

Small Sample Inference for Probabilistic Index Models

G. Amorim^{a,1,*}, O. Thas^{a,b}, K. Vermeulen^a, S. Vansteelandt^{c,d}, J. De Neve^e

^aDepartment of Mathematical Modelling Statistics and Bioinformatics, Ghent University

^bNational Institute for Applied Statistics Research Australia (NIASRA), School of Mathematics and Applied Statistics, University of Wollongong

^cDepartment of Applied Mathematics, Computer Science and Statistics, Ghent University

^dCentre for Statistical Methodology, London School of Hygiene and Tropical Medicine

^eDepartment of Data Analysis, Ghent University

Abstract

Probabilistic index models may be used to generate classical and new rank tests, with the additional advantage of supplementing them with interpretable effect size measures. The popularity of rank tests for small sample inference makes probabilistic index models also natural candidates for small sample studies. However, at present, inference for such models relies on asymptotic theory that can deliver poor approximations of the sampling distribution if the sample size is rather small. A bias-reduced version of the bootstrap and adjusted jackknife empirical likelihood are explored. It is shown that their application leads to drastic improvements in small sample inference for probabilistic index models, justifying the use of such models for reliable and informative statistical inference in small sample studies.

Keywords: Bootstrap; Empirical likelihood; Rank estimation.

1. Introduction

Probabilistic index models were introduced by Thas et al. (2012) as a class of semiparametric models that can be used to complement rank tests with interpretable effect size measures, without the need to assume a location-shift model. Bergsma et al. (2009) described similar models, which they referred to as *Bradley-Terry type models*. For more details on the connection between probabilistic index models and their Bradley-Terry type models, we refer to Bergsma et al. (2012). Probabilistic index models can also be used to generate and extend many of the well-known rank tests such as, for example, the Wilcoxon–Mann–Whitney test, Kruskal–Wallis and Friedman tests (De Neve and Thas, 2015).

A probabilistic index model parameterises the conditional probabilistic index

$$\text{pr}(Y \preceq Y^* \mid \mathbf{X}, \mathbf{X}^*) = \text{pr}(Y < Y^* \mid \mathbf{X}, \mathbf{X}^*) + 0.5\text{pr}(Y = Y^* \mid \mathbf{X}, \mathbf{X}^*),$$

in which Y and Y^* are outcomes associated with the covariates \mathbf{X} and \mathbf{X}^* , respectively, with (Y, \mathbf{X}) and (Y^*, \mathbf{X}^*) identically and independently distributed random vectors. The covariate may be vector-valued. The probabilistic index model is defined as

$$\text{pr}(Y \preceq Y^* \mid \mathbf{X}, \mathbf{X}^*) = m(\mathbf{X}, \mathbf{X}^*; \beta) = g^{-1}(\mathbf{Z}^T \beta) \quad (\mathbf{X}, \mathbf{X}^*) \in \mathcal{X}, \quad (1)$$

where β is a p -dimensional parameter vector and $m(\mathbf{X}, \mathbf{X}^*; \beta)$ is a known function that satisfies $0 \leq m(\mathbf{X}, \mathbf{X}^*; \beta) \leq 1$, $m(\mathbf{X}, \mathbf{X}; \beta) = 0.5$ and $m(\mathbf{X}, \mathbf{X}^*; \beta) + m(\mathbf{X}^*, \mathbf{X}; \beta) = 1$ for all $(\mathbf{X}, \mathbf{X}^*) \in \mathcal{X}$. The vector \mathbf{Z} depends on the regressors, e.g. $\mathbf{Z} = \mathbf{X}^* - \mathbf{X}$. The model thus takes the form of a generalized linear model by relating $\mathbf{Z}^T \beta$ to the

*Corresponding author

Email address: ggca@outlook.com (G. Amorim)

¹Dept. of Mathematical Modelling, Statistics and Bioinformatics Ghent University Coupure Links 653 B-9000 Gent, Belgium

conditional probabilistic index through a known link function $g(\cdot)$. The set \mathcal{X} is the subset of covariate pairs $(\mathbf{X}, \mathbf{X}^*)$ for which the probabilistic index model is defined.

Estimation of β in (1) was discussed in detail in Thas et al. (2012). Given a random sample of identically and independently distributed (Y_i, \mathbf{X}_i) , $i = 1, \dots, n$, the outcomes are first transformed to so-called pseudo-observations I_{ij} defined as 1 if $Y_i < Y_j$, 1/2 if $Y_i = Y_j$ and 0 otherwise ($i, j = 1, \dots, n$). For each pair $(\mathbf{X}_i, \mathbf{X}_j)$ the vector \mathbf{Z}_{ij} is constructed. An estimator of β , say $\hat{\beta}$, is then obtained by solving $\mathbf{U}_n(\beta) = \mathbf{0}$, with

$$\mathbf{U}_n(\beta) = \frac{1}{|\mathcal{I}_n|} \sum_{(i,j) \in \mathcal{I}_n} \mathbf{U}_{ij}(\beta) = \frac{1}{|\mathcal{I}_n|} \sum_{(i,j) \in \mathcal{I}_n} \mathbf{A}(\mathbf{Z}_{ij}; \beta) \{I_{ij} - g^{-1}(\mathbf{Z}_{ij}^T \beta)\}, \quad (2)$$

where $\mathbf{A}(\mathbf{Z}_{ij}; \beta)$ is a p -dimensional vector function of the regressors \mathbf{Z}_{ij} and the parameter vector β , and the sum is limited to all pairs (i, j) for which $(\mathbf{X}_i, \mathbf{X}_j)$ are in \mathcal{X} (the set of such index pairs is denoted by \mathcal{I}_n and has $|\mathcal{I}_n|$ elements). Under mild regularity conditions, the solution $\hat{\beta}$ is consistent and asymptotically normal (Thas et al., 2012; De Neve, 2013), with covariance matrix that can be consistently estimated by a sandwich estimator, which is given by

$$\hat{\Sigma}_{\hat{\beta}} = \left\{ \sum_{(i,j) \in \mathcal{I}_n} \frac{\partial \mathbf{U}_{ij}(\hat{\beta})}{\partial \beta^T} \right\}^{-1} \left\{ \sum_{(i,j) \in \mathcal{I}_n} \sum_{(k,l) \in \mathcal{I}_n} \phi_{ijkl} \mathbf{U}_{ij}(\hat{\beta}) \mathbf{U}_{kl}^T(\hat{\beta}) \right\} \left\{ \sum_{(i,j) \in \mathcal{I}_n} \frac{\partial \mathbf{U}_{ij}(\hat{\beta})}{\partial \beta^T} \right\}^{-1},$$

where ϕ_{ijkl} is an indicator variable taking value 1 if the pseudo-outcomes I_{ij} and I_{kl} share an index, and 0 otherwise.

Simulation studies in Thas et al. (2012) and De Neve (2013) confirm the asymptotic distribution theory. However, even with only two regressors, their results also indicate that $\hat{\beta}$, the sandwich estimator and the coverages of asymptotic Wald-based confidence intervals are only reliable for sample sizes of 50 or more. This is an important limitation, particularly because the above methods provide natural extensions of rank tests by, for example, allowing inference for treatment effects while controlling for covariates (De Neve and Thas, 2015; Vermeulen et al., 2015). Only for the special case of a Wilcoxon-rank sum test in randomized experiments, a covariate adjustment, based on a probabilistic index model, has been proposed for which permutation p -values are available (Vermeulen et al., 2015).

