in order to obtain the CMDPDE

$$\widehat{\boldsymbol{\theta}}_c^\beta = \left(\widehat{\mu}_{1,c}^\beta, \widehat{\mu}_{2,c}^\beta, \widehat{\mu}_{3,c}^\beta, \widehat{\mu}_{4,c}^\beta, \widehat{\rho}_c^\beta\right)^T.$$

The estimator $\widehat{\boldsymbol{\theta}}_c^\beta$ is obtained by maximizing the expression (4) with respect to $\boldsymbol{\theta}$. Firstly we are going to get

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1

Peer-reviewed version available at *Entropy* **2018**, *20*, 18; doi:10.3390/e20010018

$$\begin{aligned}\int_{\mathbb{R}^4} \frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta}}{\partial \boldsymbol{\theta}} d\boldsymbol{y} &= \frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathbb{R}^4} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta} d\boldsymbol{y} \\ &= \frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathbb{R}^4} f_{12}(\mu_1,\mu_2,\rho,y_1,y_2)^{\beta+1} f_{34}(\mu_3,\mu_4,\rho,y_3,y_4)^{\beta+1} dy_1 dy_2 dy_3 dy_4 \\ &= \frac{\partial}{\partial \boldsymbol{\theta}} \left(\int_{\mathbb{R}^2} f_{12}(\mu_1,\mu_2,\rho,y_1,y_2)^{\beta+1} dy_1 dy_2 \int_{\mathbb{R}^2} f_{34}(\mu_3,\mu_4,\rho,y_3,y_4)^{\beta+1} dy_3 dy_4 \right).\end{aligned}$$

Based on [13] (pp. 32)

$$\int_{\mathbb{R}^2} f_{12}(\mu_1,\mu_2,\rho,y_1,y_2)^{\beta+1} dy_1 dy_2 = \int_{\mathbb{R}^2} f_{34}(\mu_3,\mu_4,\rho,y_3,y_4)^{\beta+1} dy_3 dy_4 = \frac{(1-\rho^2)^{-\frac{\beta}{2}}}{\beta+1} (2\pi)^{-\beta}.$$

Then

$$\int_{\mathbb{R}^4} \frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta}}{\partial \boldsymbol{\theta}} d\,\boldsymbol{y} = \frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathbb{R}^4} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{1+\beta} d\,\boldsymbol{y} = \frac{\partial}{\partial \boldsymbol{\theta}} \frac{(1-\rho^2)^{-\beta}}{(\beta+1)^2} (2\pi)^{-2\beta}$$

and

$$\frac{\partial}{\partial \mu_i} \frac{(1-\rho^2)^{-\beta}}{(\beta+1)^2} (2\pi)^{-2\beta} = 0, \ i = 1,2,3,4,$$

while

$$\frac{\partial}{\partial \rho} \frac{(1-\rho^2)^{-\beta}}{(\beta+1)^2} (2\pi)^{-2\beta} = \frac{\beta (2\pi)^{-2\beta}}{(\beta+1)^2} \frac{2\rho}{(1-\rho^2)^{\beta+1}}.$$

Now, we are going to get

$$\frac{1}{n\beta} \sum_{i=1}^{n} \frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta}}{\partial \boldsymbol{\theta}}$$

in order to obtain the CMDPDE, $\widehat{\boldsymbol{\theta}}_c^{\beta}$, by maximizing (4) with respect to $\boldsymbol{\theta}$.

We have,

$$\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta} = f_{12}(\mu_1,\mu_2,\rho,y_1,y_2)^{\beta} f_{34}(\mu_3,\mu_4,\rho,y_3,y_4)^{\beta}.$$

Therefore,

$$\frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta}}{\partial \mu_1} = \beta f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})^{\beta-1} \left\{ -\frac{1}{2(1-\rho^2)} \left[-2(y_{1i}-\mu_1) + 2\rho(y_{2i}-\mu_2) \right] \right\} f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})^{\beta}$$

and the expression

$$\frac{1}{n\beta} \sum_{i=1}^{n} \frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta}}{\partial \mu_1} = 0$$

leads to the estimator of μ_1, given by

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1

Peer-reviewed version available at *Entropy* **2018**, *20*, 18; doi:10.3390/e20010018

$$\frac{1}{n}\sum_{i=1}^{n} f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})^{\beta-1} f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})^{\beta}\left\{-\frac{1}{2\,(1-\rho^2)}\left[-2\,(y_{1i}-\mu_1)+2\rho\,(y_{2i}-\mu_2)\right]\right\}=0. \tag{20}$$

In a similar way

$$\frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta}}{\partial \mu_2}=\beta f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})^{\beta-1}\left\{-\frac{1}{2\,(1-\rho^2)}\left[-2\,(y_{2i}-\mu_2)+2\rho\,(y_{1i}-\mu_1)\right]\right\} f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})^{\beta},$$

$$\frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta}}{\partial \mu_3}=\beta f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})^{\beta}\left\{-\frac{1}{2\,(1-\rho^2)}\left[-2\,(y_{3i}-\mu_3)+2\rho\,(y_{4i}-\mu_4)\right]\right\} f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})^{\beta-1}$$

and

$$\frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta}}{\partial \mu_4}=\beta f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})^{\beta}\left\{-\frac{1}{2\,(1-\rho^2)}\left[-2\,(y_{4i}-\mu_4)+2\rho\,(y_{3i}-\mu_3)\right]\right\} f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})^{\beta-1}.$$