To overcome the aforementioned limitations, we explore methods that are designed to give better small sample results. Resampling techniques, such as the bootstrap and jackknife, are often used as alternative approaches to increase accuracy in many statistical applications (Basu, 2001). However, they sometimes require strong computational power. For instance, direct application of the traditional non-parametric bootstrap to probabilistic index models requires solving B times (number of bootstrap samples) the nonlinear estimating equation $\mathbf{U}_n(\beta) = \mathbf{0}$ for β . Even solving the equation only once may already be computationally demanding because fitting a probabilistic index model requires modelling the pseudo-observations, resulting in an inflated number of estimating functions. We solve this issue by applying the bootstrapping U -statistics method of Jiang and Kalbfleisch (2012). This method simplifies the computational demands by resampling properly studentized terms from an asymptotic approximation of the estimating function that is a U -statistic of degree 1 or 2. It hence avoids the need to repeatedly solve nonlinear equations and our simulation results show that the resulting coverages are often close to the nominal values.

In addition to bootstrap, we also use methods based on empirical likelihood to improve small sample inference for probabilistic index models. The empirical likelihood method (Owen, 1988, 1990) maximizes a non-parametric likelihood subject to restrictions given by the estimating equations. The ratio of the maximized empirical likelihood over the maximized nonparametric likelihood, which corresponds to an unconstrained model, is known as the empirical likelihood ratio statistic for which a Wilks' theorem needs to be proven. Just as for the parametric likelihood ratio statistic, this Wilks' theorem gives the asymptotic distribution under the hypothesis that the constrained model holds true. Confidence intervals of the parameters are subsequently found by inverting the empirical likelihood ratio test. For U -statistics, the model restrictions are non-linear, leading to a computationally expensive estimation process. The jackknife empirical likelihood method (Jing et al., 2009) reduces this computational cost by rewriting the U -statistic as a sum of asymptotically linear independent terms, making the constraints linear. Its use for probabilistic index models was also suggested by Zhou (2012). However, this method, as any empirical likelihood method, requires that the constraints always have a solution, which is not necessarily true. Chen et al. (2008) proposed to adjust the empirical likelihood by including an artificial "pseudo-observation" so that a solution can always be obtained and Zhao et al. (2015) later adapted this approach to the jackknife empirical likelihood setting. In this paper we further

adapt the adjusted jackknife empirical likelihood to the probabilistic index model setting and evaluate its performance in simulation studies. This method performs generally well, but may be strongly affected by finite-sample bias. To alleviate this problem, we propose a bias-reduced adjusted jackknife empirical likelihood approach that shows good empirical results for samples as small as 20 and with coverages close to the nominal values.

2. Bias-reduced estimator for probabilistic index models

Consider the estimating function (2) with

$$\mathbf{A}(\mathbf{Z}_{ij}; \beta) = \frac{\partial}{\partial \beta} m(\mathbf{Z}_{ij}^T \beta) V^{-1} \{m(\mathbf{Z}_{ij}^T \beta)\}, \quad (3)$$

in which $V\{m(\mathbf{Z}_{ij}^T \beta)\} = m(\mathbf{Z}_{ij}^T \beta)\{1 - m(\mathbf{Z}_{ij}^T \beta)\}$ is equal to the variance of I_{ij} , under the assumption that the probabilistic index model (1) holds. Note that we treat the pseudo-observations as independent, although this does not hold true for all pseudo-observations, e.g., I_{ij} and I_{ik} , $k = i + 1, \dots, n$, are correlated because they share the outcome Y_i . This simplification, however, does not affect the consistency nor asymptotic normality of the resulting estimator of β .

The presence of β in $\mathbf{A}(\mathbf{Z}_{ij}; \beta)$ implies that $V^{-1}\{m(\mathbf{Z}_{ij}^T \beta)\}$ has to be recalculated in each iteration of the optimization procedure when solving $\mathbf{U}_n(\beta) = \mathbf{0}$. One particular choice that will be used in this paper is $V\{m(\mathbf{Z}_{ij}^T \beta_c)\} = m(\mathbf{Z}_{ij}^T \beta_c)\{1 - m(\mathbf{Z}_{ij}^T \beta_c)\}$, which is further denoted by V_c , and where β_c is a constant vector that does not change from one iteration to another. Apart from a reduction in computation time, the choice $\beta_c = \mathbf{0}$ is of particular interest as it leads to a bias-reduced estimator for β , as given in the proposition below. Importantly, note that the bias-reduced estimator is consistent, because the consistency holds for every $\mathbf{A}(\mathbf{Z}_{ij}; \beta)$ for which the regularity conditions hold (Thas et al., 2012).

Proposition 1. *Let $\mathbf{U}_n^{(c)}(\beta)$ be the estimating function (2) with $\mathbf{A}(\mathbf{Z}_{ij}; \beta)$ as in (3), but with $V\{m(\mathbf{Z}_{ij}^T \beta)\}$ replaced by V_c . Let $\hat{\beta}_c$ be the solution of $\mathbf{U}_n^{(c)}(\beta) = \mathbf{0}$. If the link function $m(\cdot)$ is such that $m(0) = 1/2$, then the second-order bias of $\hat{\beta}_c$ is minimized at $\beta_c = \mathbf{0}$.*

Proof. Let $\mathbf{U}_n^{(c)}(\beta, \beta_c)$ be the estimating function (2) with $\mathbf{A}(\mathbf{Z}_{ij}; \beta)$ as in (3), but with $V(\beta) = V\{m(\mathbf{Z}\beta)\}$ replaced by $V_c(\beta_c) = V\{m(\mathbf{Z}\beta_c)\}$. Recall that $V_c(\beta_c)$, which depends on the constant β_c , is a scalar, as the pseudo-observations are assumed to be uncorrelated. Our goal is to find the value of β_c that minimizes the asymptotic variance of the estimator that solves $\mathbf{U}_n^{(c)}(\beta, \beta_c) = \mathbf{0}$.

Let $\mathbf{U}^{(c)}(\beta, \beta_c)$ be partitioned into $(U_1^{(c)}(\beta, \beta_c), \dots, U_p^{(c)}(\beta, \beta_c))$, where p is the dimension of β , $\kappa_{rk}(\beta, \beta_c) = E(\partial U_r^{(c)}(\beta, \beta_c) / \partial \beta_k \mid \mathbf{X}_1, \dots, \mathbf{X}_n)$, $\kappa_{rkl}(\beta, \beta_c) = E(\partial^2 U_r^{(c)}(\beta, \beta_c) / \partial \beta_k \partial \beta_l \mid \mathbf{X}_1, \dots, \mathbf{X}_n)$ and $\kappa_{rk}^{(l)}(\beta, \beta_c) = \partial \kappa_{rk}(\beta, \beta_c) / \partial \beta_l$, for $r, k, l = 1, \dots, p$. If we further denote by $\mathbf{K}(\beta, \beta_c)$ the Jacobian of the estimating function, $-\{\kappa_{rk}(\beta, \beta_c)\}$ stands for its (r, k) th element and we denote by $\kappa^{rk}(\beta, \beta_c)$ the corresponding element of $\mathbf{K}^{-1}(\beta, \beta_c)$. Following Paul and Zhang (2014), we treat $\mathbf{U}^{(c)}(\beta, \beta_c)$ as a score function so that the second-order bias $b_s(\beta, \beta_c)$ for maximum likelihood estimates derived in Cordeiro and McCullagh (1991) can be directly applied. The bias of $\hat{\beta}_c$ can then be written as

$$b_s(\beta, \beta_c) = \sum_{r=1}^p \kappa^{rs}(\beta, \beta_c) \sum_{k,l=1}^p \left[\kappa_{rk}^{(l)}(\beta, \beta_c) - \frac{1}{2} \kappa_{rkl}(\beta, \beta_c) \right] \kappa^{kl}(\beta, \beta_c), \quad (s = 1, \dots, p) \quad (4)$$

and so $b_s(\beta, \beta_c)^2$ can be minimized with respect to $\beta_{c,s}$, where $\beta_{c,s}$ is the s th component of β_c . Note that all components of $b_s(\beta, \beta_c)$, namely $\kappa_{rk}(\beta, \beta_c)$, $\kappa_{rkl}(\beta, \beta_c)$ and $\kappa_{rk}^{(l)}(\beta, \beta_c)$, depend on β_c only through $V_c(\beta_c)$. For example, as

$$\kappa_{rk}(\beta, \beta_c) = - \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{\partial m_r(\mathbf{Z}_{ij}^T \beta)}{\partial \beta} V^{-1}\{m(\mathbf{Z}_{ij}^T \beta_c)\} \frac{\partial m_r(\mathbf{Z}_{ij}^T \beta)}{\partial \beta_k},$$

we have that

$$\frac{\partial \kappa_{rk}(\beta, \beta_c)}{\partial \beta_{c,s}} = - \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{\partial m_r(\mathbf{Z}_{ij}^T \beta)}{\partial \beta} \frac{\partial V^{-1}\{m(\mathbf{Z}_{ij}^T \beta_c)\}}{\partial \beta_{c,s}} \frac{\partial m_r(\mathbf{Z}_{ij}^T \beta)}{\partial \beta_k}.$$

Thus, if we set $\partial V^{-1}\{m(\mathbf{Z}_{ij}^T \beta_c)\}/\partial \beta_{c,s}$ to zero, $\partial \kappa_{rk}(\beta, \beta_c)/\partial \beta_{c,s}$ will also be equal to zero. The same argument holds for $\partial \kappa_{rk}(\beta, \beta_c)/\partial \beta_{c,s}$ and $\partial \kappa_{rk}^{(l)}(\beta, \beta_c)/\partial \beta_{c,s}$. A trivial solution for $\partial b_s^2(\beta, \beta_c)/\partial \beta_{c,s} = 0$, that is, for $\partial b_s(\beta, \beta_c)/\partial \beta_{c,s} = 0$, is obtained by setting $\partial V^{-1}\{m(\mathbf{Z}_{ij}^T \beta_c)\}/\partial \beta_{c,s}$ to zero, for $i = 1, \dots, n-1, j = i+1, \dots, n$ and $s = 1, \dots, p$. As $V\{m(\mathbf{Z}_{ij}^T \beta_c)\} = m(\mathbf{Z}_{ij}^T \beta_c)\{1 - m(\mathbf{Z}_{ij}^T \beta_c)\}$, setting $\partial V^{-1}\{m(\mathbf{Z}_{ij}^T \beta_c)\}/\partial \beta_{c,s} = 0$ implies that

$$\frac{\partial m(\mathbf{Z}_{ij}^T \beta_c)}{\partial \beta_{c,s}} \{1 - 2m(\mathbf{Z}_{ij}^T \beta_c)\} = 0.$$

If $m(0) = 1/2$, e.g., $m(\cdot)$ is the probit or logit function, the minimum is reached at $\beta_c = \mathbf{0}$. ■

Due to its bias-reduced property, we denote by $\hat{\beta}_{br}$ the estimator that solves $\mathbf{U}_n^{(c)}(\beta) = \mathbf{0}$, with $\beta_c = \mathbf{0}$. We expect that $\hat{\beta}_{br}$, due to its smaller bias, will have better small sample properties than $\hat{\beta}$. This will be empirically evaluated later in Section 4, when Wald confidence intervals based on both estimators will be considered.

3. Bootstrapping U -Statistics and Jackknife Empirical Likelihood

3.1. Bootstrapping U -statistics

The bootstrap is known to often give better small sample inference as compared to methods relying on the asymptotic distribution of the parameter estimator and its consistent variance estimator. However, it has the drawback of being computationally intensive, especially when parameters are estimated by solving estimating equations. This is even more problematic for probabilistic index models, as they require modelling the pseudo-outcomes, leading to an inflated number of estimating functions.

When parameters are estimated by solving estimating equations, Parzen et al. (1994) showed that inference can be based on resampling the terms of the estimation function, provided that these terms are asymptotically independently distributed and pivotal, that is, their asymptotic distribution does not depend on the parameter of interest. This method is not applicable to probabilistic index models, because its estimation function is not a sum of independently distributed pivots. The method of Parzen et al. (1994) was extended by Hu and Kalbfleisch (2000); they suggested to studentize the terms to make them asymptotically pivotal. Jiang and Kalbfleisch (2012) further generalized the method to estimating functions with a U -statistic structure. They showed that asymptotically correct inference can be obtained by (i) approximating the estimating function by its empirical Hájek projection, (ii) studentizing the resulting terms so as to obtain asymptotic pivots, and (iii) appropriately resampling these terms to obtain an estimate of the sampling distribution of the pivots. From this distribution, approximate inference of the parameters can be derived. They abbreviated their method, which is applied to a studentized estimating function, as EF_t . Empirical evaluations in Jiang and Kalbfleisch (2012) for least squares regression and Wilcoxon rank regression show very good results for the coverage of confidence intervals for small sample sizes.

As in Thas et al. (2012), we set $\mathcal{I}_n = \{i, j = 1, \dots, n : i < j\}$. Upon writing the estimating function (2) as

$$\mathbf{U}_n(\beta) = \binom{2}{n}^{-1} \sum_{1 \leq i < j \leq n} \mathbf{h}(\mathbf{W}_i, \mathbf{W}_j; \beta) \quad (5)$$

with $\mathbf{W}_i^T = (Y_i, \mathbf{X}_i^T)$ and $\mathbf{h}(\mathbf{W}_i, \mathbf{W}_j; \beta) = \mathbf{A}(\mathbf{Z}_{ij}^T \beta) \{I_{ij} - g^{-1}(\mathbf{Z}_{ij}^T \beta)\}$, $\mathbf{U}_n(\beta)$ is recognized as a one-sample U -statistic with kernel function $\mathbf{h}(\cdot, \cdot; \beta)$ of degree 2. As in Jiang and Kalbfleisch (2012), we define $\mathbf{H}_i(\beta) = (n-1)^{-1} \sum_{j: j \neq i} \mathbf{h}(\mathbf{W}_i, \mathbf{W}_j; \beta)$, $i = 1, \dots, n$, which are asymptotically uncorrelated. It follows that $\mathbf{U}_n(\beta) = n^{-1} \sum_{i=1}^n \mathbf{H}_i(\beta)$ represents the estimating function as a sum of asymptotically independent terms (Hájek projection), hence allowing for straightforward variance estimation. Let $\mathbf{U}_{n,t}(\beta)$ denote the studentized $\mathbf{U}_n(\beta)$, i.e., $\mathbf{U}_{n,t}(\beta) = \{\mathbf{S}(\beta)\}^{-1/2} \mathbf{U}_n(\beta)$ where $\mathbf{S}(\beta) = 4\{n^2(n-1)\}^{-1} \sum_{1 \leq i < j \leq n} \{\mathbf{H}_i(\beta) - \mathbf{H}_j(\beta)\} \{\mathbf{H}_i(\beta) - \mathbf{H}_j(\beta)\}^T$ is the variance estimator of $\text{var}\{\mathbf{U}_n(\beta)\}$ (Jiang and Kalbfleisch, 2012). A detailed description of the EF_t -method applied to probabilistic index models is as follows:

1. Generate (M_1, \dots, M_n) from a multinomial distribution with parameters n and $(1/n, \dots, 1/n)$.
2. Compute $\hat{\mathbf{H}}_i^*(\hat{\beta}) = (n-1)^{-1} \sum_{j:j \neq i} M_j \mathbf{h}(\mathbf{W}_i, \mathbf{W}_j; \hat{\beta})$ and $\hat{\mathbf{U}}_n^*(\hat{\beta}) = n^{-1} \sum_{i=1}^n M_i \hat{\mathbf{H}}_i^*(\hat{\beta})$.
3. Compute $\mathbf{S}^*(\hat{\beta}) = 4\{n^2(n-1)\}^{-1} \sum_{1 \leq i < j \leq n} M_i M_j \{\hat{\mathbf{H}}_i^*(\hat{\beta}) - \hat{\mathbf{H}}_j^*(\hat{\beta})\} \{\hat{\mathbf{H}}_i^*(\hat{\beta}) - \hat{\mathbf{H}}_j^*(\hat{\beta})\}^T$, which is a variance estimator of $\mathbf{U}_n^*(\hat{\beta})$.
4. Compute the studentized statistic $\hat{\mathbf{U}}_{n,t}^*(\hat{\beta}) = \{\mathbf{S}^*(\hat{\beta})\}^{-1/2} \hat{\mathbf{U}}_n^*(\hat{\beta})$.
5. Repeat steps 1-4 a large number B times.

Theorem 2.3 in Jiang and Kalbfleisch (2012) states that $\hat{\mathbf{U}}_{n,t}^*(\hat{\beta})$ converges weakly to a p -variate standard normal distribution as $n \rightarrow \infty$.

For a scalar β , $1 - \alpha$ confidence intervals are defined as $\{\beta : \hat{U}_{n,t,\alpha/2} < U_{n,t}(\beta) < \hat{U}_{n,t,1-\alpha/2}\}$, where $\hat{U}_{n,t,\alpha/2}$ and $\hat{U}_{n,t,1-\alpha/2}$ are the estimated $\alpha/2$ and $1 - \alpha/2$ quantiles of $U_{n,t}(\beta)$, respectively. We can either take $\hat{U}_{n,t,\alpha/2}$ equal to the $\alpha/2$ quantile of the standard normal distribution, relying thus on the asymptotic distribution of $U_{n,t}(\beta)$, or take $\hat{U}_{n,t,\alpha/2}$ to be the empirical $\alpha/2$ quantile from the replications of the above procedure. If $U_{n,t}(\beta)$ is monotone non-decreasing in β , a confidence interval for β is $[\hat{\beta}_{\alpha/2}, \hat{\beta}_{1-\alpha/2}]$, where $\hat{\beta}_{\alpha/2}$ and $\hat{\beta}_{1-\alpha/2}$ are the solutions to $U_{n,t}(\beta) = \hat{U}_{n,t,1-\alpha/2}$ and $U_{n,t}(\beta) = \hat{U}_{n,t,\alpha/2}$, respectively.

For a p -vector β , an approximate confidence interval for its i th component β_i is defined as $\{\beta_i : \beta_{i,\alpha/2}^* < \beta_i < \beta_{i,1-\alpha/2}^*\}$, for $i = 1, \dots, p$. Here, $\beta_{i,\alpha/2}^*$ and $\beta_{i,1-\alpha/2}^*$ are the empirical $\alpha/2$ and $(1 - \alpha/2)$ quantiles of the set $\{\beta_i^{*(b)} : U_{n,t}(\beta^{*(b)}) = u^{(b)}, b = 1, \dots, B\}$, respectively. If the normal approximation of $U_{n,t}(\beta)$ is to be used, we can implement this by repeatedly generating the p components of $u^{(b)}$ as independent standard normal variates a large number B times. Otherwise, if the normal approximation is not used, we take $u^{(b)} = U_{n,t}^*(\hat{\beta})$ with $U_{n,t}^*(\hat{\beta})$ being the b th resampled value.

As for the probabilistic index model, using the constant V_c in (5), with $\beta_c = 0$, can potentially improve the finite-sample inference. Therefore, we denote as the bias-reduced bootstrapping U -statistics the method explained above but with $U_n^{(c)}(\beta)$ and $\hat{\beta}_{br}$ instead of $U_n(\beta)$ and $\hat{\beta}$.

3.2. Adjusted Jackknife Empirical Likelihood

The empirical likelihood function is constructed by restricting the sample space to the support of the sample observations, and imposing a multinomial distribution on these n sample points. The probability vector of the multinomial distribution is denoted by $\mathbf{w}^T = (w_1, w_2, \dots, w_n)$, with $\sum_{i=1}^n w_i = 1$. For a given \mathbf{w} , the likelihood is simply $L = \prod_{i=1}^n w_i$. If no further restrictions are imposed, the w_i 's that maximize L subject to the sum constraint are given by $w_i = n^{-1}$, for all $i = 1, \dots, n$. The resulting likelihood $L = \prod_{i=1}^n w_i = n^{-n}$ is known as the nonparametric likelihood. A statistical model typically imposes constraints on the distribution of the sample observations which we express here by using an estimating equation. This leads to the multinomial likelihood $\prod_{i=1}^n w_i$ maximized over \mathbf{w} subject to the sum and model constraints; this is the empirical likelihood function. The appeal of the empirical likelihood method is that it delivers an empirical likelihood ratio statistic, for which often Wilks' theorem holds so that confidence intervals can be computed from quantiles of a chi-squared distribution. This method avoids the estimation of the covariance matrix of the parameter estimators, and the resulting confidence intervals are not necessarily symmetric about the parameter estimates (in contrast to Wald confidence intervals). We now explicitly work out the empirical likelihood for the probabilistic index model.