Therefore the equations

$$\frac{1}{n\beta}\sum_{i=1}^{n}\frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta}}{\partial \mu_2}=0,\ \frac{1}{n\beta}\sum_{i=1}^{n}\frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta}}{\partial \mu_3}=0 \text{ and } \frac{1}{n\beta}\sum_{i=1}^{n}\frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta}}{\partial \mu_4}=0$$

lead to the estimators of μ_2, μ_3 and μ_4, which should be read as follows

$$\frac{1}{n}\sum_{i=1}^{n} f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})^{\beta-1} f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})^{\beta}\left\{-\frac{1}{2\,(1-\rho^2)}\left[-2\,(y_{2i}-\mu_2)+2\rho\,(y_{1i}-\mu_1)\right]\right\}=0, \tag{21}$$

$$\frac{1}{n}\sum_{i=1}^{n} f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})^{\beta-1} f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})^{\beta}\left\{-\frac{1}{2\,(1-\rho^2)}\left[-2\,(y_{3i}-\mu_3)+2\rho\,(y_{4i}-\mu 4)\right]\right\}=0 \tag{22}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})^{\beta} f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})^{\beta}\left\{-\frac{1}{2\,(1-\rho^2)}\left[-2\,(y_{4i}-\mu_4)+2\rho\,(y_{3i}-\mu_3)\right]\right\}=0. \tag{23}$$

Now it is necessary to get

$$\begin{aligned}\frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta}}{\partial \rho} &= \frac{\partial f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})^{\beta} f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})^{\beta}}{\partial \rho}\\ &= \beta f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})^{\beta-1} f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})^{\beta}\frac{\partial f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})}{\partial \rho}\\ &\quad+\beta f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})^{\beta} f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})^{\beta-1}\frac{\partial f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})}{\partial \rho}.\end{aligned}$$

Peer-reviewed version available at *Entropy* **2018**, *20*, 18; doi:10.3390/e20010018

But $\frac{\partial f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})}{\partial\rho}$ is given by

$$
\begin{aligned}
& \frac{1}{2\pi}\frac{(-1)}{(1-\rho^2)}\frac{(-2\rho)}{2\,(1-\rho^2)^{\frac{1}{2}}}\exp\left\{\frac{(-1)}{2\,(1-\rho^2)}\left[(y_{1i}-\mu_1)^2-2\rho\,(y_{1i}-\mu_1)\,(y_{2i}-\mu_2)+(y_{2i}-\mu_2)^2\right]\right\} \\
& +\frac{1}{2\pi\,(1-\rho^2)^{\frac{1}{2}}}\exp\left\{\frac{(-1)}{2\,(1-\rho^2)}\left[(y_{1i}-\mu_1)^2-2\rho\,(y_{1i}-\mu_1)\,(y_{2i}-\mu_2)+(y_{2i}-\mu_2)^2\right]\right\} \\
& \left[\frac{-\rho}{(1-\rho^2)^2}\left((y_{1i}-\mu_1)^2-2\rho\,(y_{1i}-\mu_1)\,(y_{2i}-\mu_2)+(y_{2i}-\mu_2)^2\right)+\frac{1}{(1-\rho^2)}\,(y_{1i}-\mu_1)\,(y_{2i}-\mu_2)\right] \\
= & \frac{\rho}{1-\rho^2}f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})+f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i}) \\
& \left[\frac{-\rho}{(1-\rho^2)^2}\left((y_{1i}-\mu_1)^2-2\rho\,(y_{1i}-\mu_1)\,(y_{2i}-\mu_2)+(y_{2i}-\mu_2)^2\right)+\frac{1}{(1-\rho^2)}\,(y_{1i}-\mu_1)\,(y_{2i}-\mu_2)\right] \\
= & f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})\frac{\rho}{1-\rho^2}\left[1-\frac{1}{1-\rho^2}\left((y_{1i}-\mu_1)^2-2\rho\,(y_{1i}-\mu_1)\,(y_{2i}-\mu_2)+(y_{2i}-\mu_2)^2\right)\right. \\
& \left.+\frac{1}{\rho}\,(y_{1i}-\mu_1)\,(y_{2i}-\mu_2)\right].
\end{aligned}
$$

In a similar way $\frac{\partial f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})}{\partial\rho}$ is given by

$$
\begin{aligned}
& f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})\frac{\rho}{1-\rho^2}\left[1-\frac{1}{1-\rho^2}\left((y_{3i}-\mu_3)^2-2\rho\,(y_{3i}-\mu_3)\,(y_{4i}-\mu_4)+(y_{4i}-\mu_4)^2\right)\right. \\
& \left.+\frac{1}{\rho}\,(y_{3i}-\mu_3)\,(y_{4i}-\mu_4)\right].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{\partial\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^\beta}{\partial\rho} = & \frac{\rho}{1-\rho^2}\beta f_{12}(\mu_1,\mu_2,\rho,y_{1i},y_{2i})^\beta f_{34}(\mu_3,\mu_4,\rho,y_{3i},y_{4i})^\beta \\
& \left\{2+\frac{1}{\rho}\left\{(y_{1i}-\mu_1)\,(y_{2i}-\mu_2)+(y_{3i}-\mu_3)\,(y_{4i}-\mu_4)\right\}\right. \\
& -\frac{1}{1-\rho^2}\left((y_{1i}-\mu_1)^2-2\rho\,(y_{1i}-\mu_1)\,(y_{2i}-\mu_2)+(y_{2i}-\mu_2)^2\right) \\
& \left.-\frac{1}{1-\rho^2}\left((y_{3i}-\mu_3)^2-2\rho\,(y_{3i}-\mu_3)\,(y_{4i}-\mu_4)+(y_{4i}-\mu_4)^2\right)\right\}.
\end{aligned} \tag{24}
$$

So the equation in relation to ρ is given by

$$
\frac{1}{n\beta}\sum_{i=1}^{n}\frac{\partial\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^\beta}{\partial\boldsymbol{\rho}}-\frac{1}{\beta+1}\int_{\mathbb{R}^m}\frac{\partial\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta+1}}{\partial\boldsymbol{\rho}}\boldsymbol{dy}=0
$$

being

$$
\int_{\mathbb{R}^m}\frac{\partial\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^{\beta+1}}{\partial\boldsymbol{\theta}}\boldsymbol{dy}=\frac{\beta(2\pi)^{-2\beta}}{(\beta+1)^2}\frac{2\rho}{(1-\rho^2)^{\beta+1}} \tag{25}
$$

and

$$
\frac{\partial\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y}_i)^\beta}{\partial\rho}
$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1

Peer-reviewed version available at Entropy 2018, 20, 18; doi:10.3390/e20010018

was given in (24).