Upon writing the estimating function as in (5), the model constraint can be written as $E\{\mathbf{h}(\mathbf{W}, \mathbf{W}^*; \beta)\} = \mathbf{0}$, for some β . This expectation can be defined with respect to the empirical likelihood, which is a multinomial distribution with probability vector \mathbf{w} , and thus

$$E\{\mathbf{h}(\mathbf{W}, \mathbf{W}^*; \beta)\} = \sum_{i,j=1}^n w_i w_j \mathbf{h}(\mathbf{W}_i, \mathbf{W}_j; \beta) = \mathbf{0}. \quad (6)$$

Hence, the empirical likelihood function for the probabilistic index model becomes

$$L(\beta) = \max_{\mathbf{w}} \left\{ \prod_{i=1}^n w_i : \sum_{i=1}^n w_i = 1, \sum_{i,j=1}^n w_i w_j \mathbf{h}(\mathbf{W}_i, \mathbf{W}_j; \beta) = \mathbf{0} \right\}. \quad (7)$$

The constrained maximization involved in (7) is non-linear and computationally intensive. Wood et al. (1996) suggested an approximation by sequentially linearizing the constraints. The method is ad-hoc and is computationally intensive because it typically requires several iterations. A computationally more efficient method was proposed by Jing et al. (2009). It essentially consists of rewriting $U_n(\beta)$ as a sum of asymptotically independent terms, say $\hat{\mathbf{H}}_i(\beta)$, for $i = 1, \dots, n$, and replacing the model constraint in (7) by $\sum_{i=1}^n w_i \hat{\mathbf{H}}_i(\beta) = \mathbf{0}$, which is linear in the w_i s, therefore making the maximization problem easier. Jing et al. (2009) constructed the $\hat{\mathbf{H}}_i(\beta)$ as the jackknife pseudo-values, defined as

$$\hat{\mathbf{H}}_i(\beta) = n\mathbf{U}_n(\beta) - (n-1)\mathbf{U}_{n-1}^{(-i)}(\beta), \quad (8)$$

where $\mathbf{U}_{n-1}^{(-i)}(\beta)$ is the U -statistic of Equation (5) but with observation \mathbf{W}_i removed from the sample, so that $\bar{\mathbf{H}}_n(\beta) = n^{-1} \sum_{i=1}^n \hat{\mathbf{H}}_i(\beta) = \mathbf{U}_n(\beta)$. The method that results from this approach is known as the jackknife empirical likelihood method. Jing et al. (2009) and Lin (2013) showed that Wilks' theorem is still valid for the jackknife empirical likelihood statistic.

As noticed by a referee, an alternative and possibly faster procedure exists in applying the modified Aitchison/Silvey Lagrangian method of Bergsma et al. (2009), which is an adaptation of methods by Aitchison and Silvey (1958); Lang and Agresti (1994) and Lang (1996) for maximizing a multinomial likelihood subject to constraints (linear and non-linear). We may evaluate this method in future work.

A necessary condition for the maximum of the empirical likelihood or the jackknife empirical likelihood to exist is that $\mathbf{0}$ is an interior point of the convex hull of the n points $\hat{\mathbf{H}}_i(\beta)$. Because a maximum is often difficult to find, Chen et al. (2008) proposed to add one extra pseudo-observation, say $\hat{\mathbf{H}}_{n+1}(\beta)$, so that $\mathbf{0}$ is always included in the resulting convex hull of $n+1$ points. The extra point is given by $\hat{\mathbf{H}}_{n+1}(\beta) = -a_n \bar{\mathbf{H}}_n(\beta)$, for some $a_n > 0$. The resulting empirical likelihood is known as the adjusted empirical likelihood. Chen et al. (2008) showed that if $a_n = o_p(n^{2/3})$, it retains the first order asymptotic properties of the classical empirical likelihood. They suggested taking $a_n = \max\{1, \log(n)/2\}$, which worked well in a number of examples presented in their paper. Liu and Chen (2010) later recommended using $a_n = b/2$, where $b = \tau_4/2 - \tau_3^2/3$, with $\tau_r = E\{\mathbf{h}(\mathbf{X}_i, \mathbf{X}_j; \beta)^r\}$, is the Bartlett correction factor for the original empirical likelihood, so that for a scalar estimating function the resulting adjusted empirical likelihood achieves the higher-order precision of the Bartlett corrected empirical likelihood. If the estimating function is of dimension $q > 1$, the authors propose adding an extra pseudo-observation $a_{n+1} \bar{\mathbf{H}}_n(\beta)$, with $a_{n+1} - a_n = b$.

The adjusted empirical likelihood was originally constructed from an empirical likelihood with a linear constraint. The same procedure, however, is also valid when applied to the jackknife empirical likelihood. This has been demonstrated by Zhao et al. (2015) and Wang et al. (2015) for building confidence regions for the mean absolute deviation of random variables and for the Gini index, respectively. In the probabilistic index model setting, the adjusted jackknife empirical likelihood can be written as

$$L(\beta) = \max_{\mathbf{w}} \left\{ \prod_{i=1}^{n+1} w_i : \sum_{i=1}^{n+1} w_i = 1, \sum_{i=1}^{n+1} w_i \hat{\mathbf{H}}_i(\beta) = \mathbf{0} \right\}. \quad (9)$$

The maximization in (9) has an analytical solution which is obtained by introducing Lagrange multipliers $\boldsymbol{\lambda}^T = (\lambda_1, \dots, \lambda_p)$ for dealing with the linear constraints. The maximum is reached at $w_i = \{1 + \boldsymbol{\lambda}^T \hat{\mathbf{H}}_i(\beta)\}^{-1}$, where the Lagrange multiplier is the solution to

$$\frac{1}{(n+1)} \sum_{i=1}^{n+1} \frac{\hat{\mathbf{H}}_i(\beta)}{1 + \boldsymbol{\lambda}^T \hat{\mathbf{H}}_i(\beta)} = \mathbf{0}. \quad (10)$$

After plugging the weights w_i back into equation (9), the adjusted jackknife empirical likelihood ratio statistic becomes

$$R(\beta) = \frac{L(\beta)}{(n+1)^{-(n+1)}} = \prod_{i=1}^{n+1} \{1 + \lambda^T \hat{\mathbf{H}}_i(\beta)\}^{-1}. \quad (11)$$

Zhao et al. (2015) and Wang et al. (2015) further showed that Wilks' theorem is still valid for the adjusted jackknife empirical likelihood ratio statistic and their simulation results suggest that the confidence intervals obtained by inverting the test lead to improvements in coverage when compared to the standard jackknife empirical likelihood method. However, theory is available only for the $p = 1$ setting. Our Theorem 1 below, whose proof is given in the Supplementary Material (section S.1), generalizes the results to $p \geq 1$ dimensional parameters.

Theorem 1. *Let β_0 be the true value of β and assume that \mathbf{h} is a square-integrable kernel function, $\text{var}[E\{\mathbf{h}(\mathbf{W}_1, \mathbf{W}_2; \beta) | \mathbf{W}_1\}]$ is finite and nonsingular, and $a_n = o_p(n^{2/3})$. Then, with $R(\beta)$ as defined in Equation (11),*

$$-2 \log \{R(\beta_0)\} \rightarrow \chi_p^2$$

in distribution, as $n \rightarrow \infty$.

All empirical likelihood methods discussed so far, come with their own Wilks' theorem that provides the chi-square distribution as an asymptotic approximation. Owen (1990) argued that the chi-square distribution may be replaced by a scaled F -distribution to improve the small sample coverage of confidence intervals. In particular, for the setting of Theorem 1, χ_p^2 can be replaced by $(n-1)p/(n-p)F_{p,n-p}$ (Owen, 1990). The simulation study of Section 4 includes a comparison of empirical likelihood methods relying on both approximations.