Finally,

$$\widehat{\boldsymbol{\theta}}_c^{\beta} = \left(\widehat{\mu}_{1,c}^{\beta}, \widehat{\mu}_{2,c}^{\beta}, \widehat{\mu}_{3,c}^{\beta}, \widehat{\mu}_{4,c}^{\beta}, \widehat{\rho}_c^{\beta}\right)^T$$

will be obtained as the solution of the system of equations given by (20), (21), (22), (23) and (25).

After some heavy algebraic manipulations specified in Appendix, Section A.4, the sensitivity and variability matrices are given by

$$\boldsymbol{H}_{\beta}(\boldsymbol{\theta}) = \frac{C_{\beta}}{(\beta+1)(1-\rho^2)} \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 \\ -\rho & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -\rho & 0 \\ 0 & 0 & -\rho & 1 & 0 \\ 0 & 0 & 0 & 0 & 2\frac{(\rho^2+1)+2\rho^2\beta^2}{(1-\rho^2)(1+\beta)} \end{pmatrix} \tag{26}$$

and

$$\boldsymbol{J}_{\beta}(\boldsymbol{\theta}) = \boldsymbol{H}_{2\beta}(\boldsymbol{\theta}) - \boldsymbol{\xi}_{\beta}(\boldsymbol{\theta})^T \boldsymbol{\xi}_{\beta}(\boldsymbol{\theta}), \tag{27}$$

where $C_{\beta} = \frac{1}{(\beta+1)^2}\left(\frac{1}{(2\pi)^2(1-\rho^2)}\right)^{\beta}$ and $\boldsymbol{\xi}_{\beta}(\boldsymbol{\theta}) = (0, 0, 0, 0, \frac{2\rho\beta C_{\beta}}{(\beta+1)(1-\rho^2)})^T$.

3.1. Simulation Study

A simulation study, developed by using the R statistical programming environment, is presented in order to study the behavior of the CMDPDE as well as the behavior of the Wald-type test statistics based on them. The theoretical model studied in the previous example is considered. The parameters in the model are

$$\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3, \mu_4, \rho)^T$$

and we are interested in studying the behavior of the CMDPDE

$$\widehat{\boldsymbol{\theta}}_c^{\beta} = \left(\widehat{\mu}_{1,c}^{\beta}, \widehat{\mu}_{2,c}^{\beta}, \widehat{\mu}_{3,c}^{\beta}, \widehat{\mu}_{4,c}^{\beta}, \widehat{\rho}_c^{\beta}\right)^T$$

as well as the behavior of the Wald-type test statistics for testing

$$H_0 : \rho = \rho_0 \quad \text{against} \quad H_1 : \rho \neq \rho_0. \tag{28}$$

Through $R = 10,000$ replications of the simulation experiment we compare, for different values of β, the corresponding CMDPDE through the root of the mean square errors (RMSE), when the true value of the parameters is $\boldsymbol{\theta}^* = (0, 0, 0, 0, \rho^*)$ and $\rho^* \in \{-0.1, 0, 0.15\}$. We pay special attention to the problem of the existence of some outliers in the sample, generating a 5% of the samples with $\tilde{\boldsymbol{\theta}} = (1, 3, -2, -1, \tilde{\rho})$ and $\tilde{\rho} \in \{-0.15, 0.1, 0.2\}$, respectively. Notice that, although the case $\rho^* = 0$ has been considered, this case is less important since taking into account the way of the theoretical model under consideration and having the case of independent observations, the composite likelihood theory is useless. Results are presented in Table 1 and Table 2. Two points deserve our attention. The first one is that, as expected, RMSEs for contaminated data are always greater than RMSEs for pure data and that the RMSEs decrease when the sample size n increases. The second is that, while in pure data RMSEs are greater for big values of β, when working with contaminated data the CMDPDE with medium-low values of β ($\beta \in \{0.1, 0.2, 0.3\}$) present the best behavior in terms of efficiency.

For a nominal size $\alpha = 0.05$, with the model under the null hypothesis given in (28), the estimated significance levels for different Wald-type test statistics are given by

$$\widehat{\alpha}_n^{(\beta)}(\rho_0) = \widehat{\Pr}(W_n^{\beta} > \chi^2_{1,0.05}|H_0) = \frac{\sum_{i=1}^{R} I(W_{n,i}^{\beta}) > \chi^2_{1,0.05}|\rho_0)}{R},$$

with $I(S)$ being the indicator function (with value 1 if S is true and 0 otherwise). Empirical levels with the same previous parameter values are presented in Table 3 (pure data) and Table 4 (5% of outliers). While medium-high values of β are not recommended at all, CMLE is the best when working with pure data. However the lack of robustness of CMLE test is impressive, as it can be seen in Table 4. The effect of contamination in medium-low values of β is much lighter, while for medium-high values of β it can return deceptively beneficial.

For finite sample sizes and nominal size $\alpha = 0.05$, the simulated powers are obtained under H_1 in (28), when $\rho^* \in \{-0.1, 0, 0.1\}$, $\tilde{\rho} = 0.2$ and $\rho_0 = 0.15$ (Table 5 and Table 6). The (simulated) power for different composite Wald-type test statistics is obtained by

$$\beta_n^{(\beta)}(\rho_0, \rho^*) = \Pr(W_n^{\beta} > \chi^2_{1,0.05}|H_1) \text{ and } \widehat{\beta}_n^{(\lambda)}(\rho_0, \rho^*) = \frac{\sum_{i=1}^{R} I(W_{n,i}^{\beta} > \chi^2_{1,0.05}|\rho_0, \rho^*)}{R}.$$

As expected, when we get closer to the null hypothesis and when decreasing the sample sizes, the power decreases. With pure data the best behavior is obtained with $\beta = 0$ and with contaminated data the best results are obtained for medium values of β.