The result of Theorem 1 relies on the convergence of $\bar{\mathbf{H}}_n^T(\beta_0) \hat{\Sigma}_H^{-1}(\beta_0) \bar{\mathbf{H}}_n(\beta_0)$ to a chi-square distribution with p degrees of freedom, where $\hat{\Sigma}_H(\beta) = n^{-1} \sum_{i=1}^n \hat{\mathbf{H}}_i(\beta) \hat{\mathbf{H}}_i^T(\beta)$ is a consistent estimator of $\text{var}\{U_n(\beta)\}$ (Lee, 1990). This approximation can be further improved by minimizing the finite-sample bias of $\hat{\Sigma}_H(\beta)$, as suggested in the proposition below whose proof is also given in the Supplementary Material (Section S.2).

Proposition 2. *Let $\hat{\Sigma}_H^{(c)}(\beta) = n^{-1} \sum_{i=1}^n \hat{\mathbf{H}}_i^{(c)}(\beta) [\hat{\mathbf{H}}_i^{(c)}(\beta)]^T$, where $\hat{\mathbf{H}}_i^{(c)}(\beta)$, for $i = 1, \dots, n$, are the jackknife pseudo-values based on $U_n^{(c)}(\beta)$ with $\beta_c = \mathbf{0}$. If the regularity conditions of Theorem 1 are satisfied, then $\text{abs}[\text{bias}\{\hat{\Sigma}_H^{(c)}(\beta_0)\}] \leq \text{abs}[\text{bias}\{\hat{\Sigma}_H(\beta_0)\}]$, where $\text{abs}\{\mathbf{H}(\cdot)\}$ stands for the absolute value of the components of $\mathbf{H}(\cdot)$ and the inequality holds component-wise.*

Since the bias-reduced estimator $\hat{\beta}_{\text{br}}$ is consistent, Theorem 1 also holds with this estimator. For completeness, this result is stated in the following theorem (the proof follows the same steps as for Theorem 1 and is thus omitted).

Theorem 2. *Let $\hat{\mathbf{H}}_i^{(c)}(\beta)$ be defined as in Proposition 2, let β_0 be the true value of β and assume that the regularity conditions of Theorem 1 are satisfied. Then, with $R^{(c)}(\beta) = \prod_{i=1}^{n+1} \{1 + \lambda^T \hat{\mathbf{H}}_i^{(c)}(\beta)\}^{-1}$,*

$$-2 \log \{R^{(c)}(\beta_0)\} \rightarrow \chi_p^2$$

in distribution, as $n \rightarrow \infty$.

Because of the reduced bias (Proposition 2), we expect that the empirical likelihood ratios may lead to a better chi-square approximation in finite-sample when compared to the original jackknife and adjusted jackknife empirical likelihood methods. Their finite sample properties will be empirically investigated in Section 4.

3.3. Profile Adjusted Jackknife Empirical Likelihood

In this section we introduce a more flexible adjusted jackknife empirical likelihood that allows us to make inference on a subset of β only. To this end, let β be partitioned as $\beta^T = (\xi^T, \gamma^T)$, where ξ is a q_1 -dimensional vector and γ a $q_2 (= p - q_1)$ -dimensional vector. We are only interested in inference on ξ , so that γ is of no interest in its own and is treated as a nuisance parameter.

Qin and Lawless (1994) first showed that Wilks' theorem is valid when using the profile empirical likelihood $L(\xi) = L\{\xi, \hat{\gamma}(\xi)\}$, where $\hat{\gamma}(\xi)$ maximizes $L(\beta) = L((\xi^t, \gamma^t)^t)$ with ξ held fixed. An alternative approach that

reduces the computational burden of estimating $\hat{\gamma}(\xi)$ for each ξ was recently introduced by Li et al. (2011). The authors proposed a two-step jackknife empirical likelihood procedure in which the nuisance parameter γ is first estimated from q_2 -dimensional estimating equations and the jackknife empirical likelihood is subsequently applied to the remaining q_1 estimating equations, with $\hat{\gamma}(\xi)$ replaced by the estimate of γ , say $\tilde{\gamma}(\xi)$. Wilks' theorem was shown to hold for this setting. Here we extend it to the adjusted jackknife empirical likelihood method.

First, consider the partition $U_n^T(\beta) = \{U_{n,1}^T(\beta), U_{n,2}^T(\beta)\}$ in which the two components correspond to the q_1 and q_2 -dimensional parts of $U_n(\beta)$, relating to ξ and γ , respectively. Let the jackknife pseudo-values be defined as $\hat{H}_{i,1}(\xi) = nU_{n,1}\{\xi, \tilde{\gamma}(\xi)\} - (n-1)U_{n-1,1}^{(-i)}\{\xi, \tilde{\gamma}^{(-i)}(\xi)\}$, $i = 1, \dots, n$, where $\tilde{\gamma}(\xi)$ and $\tilde{\gamma}^{(-i)}(\xi)$ are the solutions of $U_{n,2}(\xi, \gamma) = \mathbf{0}$ and $U_{n-1,2}^{(-i)}(\xi, \gamma) = \mathbf{0}$, respectively, with ξ held fixed, and $\hat{H}_{n+1,1}(\xi) = -a_n \bar{H}_{n,1}(\xi) = -a_n n^{-1} \sum_{i=1}^n \hat{H}_{i,1}(\xi)$, for some $a_n > 0$. Note that as the probabilistic index model estimator $\hat{\beta}$ satisfies $U_n(\hat{\beta}) = \mathbf{0}$ and since $n^{-1} \sum_{i=1}^n \hat{H}_{i,1}(\xi) = U_{n,1}\{\xi, \tilde{\gamma}(\xi)\}$, we have that $U_{n,1}\{\hat{\xi}, \tilde{\gamma}(\hat{\xi})\} = \mathbf{0}$. This leads to the profile adjusted jackknife empirical likelihood

$$L(\xi) = \max_w \left\{ \prod_{i=1}^{n+1} w_i : \sum_{i=1}^{n+1} w_i = 1, \sum_{i=1}^{n+1} w_i \hat{H}_{i,1}(\xi) = \mathbf{0} \right\}.$$

The following theorem confirms the validity of the Wilks' theorem.

Theorem 3. *Let ξ_0 be the true value of ξ and assume that the regularity conditions of Theorem 1 hold. Then,*

$$-2 \log \{R(\xi_0)\} = -2 \log \left\{ L(\xi_0) / (n+1)^{-(n+1)} \right\} \rightarrow \chi_{q_1}^2,$$

in distribution, as $n \rightarrow \infty$.

This result can be proved by first observing that $\max_{1 \leq i \leq n} \|\hat{H}_{i,1}(\xi)\| = o_P(n^{1/2})$ (Li et al., 2011). The remainder of the proof is similar to that of Theorem 1 and is thus omitted.

4. Simulation Study

4.1. Description

In this section we empirically evaluate the methods of Section 3 for small sample inference. We generate data from the linear model

$$Y = \mathbf{X}^T \alpha + \epsilon. \quad (12)$$

Our goal is to make inference about β , the parameter of the probabilistic index model (1). Following Thas et al. (2012) the parameters α and β are related. Specifically, if ϵ follows a standard normal distribution, one can show that $\beta = \alpha/\sqrt{2}$, $\mathbf{Z}_{ij} = \mathbf{X}_j - \mathbf{X}_i$ and $\mathbf{g}^{-1}(\cdot)$ is the probit link function. If ϵ follows a Gumbel distribution, $\beta = \alpha$, $\mathbf{Z}_{ij} = \mathbf{X}_j - \mathbf{X}_i$ and $\mathbf{g}^{-1}(\cdot)$ is the logit link function. This simple relationship allows us to simulate from (12) and assess the performance of inference on the parameter β . In this section we present results when the error term follows a standard normal distribution. Results for a Gumbel distributed error are available in section S.4 of the Supplementary Material.

The focus of this simulation study is on the coverage of confidence intervals for β . Recall that for both $\hat{\beta}$ and $\hat{\beta}_{br}$, which are the solutions to $U_n(\beta) = \mathbf{0}$ and $U_n^{(c)}(\beta) = \mathbf{0}$, respectively, standard Wald intervals are available. These intervals are compared with respect to length and coverage, to those obtained from the following methods: the bootstrap and bias-reduced bootstrap method of Jiang and Kalbfleisch (2012), the jackknife and bias-reduced jackknife empirical likelihood and the adjusted and bias-reduced adjusted jackknife empirical likelihood. For the latter, we considered the empirical $a_n = \max\{1, \log(n)/2\}$, as well as $a_n = b/2$, which is related to the Bartlett correction for the original empirical likelihood, as discussed in Section 3.2. The adjusted jackknife empirical likelihood methods are implemented both with the χ^2 and the scaled F distributions for calibration. For the bootstrap methods, we assign the resampling weights as independent Gamma(1,1), as in Jiang and Kalbfleisch (2012), and use quantiles from the bootstrap replications to generate the bootstrap confidence interval.

Table 1: Empirical coverages (in %) for scalar β , based on 2,500 Monte Carlo simulations for Model (12) with $\alpha = 0$ and $\alpha = 2\sqrt{2}$ and a normal error distribution. Results are presented for nominal coverages of 95% and different sample sizes.

	sample size											
	20	25	35	50	75	100	20	25	35	50	75	100
	$\alpha = 0$						$\alpha = 2\sqrt{2}$					
PIM _w	89.8	91.9	92.6	93.4	93.6	94.8	87.4	88.6	90.8	92.4	93.2	94.5
BR-PIM _w	90.1	91.8	92.5	93.3	93.6	94.7	90.2	90.5	92.3	93.4	93.7	94.7
ATS ₁	89.8	91.9	92.6	93.4	93.6	94.8	87.4	88.6	90.8	92.4	93.2	94.5
BR-ATS ₁	90.1	91.8	92.5	93.3	93.6	94.7	90.2	90.5	92.3	93.4	93.7	94.7
ATS ₂	91.8	92.8	93.6	93.9	94.1	94.9	89.0	90.2	92.0	92.9	93.8	94.6
BR-ATS ₂	91.8	92.9	93.7	93.9	94.2	94.8	92.1	92.2	93.3	94.3	94.0	94.8
EF _t	96.4	96.4	95.7	95.3	94.8	95.4	90.8	91.5	92.6	93.4	93.7	94.8
BR-EF _t	95.9	96.5	95.6	95.2	95.0	95.2	92.6	93.0	93.7	94.5	94.6	95.0
JEL	93.2	94.3	94.6	94.9	94.8	95.3	90.1	91.6	92.7	93.8	94.2	95.1
BR-JEL	93.2	94.3	94.6	94.9	94.8	95.3	93.2	93.9	94.4	94.9	95.0	95.4
AJEL	95.2	96.0	95.7	95.8	95.4	95.8	91.4	92.7	93.8	94.5	94.8	95.4
BR-AJEL	95.2	96.0	95.7	95.8	95.4	95.8	94.7	95.2	95.5	96.1	95.8	95.6
AJEL _b	94.0	95.0	95.2	95.3	95.2	95.5	90.6	92.2	93.2	94.3	94.6	95.2
BR-AJEL _b	94.0	95.0	95.2	95.3	95.2	95.5	93.9	94.4	94.8	95.4	95.4	95.6
AJEL _b ^F	95.7	96.2	95.7	95.7	95.4	95.7	92.1	93.0	94.0	94.7	94.9	95.4
BR-AJEL _b ^F	95.7	96.2	95.7	95.7	95.4	95.7	94.9	95.5	95.6	96.2	95.9	95.6

NOTE: (BR-)PIM_w, standard Wald intervals based on $(\hat{\beta}_{br})\hat{\beta}$; (BR-)ATS₁, (bias-reduced) anova-type-statistic of type 1; (BR-)ATS₂ anova-type-statistic of type 2; (BR-)EF_t, (bias-reduced) bootstrapping U -statistics; (BR-)JEL, (bias-reduced) jackknife empirical likelihood; (BR-)AJEL, (bias-reduced) adjusted jackknife empirical likelihood with $a_n = \max\{0, \log(n)/2\}$; (BR-)AJEL_b, (bias-reduced) adjusted jackknife empirical likelihood with $a_n = b/2$; (BR-)AJEL_b^F, (bias-reduced) AJEL_b with a scaled F -distribution.

Two more methods are included in the simulation study. Brunner et al. (1997) and Brunner et al. (2016) argued that small sample performance of rank methods that rely on Wald-type statistics may suffer from problems related to the inversion of a large covariance matrix. They proposed alternative statistics, which they refer to as ANOVA-type statistics, that avoid such matrix inversions, and instead rely on asymptotic approximations by scaled chi-square (ANOVA-type-statistic of type 1) or F distributions (ANOVA-type-statistic of type 2) with estimated degrees of freedom. Their ANOVA-type statistics are easy to construct for probabilistic index models and are based on the test statistic $\hat{\beta}^T T \hat{\beta}$, where T is the identity matrix with dimensions $p \times p$; details are provided in Supplementary Material (section S.3). These statistics are test statistics, and confidence intervals are obtained by inverting the tests.

4.2. Results for scalar α

The empirical evaluation is based on 2,500 Monte Carlo simulation runs, with 999 bootstrap samples. Several sample sizes n and two values for α (and hence β) are considered. We generate data from (12), with X taking equally spaced values between 0.1 and 1. The error term is simulated from a standard normal distribution. The empirical coverages of the confidence intervals are presented in Table 1 and their respective average lengths are displayed in Table 2. Figure 1 summarizes Tables 1 and Table 2 by comparing only 3 methods: the Wald-type, the ANOVA-type 2 statistics and the bias-reduced adjusted jackknife empirical likelihood with the scaled F -distribution approximation.

Wald-based confidence intervals show poor coverage in most cases and especially when the sample size n is equal to 25 or lower. The large finite-sample bias of $\hat{\beta}$ (Thas et al., 2012) partly explains the poor coverages when the parameter β (and thus α) deviates from zero. The bias-reduced estimator $\hat{\beta}_{br}$ reduces bias and the associated Wald confidence intervals show better results, but this improvement is limited. This may be to some extent explained by both confidence intervals relying on the asymptotic normality of $\hat{\beta}$ and $\hat{\beta}_{br}$. Figure S.1 in the Supplementary Material shows normal quantile-quantile plots of 2,500 estimates $\hat{\beta}$ from a simulation study. The coverages of the confidence intervals based on the ANOVA-type statistics are only slightly better than the Wald-based intervals and substantially worse than that obtained by the bias-reduced Bartlett-corrected jackknife empirical likelihood method with the scaled F -distribution approximation. This is clearly seen on Figure 1.