Table 1. RMSEs for pure data

	$n = 100$			$n = 200$			$n = 300$		
	$\rho = -0.1$	$\rho = 0$	$\rho = 0.15$	$\rho = -0.1$	$\rho = 0$	$\rho = 0.15$	$\rho = -0.1$	$\rho = 0$	$\rho = 0.15$
$\beta = 0$	0.0958	0.0950	0.0948	0.0683	0.0668	0.0666	0.0553	0.0552	0.0551
$\beta = 0.1$	0.0972	0.0961	0.0966	0.0693	0.0676	0.0677	0.0560	0.0559	0.0561
$\beta = 0.2$	0.1009	0.0991	0.1007	0.0718	0.0697	0.0704	0.0581	0.0575	0.0585
$\beta = 0.3$	0.1061	0.1034	0.1062	0.0754	0.0727	0.0742	0.0612	0.0599	0.0619
$\beta = 0.4$	0.1123	0.1087	0.1127	0.0797	0.0762	0.0787	0.0649	0.0628	0.0659
$\beta = 0.5$	0.1195	0.1147	0.1200	0.0845	0.0803	0.0837	0.0691	0.0661	0.0702
$\beta = 0.6$	0.1274	0.1215	0.1280	0.0898	0.0848	0.0892	0.0737	0.0697	0.0748
$\beta = 0.7$	0.1361	0.1291	0.1369	0.0955	0.0897	0.0952	0.0786	0.0736	0.0797
$\beta = 0.8$	0.1456	0.1374	0.1467	0.1015	0.0905	0.1016	0.0839	0.0778	0.0849

Table 2. RMSEs for contaminated data

	$n = 100$			$n = 200$			$n = 300$		
	$\rho = -0.1$	$\rho = 0$	$\rho = 0.15$	$\rho = -0.1$	$\rho = 0$	$\rho = 0.15$	$\rho = -0.1$	$\rho = 0$	$\rho = 0.15$
$\beta = 0$	0.1371	0.1336	0.1287	0.121	0.1167	0.1113	0.1144	0.1098	0.1047
$\beta = 0.1$	0.1105	0.1104	0.1081	0.0875	0.0874	0.0843	0.0778	0.0786	0.0748
$\beta = 0.2$	0.1061	0.1053	0.1047	0.0783	0.0777	0.0759	0.0660	0.0669	0.0643
$\beta = 0.3$	0.1091	0.1072	0.1083	0.0783	0.0766	0.0761	0.0646	0.0645	0.0635
$\beta = 0.4$	0.1147	0.1118	0.1146	0.0814	0.0788	0.0798	0.0668	0.0657	0.0665
$\beta = 0.5$	0.1215	0.1176	0.1220	0.0858	0.0823	0.0848	0.0703	0.0683	0.0709
$\beta = 0.6$	0.1292	0.1242	0.1302	0.0907	0.0864	0.0905	0.0744	0.0716	0.0758
$\beta = 0.7$	0.1375	0.1315	0.1391	0.0961	0.0911	0.0966	0.0790	0.0753	0.0810
$\beta = 0.8$	0.1465	0.1396	0.1486	0.1018	0.0962	0.1031	0.0838	0.0794	0.0863

4. Conclusions

The likelihood function is the basis of the maximum likelihood method in estimation theory and it also plays a key role in the development of log-likelihood ratio tests. However, it is not so tractable in many cases, in practice. Maximum likelihood estimators are based on the likelihood function and they can be easily obtained, however, there are cases where they do not exist or they cannot

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1
Peer-reviewed version available at *Entropy* **2018**, *20*, 18; doi:10.3390/e20010018

Table 3. Levels for pure data

	$n = 100$			$n = 200$			$n = 300$		
	$\rho_0 = -0.1$	$\rho_0 = 0$	$\rho_0 = 0.15$	$\rho_0 = -0.1$	$\rho_0 = 0$	$\rho_0 = 0.15$	$\rho_0 = -0.1$	$\rho_0 = 0$	$\rho_0 = 0.15$
$\beta = 0$	0.067	0.059	0.070	0.068	0.046	0.062	0.072	0.045	0.075
$\beta = 0.1$	0.067	0.060	0.072	0.062	0.046	0.070	0.085	0.045	0.079
$\beta = 0.2$	0.072	0.061	0.084	0.069	0.051	0.084	0.097	0.049	0.102
$\beta = 0.3$	0.081	0.062	0.093	0.084	0.053	0.100	0.112	0.051	0.121
$\beta = 0.4$	0.094	0.069	0.099	0.103	0.055	0.111	0.127	0.055	0.142
$\beta = 0.5$	0.105	0.071	0.111	0.118	0.056	0.122	0.149	0.051	0.155
$\beta = 0.6$	0.122	0.083	0.129	0.131	0.062	0.136	0.167	0.051	0.165
$\beta = 0.7$	0.135	0.088	0.141	0.139	0.063	0.146	0.181	0.055	0.177
$\beta = 0.8$	0.153	0.099	0.158	0.151	0.071	0.156	0.198	0.056	0.179

Table 4. Levels for contaminated data

	$n = 100$			$n = 200$			$n = 300$		
	$\rho_0 = -0.1$	$\rho_0 = 0$	$\rho_0 = 0.15$	$\rho_0 = -0.1$	$\rho_0 = 0$	$\rho_0 = 0.15$	$\rho_0 = -0.1$	$\rho_0 = 0$	$\rho_0 = 0.15$
$\beta = 0$	0.357	0.223	0.081	0.638	0.429	0.155	0.788	0.623	0.24 0
$\beta = 0.1$	0.121	0.113	0.056	0.207	0.191	0.077	0.287	0.284	0.100
$\beta = 0.2$	0.065	0.074	0.048	0.066	0.099	0.049	0.086	0.129	0.059
$\beta = 0.3$	0.057	0.067	0.071	0.057	0.066	0.059	0.065	0.077	0.073
$\beta = 0.4$	0.075	0.066	0.087	0.067	0.058	0.081	0.079	0.060	0.095
$\beta = 0.5$	0.090	0.062	0.107	0.080	0.061	0.110	0.105	0.051	0.128
$\beta = 0.6$	0.096	0.063	0.126	0.095	0.063	0.131	0.117	0.049	0.151
$\beta = 0.7$	0.109	0.073	0.137	0.101	0.061	0.141	0.127	0.047	0.159
$\beta = 0.8$	0.125	0.083	0.147	0.109	0.061	0.149	0.141	0.049	0.171

Table 5. Powers for pure data, $\rho^* = 0.15$

	$n = 100$			$n = 200$			$n = 300$		
	$\rho_0 = -0.1$	$\rho_0 = 0$	$\rho_0 = 0.1$	$\rho_0 = -0.1$	$\rho_0 = 0$	$\rho_0 = 0.1$	$\rho_0 = -0.1$	$\rho_0 = 0$	$\rho_0 = 0.1$
$\beta = 0$	0.945	0.603	0.141	1	0.871	0.180	1	0.962	0.265
$\beta = 0.1$	0.954	0.588	0.157	1	0.863	0.207	1	0.96	0.299
$\beta = 0.2$	0.952	0.557	0.158	1	0.825	0.213	1	0.944	0.315
$\beta = 0.3$	0.941	0.510	0.153	0.999	0.783	0.213	1	0.913	0.313
$\beta = 0.4$	0.925	0.465	0.154	0.999	0.734	0.210	1	0.885	0.301
$\beta = 0.5$	0.904	0.424	0.159	0.996	0.677	0.202	1	0.845	0.289
$\beta = 0.6$	0.873	0.395	0.153	0.990	0.618	0.197	0.999	0.789	0.277
$\beta = 0.7$	0.830	0.361	0.153	0.985	0.555	0.183	0.999	0.733	0.261
$\beta = 0.8$	0.789	0.322	0.161	0.974	0.499	0.179	0.997	0.678	0.246

Table 6. Powers for contaminated data, $\rho^* = 0.15$

	$n = 100$			$n = 200$			$n = 300$		
	$\rho_0 = -0.1$	$\rho_0 = 0$	$\rho_0 = 0.1$	$\rho_0 = -0.1$	$\rho_0 = 0$	$\rho_0 = 0.1$	$\rho_0 = -0.1$	$\rho_0 = 0$	$\rho_0 = 0.1$
$\beta = 0$	0.424	0.090	0.029	0.746	0.141	0.030	0.919	0.246	0.037
$\beta = 0.1$	0.716	0.222	0.041	0.954	0.397	0.029	0.994	0.569	0.037
$\beta = 0.2$	0.838	0.333	0.071	0.989	0.555	0.075	0.999	0.744	0.096
$\beta = 0.3$	0.881	0.383	0.105	0.993	0.633	0.121	0.999	0.803	0.161
$\beta = 0.4$	0.879	0.393	0.129	0.993	0.642	0.150	0.999	0.809	0.213
$\beta = 0.5$	0.865	0.381	0.135	0.992	0.621	0.168	0.999	0.797	0.241
$\beta = 0.6$	0.836	0.357	0.149	0.984	0.583	0.174	0.998	0.769	0.252
$\beta = 0.7$	0.808	0.332	0.146	0.980	0.531	0.173	0.997	0.713	0.256
$\beta = 0.8$	0.773	0.309	0.152	0.961	0.487	0.173	0.995	0.657	0.243

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1

Peer-reviewed version available at Entropy 2018, 20, 18; doi:10.3390/e20010018

be obtained. In such a case, composite likelihood methods constitute an appealing methodology in the area of estimation and testing of hypotheses. On the other hand, distance or divergence based on methods of estimation and testing have increasingly become fundamental tools in the field of mathematical statistics. The work in [15] is the first, to the best of our knowledge, which links the notion of composite likelihood with divergence based on methods for testing statistical hypotheses.

In this paper, MDPDE are introduced and they are exploited to develop Wald type test statistics for testing simple or composite null hypotheses, in a composite likelihood framework. The validity of the proposed procedures is investigated by means of simulations. The simulation results point out the robustness of the proposed information theoretic procedures in estimation and testing, in the composite likelihood context. There are several areas where the notions of divergence and composite likelihood are crucial, including spatial statistics and time series analysis. These are areas of interest and they will be maybe explored elsewhere.

Acknowledgments: This research is supported by Grant MTM2015-67057-P, from Ministerio de Economia y Competitividad (Spain).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MLE	Maximum likelihood estimator
CMLE	Composite maximum likelihood estimator
DPD	Density power divergence
MDPDE	Minimum density power divergence estimator
CMDPDE	Composite minimum density power divergence estimator

Appendix Proof of Results

Appendix A.1 Proof of Theorem 2

The result follows in a straightforward manner because of the asymptotic normality of $\widehat{\boldsymbol{\theta}}_c^{\beta}$,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_c^{\beta} - \boldsymbol{\theta}_0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{H}_{\beta}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{J}_{\beta}(\boldsymbol{\theta}_0)\boldsymbol{H}_{\beta}^{-1}(\boldsymbol{\theta}_0)\right).$$

Appendix A.2 Proof of Theorem 3

A first order Taylor expansion of $l(\boldsymbol{\theta})$ at $\widehat{\boldsymbol{\theta}}_c^{\beta}$ around $\boldsymbol{\theta}^*$ gives

$$l\left(\widehat{\boldsymbol{\theta}}_c^{\beta}\right) - l(\boldsymbol{\theta}^*) = \left(\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\left(\widehat{\boldsymbol{\theta}}_c^{\beta} - \boldsymbol{\theta}^*\right) + o_p\left(\left\|\widehat{\boldsymbol{\theta}}_c^{\beta} - \boldsymbol{\theta}^*\right\|\right).$$

Now the result follows because the asymptotic distribution of $\left(l\left(\widehat{\boldsymbol{\theta}}_c^{\beta}\right) - l(\boldsymbol{\theta}^*)\right)$ coincides with the asymptotic distribution of $\sqrt{n}\left(\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\left(\widehat{\boldsymbol{\theta}}_c^{\beta} - \boldsymbol{\theta}^*\right)$.

Appendix A.3 Proof of Theorem 5

We have

$$\begin{aligned} \boldsymbol{g}(\widehat{\boldsymbol{\theta}}_c^{\beta}) &= \boldsymbol{g}(\boldsymbol{\theta}_0) + \boldsymbol{G}(\boldsymbol{\theta}_0)^T\left(\widehat{\boldsymbol{\theta}}_c^{\beta} - \boldsymbol{\theta}_0\right) + o_p\left(\left\|\widehat{\boldsymbol{\theta}}_c^{\beta} - \boldsymbol{\theta}_0\right\|\right) \\ &= \boldsymbol{G}^T(\boldsymbol{\theta}_0)\left(\widehat{\boldsymbol{\theta}}_c^{\beta} - \boldsymbol{\theta}_0\right) + o_p\left(\left\|\widehat{\boldsymbol{\theta}}_c^{\beta} - \boldsymbol{\theta}_0\right\|\right), \end{aligned}$$

because $\boldsymbol{g}(\boldsymbol{\theta}_0) = \mathbf{0}_r$.

Therefore

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1

Peer-reviewed version available at Entropy 2018, 20, 18; doi:10.3390/e20010018

$$\sqrt{n}\boldsymbol{g}\left(\widehat{\boldsymbol{\theta}}_c^{\beta}\right) \underset{n\longrightarrow\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(\mathbf{0}, \boldsymbol{G}_\beta(\boldsymbol{\theta}_0)^T \boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{J}_\beta(\boldsymbol{\theta}_0)\boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}_\beta(\boldsymbol{\theta}_0))$$

because

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_c^{\beta} - \boldsymbol{\theta}_0\right) \underset{n\longrightarrow\infty}{\xrightarrow{\mathcal{L}}} N(\mathbf{0}, \boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{J}_\beta(\boldsymbol{\theta}_0)\boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0)).$$

Now

$$W_{n,\beta} = n\boldsymbol{g}\left(\widehat{\boldsymbol{\theta}}_\beta\right)^T \left[\boldsymbol{G}^T(\boldsymbol{\theta}_0)\boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{J}_\beta(\boldsymbol{\theta}_0)\boldsymbol{H}_\beta^{-1}(\boldsymbol{\theta}_0)\boldsymbol{G}(\boldsymbol{\theta}_0)\right]^{-1} \boldsymbol{g}\left(\widehat{\boldsymbol{\theta}}_\beta\right) \underset{n\longrightarrow\infty}{\xrightarrow{\mathcal{L}}} \chi_r^2.$$

Appendix A.4 Computation of Sensitivity and Variability Matrices in the Numerical Example

We want to compute

$$\begin{aligned}
\boldsymbol{H}_\beta(\boldsymbol{\theta}) &= \int_{\mathbb{R}^m} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})d\boldsymbol{y} \\
\boldsymbol{J}_\beta(\boldsymbol{\theta}) &= \int_{\mathbb{R}^m} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{2\beta+1}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})d\boldsymbol{y} \\
&\quad - \int_{\mathbb{R}^m} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})d\boldsymbol{y} \int_{\mathbb{R}^m} (\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}))^T \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}d\boldsymbol{y}.
\end{aligned}$$

First of all, we can see that

$$\begin{aligned}
\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1} &= \left(f_{A_1}(\boldsymbol{\theta},\boldsymbol{y})f_{A_2}(\boldsymbol{\theta},\boldsymbol{y})\right)^{\beta+1} \\
&= \left(\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\tfrac{1}{2(1-\rho^2)}Q(y_1,y_2)\right\}\cdot\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\tfrac{1}{2(1-\rho^2)}Q(y_3,y_4)\right\}\right)^{\beta+1} \\
&= \left(\frac{1}{(2\pi)^2(1-\rho^2)}\right)^{\beta+1}\exp\left\{-\tfrac{\beta+1}{2(1-\rho^2)}\left[Q(y_1,y_2)+Q(y_3,y_4)\right]\right\} \\
&= \frac{1}{(\beta+1)^2}\left(\frac{1}{(2\pi)^2(1-\rho^2)}\right)^{\beta}\frac{(\beta+1)^2}{(2\pi)^2(1-\rho^2)}\exp\left\{-\tfrac{\beta+1}{2(1-\rho^2)}\left[Q(y_1,y_2)+Q(y_3,y_4)\right]\right\} \\
&= C_\beta\cdot\mathcal{CL}_\beta^*,
\end{aligned}$$

where $C_\beta = \frac{1}{(\beta+1)^2}\left(\frac{1}{(2\pi)^2(1-\rho^2)}\right)^\beta$ and $\mathcal{CL}_\beta^* = \mathcal{CL}_\beta(\boldsymbol{\theta},\boldsymbol{y})^* \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}^*)$, with $\boldsymbol{\Sigma}^* = \frac{1}{\beta+1}\boldsymbol{\Sigma}$.

While $\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}) = \dfrac{\partial\log\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})}{\partial\boldsymbol{\theta}}$ we will denote as $\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*$ to $\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^* = \dfrac{\partial\log\mathcal{CL}_\beta^*}{\partial\boldsymbol{\theta}}$. Then

$$\begin{aligned}
\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}) &= \frac{\partial\log\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})}{\partial\boldsymbol{\theta}} = \frac{1}{\beta+1}\frac{\partial\log\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}}{\partial\boldsymbol{\theta}} = \frac{1}{\beta+1}\frac{\partial\log(C_\beta\cdot\mathcal{CL}_\beta^*)}{\partial\boldsymbol{\theta}} \\
&= \frac{1}{\beta+1}\left(\frac{\partial\log C_\beta}{\partial\boldsymbol{\theta}}+\frac{\partial\log\mathcal{CL}_\beta^*}{\partial\boldsymbol{\theta}}\right) = \frac{1}{\beta+1}\left(\frac{\partial\log C_\beta}{\partial\boldsymbol{\theta}}+\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*\right).
\end{aligned} \tag{29}$$

Further,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1

Peer-reviewed version available at Entropy 2018, 20, 18; doi:10.3390/e20010018

$$\begin{aligned}\int_{\mathbb{R}^m} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})d\boldsymbol{y} &= \int_{\mathbb{R}^m} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}\frac{\partial \log \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})}{\partial \boldsymbol{\theta}}d\boldsymbol{y} = \int_{\mathbb{R}^m} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta}\frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})}{\partial \boldsymbol{\theta}}d\boldsymbol{y} \\ &= \int_{\mathbb{R}^m} \frac{1}{\beta+1}\frac{\partial \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}}{\partial \boldsymbol{\theta}}d\boldsymbol{y} = \frac{1}{\beta+1}\frac{\partial}{\partial \boldsymbol{\theta}}\int_{\mathbb{R}^m} \mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}d\boldsymbol{y} \\ &= \frac{1}{\beta+1}\frac{\partial C_\beta}{\partial \boldsymbol{\theta}} = (0,0,0,0,\frac{2\rho\beta C_\beta}{(\beta+1)(1-\rho^2)})^T = \boldsymbol{\xi}_\beta(\boldsymbol{\theta}). \end{aligned} \tag{30}$$

307 Now

$$\begin{aligned}&\int_{\mathbb{R}^4} \mathcal{CL}^{\beta+1}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})d\boldsymbol{y} \\ &= \int_{\mathbb{R}^4} (C_\beta\cdot \mathcal{CL}_\beta^*)\frac{1}{(\beta+1)^2}\left(\frac{\partial \log C_\beta}{\partial \boldsymbol{\theta}}+\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*\right)^T\left(\frac{\partial \log C_\beta}{\partial \boldsymbol{\theta}}+\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*\right)d\boldsymbol{y} \\ &= \frac{C_\beta}{(\beta+1)^2}\int_{\mathbb{R}^4}\left[\left(\frac{\partial \log C_\beta}{\partial \boldsymbol{\theta}}\right)^T\left(\frac{\partial \log C_\beta}{\partial \boldsymbol{\theta}}\right)\mathcal{CL}_\beta^*\right. \\ &\left.+\mathcal{CL}_\beta^*\left(\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*\right)^T\frac{\partial \log C_\beta}{\partial \boldsymbol{\theta}}+\mathcal{CL}_\beta^*\left(\frac{\partial \log C_\beta}{\partial \boldsymbol{\theta}}\right)^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*+\mathcal{CL}_\beta^*(\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*)^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*\right]d\boldsymbol{y} \\ &= \frac{C_\beta}{(\beta+1)^2}\left[\left(\frac{\partial \log C_\beta}{\partial \boldsymbol{\theta}}\right)^T\left(\frac{\partial \log C_\beta}{\partial \boldsymbol{\theta}}\right)\int_{\mathbb{R}^4}\mathcal{CL}_\beta^*d\boldsymbol{y}+\left(\int_{\mathbb{R}^4}\mathcal{CL}_\beta^*\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*d\boldsymbol{y}\right)^T\left(\frac{\partial \log C_\beta}{\partial \boldsymbol{\theta}}\right)\right. \\ &\left.+\left(\frac{\partial \log C_\beta}{\partial \boldsymbol{\theta}}\right)^T\int_{\mathbb{R}^4}\mathcal{CL}_\beta^*\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*d\boldsymbol{y}+\int_{\mathbb{R}^4}\mathcal{CL}_\beta^*(\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*)^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*d\boldsymbol{y}\right] \\ &= \frac{C_\beta}{(\beta+1)^2}\left[\boldsymbol{K}^T\boldsymbol{K}+\left(\int_{\mathbb{R}^4}\mathcal{CL}_\beta^*\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*d\boldsymbol{y}\right)^T\boldsymbol{K}+\boldsymbol{K}^T\int_{\mathbb{R}^4}\mathcal{CL}_\beta^*\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*d\boldsymbol{y}+\int_{\mathbb{R}^4}\mathcal{CL}_\beta^*(\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*)^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*d\boldsymbol{y}\right],\end{aligned} \tag{31}$$

308 where $\boldsymbol{K} = \frac{\partial \log C_\beta}{\partial \boldsymbol{\theta}} = (0,0,0,0,\frac{2\rho\cdot\beta}{1-\rho^2})$. But

$$\begin{aligned}\int_{\mathbb{R}^4}\mathcal{CL}_\beta^*\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*d\boldsymbol{y} &= \int_{\mathbb{R}^4}\left(\frac{1}{C_\beta}\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}\right)\left[(\beta+1)\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})-\frac{\partial \log C_\beta}{\partial \boldsymbol{\theta}}\right]d\boldsymbol{y} \\ &= \frac{\beta+1}{C_\beta}\left[\int_{\mathbb{R}^4}\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})d\boldsymbol{y}\right]-\frac{\boldsymbol{K}}{C_\beta}\int_{\mathbb{R}^4}\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}d\boldsymbol{y} \\ &= \frac{1}{C_\beta}\frac{\partial C_\beta}{\partial \boldsymbol{\theta}}-\boldsymbol{K}=\boldsymbol{K}-\boldsymbol{K}=\boldsymbol{0},\end{aligned}$$

309 and thus (31) can be expressed as

$$\int_{\mathbb{R}^4}\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})^{\beta+1}\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})d\boldsymbol{y}=\frac{C_\beta}{(\beta+1)^2}\left[\boldsymbol{K}^T\boldsymbol{K}+\int_{\mathbb{R}^4}\mathcal{CL}_\beta^*(\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*)^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*d\boldsymbol{y}\right].$$

310 On the other hand, it is not difficult to prove that

$$\int_{\mathbb{R}^4}\mathcal{CL}_\beta^*(\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*)^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})^*d\boldsymbol{y}=\boldsymbol{C}\cdot\int_{\mathbb{R}^4}\mathcal{CL}(\boldsymbol{\theta},\boldsymbol{y})(\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y}))^T\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{y})d\boldsymbol{y}=\boldsymbol{C}\cdot\boldsymbol{H}_0(\boldsymbol{\theta}),$$

311 where $\boldsymbol{C} = diag(\beta+1,\beta+1,\beta+1,\beta+1,1)$ and ([15])

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 November 2017 doi:10.20944/preprints201711.0039.v1

Peer-reviewed version available at Entropy 2018, 20, 18; doi:10.3390/e20010018

$$\boldsymbol{H}_0(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{1-\rho^2} & \frac{-\rho}{1-\rho^2} & 0 & 0 & 0 \\ \frac{-\rho}{1-\rho^2} & \frac{1}{1-\rho^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{1-\rho^2} & \frac{-\rho}{1-\rho^2} & 0 \\ 0 & 0 & \frac{-\rho}{1-\rho^2} & \frac{1}{1-\rho^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{2(\rho^2+1)}{(1-\rho^2)^2} \end{pmatrix}. \quad (32)$$

So

$$\boldsymbol{H}_\beta(\boldsymbol{\theta}) = \frac{C_\beta}{(\beta+1)^2}\left[\boldsymbol{C}\cdot\boldsymbol{H}_0(\boldsymbol{\theta}) + \boldsymbol{K}^T\boldsymbol{K}\right],$$

this is

$$\boldsymbol{H}_\beta(\boldsymbol{\theta}) = \frac{C_\beta}{(\beta+1)(1-\rho^2)} \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 \\ -\rho & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -\rho & 0 \\ 0 & 0 & -\rho & 1 & 0 \\ 0 & 0 & 0 & 0 & 2\frac{(\rho^2+1)+2\rho^2\beta^2}{(1-\rho^2)(1+\beta)} \end{pmatrix}. \quad (33)$$

Note that, for $\beta = 0$, (33) equals to (32).

On the other hand, the expression of the variability matrix $\boldsymbol{J}_\beta(\boldsymbol{\theta})$ can be obtained from expressions (26) and (30) as

$$\boldsymbol{J}_\beta(\boldsymbol{\theta}) = \boldsymbol{H}_{2\beta}(\boldsymbol{\theta}) - \boldsymbol{\xi}_\beta(\boldsymbol{\theta})^T\boldsymbol{\xi}_\beta(\boldsymbol{\theta}). \quad (34)$$

References

1. Basu, A.; Harris, I.R.; Hjort, N.L. and Jones, M.C. Robust and efficient estimation by minimizing a density power divergence. *Biometrika*, **1998**, *85*, 549–559.
2. Basu, A.; Mandal, A.; Martín, N. and Pardo, L. Testing statistical hypotheses based on the density power divergence. *Ann. Inst. Stat. Math.*, **2013**, *65*, 319–348
3. Basu, A.; Mandal, A.; Martín, N. and Pardo, L. Robust tests for the equality of two normal means based on the density power divergence. *Metrika*, **2015**, *78*, 611–634.
4. Basu, A.; Mandal, A.; Martín, N. and Pardo, L. Generalized Wald-type tests based on minimum density power divergence estimators. *Statistics*, **2016**, *50*, 1, 1-26.
5. Basu, A.; Ghosh, A. Mandal; Martín, N. and Pardo, L. A Wald-type test statistic for testing linear hypothesis in logistic regression models based on minimum density power divergence estimator. *Electon. J. Stat.*, **2017**, *11*, 2, 2741–2772.
6. Ghosh, A.; Mandal, A.; Martín, N. and Pardo, L. Influence analysis of robust Wald-type tests. *J. Multivariate Anal.*, **2016**, *147*, 102–126.
7. Varin, C.; Reid, N. and Firth, D. An overview of composite likelihood methods. *Stat. Sin.*, **2011**, *21*, 1, 4-42.
8. Xu, X. and Reid, N. On the robustness of maximum composite estimate. *J. Stat. Plan. Inference.*, **2011**, *141*, 3047-3054.
9. Joe, H., Reid, N.; Somg, P.X.; Firth, D. and Varin, C. Composite likelihood methods. *Report on the Workshop on Composite Likelihood*. **2012** Available at http://www.birs.ca/events/2012/5-day-workshops/12w5046.
10. Lindsay, G. Composite likelihood methods. *Contemp. Math.*, **1998**, *80*, 221-239.
11. Basu, A.; Shioya, H. and Park, C. *Statistical inference. The minimum distance approach*. Chapman & Hall/CRC. Boca Raton, 2011.
12. Maronna, R. A., Martin, R. D. and Yohai, V. J. *Time Series, in Robust Statistics: Theory and Methods*, John Wiley & Sons, Ltd, Chichester, UK., 2006.
13. Pardo, L. *Statistical inference based on divergence measures*. Chapman & Hall/CRC. Boca Raton, 2006.
14. Serfling, Robert J. *Approximation Theorems of Mathematical Statistics*. New York: Wiley, 1980.

15. Martín, N.; Pardo, L. and Zografos, K. On divergence tests for composite hypotheses under composite likelihood. *Stat. Pap.*, **2017** . Available online: https://doi.org/10.1007/s00362-017-0900-1.