Table 2: Average lengths of confidence intervals for scalar β , based on 2,500 Monte Carlo simulations for Model (12) with $\alpha = 0$ and $\alpha = 2\sqrt{2}$ and a normal error distribution. The associated empirical coverages are displayed in Table 1.

	sample size						sample size					
	20	25	35	50	75	100	20	25	35	50	75	100
	$\alpha = 0$						$\alpha = 2\sqrt{2}$					
PIM _w	2.22	2.03	1.75	1.49	1.23	1.07	2.40	2.21	1.93	1.65	1.37	1.20
BR-PIM _w	2.22	2.03	1.75	1.49	1.23	1.07	2.51	2.27	1.96	1.66	1.37	1.20
ATS ₁	2.20	2.01	1.74	1.47	1.21	1.06	2.38	2.19	1.91	1.63	1.36	1.18
BR-ATS ₁	2.21	2.02	1.74	1.48	1.22	1.06	2.49	2.26	1.95	1.65	1.36	1.19
ATS ₂	2.36	2.13	1.81	1.51	1.24	1.07	2.55	2.32	1.99	1.68	1.39	1.20
BR-ATS ₂	2.36	2.13	1.81	1.52	1.24	1.08	2.66	2.38	2.02	1.69	1.39	1.21
EF _t	3.13	2.60	2.05	1.65	1.31	1.12	3.50	2.87	2.25	1.82	1.45	1.25
BR-EF _t	3.17	2.60	2.05	1.65	1.31	1.12	3.67	2.96	2.29	1.82	1.46	1.25
JEL	2.65	2.28	1.86	1.54	1.22	1.06	3.64	2.88	2.27	1.83	1.47	1.26
BR-JEL	2.62	2.27	1.86	1.54	1.22	1.06	3.89	2.91	2.27	1.82	1.46	1.26
AJEL	3.13	2.57	2.01	1.63	1.26	1.09	4.81	2.98	2.46	1.93	1.52	1.29
BR-AJEL	3.07	2.54	2.01	1.63	1.26	1.09	4.84	3.26	2.43	1.91	1.50	1.29
AJEL _b	2.84	2.39	1.93	1.57	1.23	1.07	4.94	3.33	2.38	1.89	1.50	1.28
BR-AJEL _b	2.77	2.37	1.92	1.57	1.23	1.07	4.15	3.02	2.33	1.86	1.48	1.27
AJEL _b ^F	3.18	2.58	2.01	1.62	1.26	1.09	5.11	3.44	2.49	1.94	1.52	1.30
BR-AJEL _b ^F	3.11	2.55	2.00	1.62	1.26	1.09	4.93	3.25	2.43	1.91	1.50	1.28

NOTE: (BR-)PIM_w, standard Wald intervals based on $(\hat{\beta}_{br}) \hat{\beta}$; (BR-)ATS₁, (bias-reduced) anova-type-statistic of type 1; (BR-)ATS₂ anova-type-statistic of type 2; (BR-)EF_t, (bias-reduced) bootstrapping U -statistics; (BR-)JEL, (bias-reduced) jackknife empirical likelihood; (BR-)AJEL, (bias-reduced) adjusted jackknife empirical likelihood with $a_n = \max\{0, \log(n)/2\}$; (BR-)AJEL_b, (bias-reduced) adjusted jackknife empirical likelihood with $a_n = b/2$; (BR-)AJEL_b^F, (bias-reduced) AJEL_b with a scaled F -distribution.

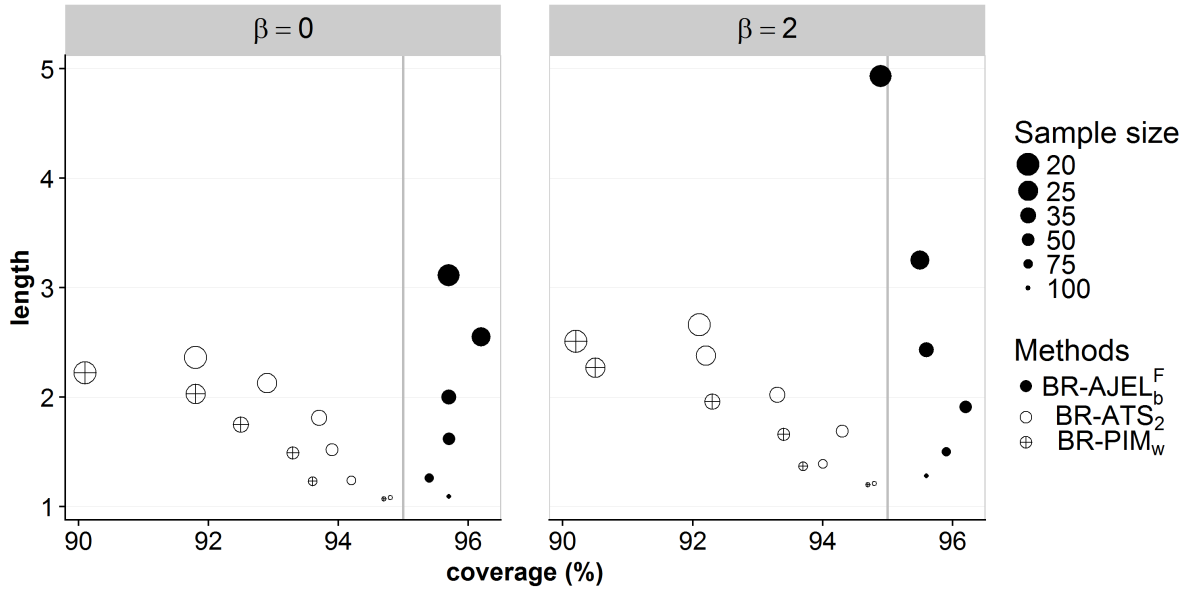


Figure 1: Empirical coverages of the confidence interval based on $\hat{\beta}_{br}$, the ANOVA-type-test ATS₂ and finally on the bias-reduced adjusted jackknife empirical likelihood with the scaled F -distribution approximation BR-AJEL_b^F. Results are based on 2,500 Monte Carlo runs and on Model (12) with normal error distribution and for different sample sizes. The left panel corresponds to $\alpha = 0$ and the right to $\alpha = 2\sqrt{2}$ in model (12).

Table 3: Empirical coverages (in %) for β_1 , based on 2,500 Monte Carlo simulations for Model (12) with $\alpha^T = (0, 0)$, $(\sqrt{2}, 0)$ and a normal error distribution. Results are presented for a nominal coverage (NC) of 95% and sample sizes of $n = 20, 40, 60$ and 80 .

	sample size							
	20	40	60	80	20	40	60	80
	$\alpha = (0, 0)$				$\alpha = (\sqrt{2}, 0)$			
JEL	92.4	94.4	94.4	94.9	92.0	92.1	92.8	93.5
BR-JEL	91.8	94.1	94.2	94.7	94.6	95.1	95.4	95.8
AJEL	94.7	95.4	95.0	95.5	93.0	92.9	93.6	94.4
BR-AJEL	94.1	95.1	95.0	95.4	96.0	96.0	95.9	96.1
AJEL _b	93.5	94.9	94.8	95.4	92.4	92.4	93.3	94.0
BR-AJEL _b	93.1	94.7	94.6	95.2	95.4	95.4	95.6	95.8
AJEL _b ^F	95.2	95.6	95.1	95.6	93.1	92.9	93.8	94.5
BR-AJEL _b ^F	94.4	95.3	95.0	95.4	96.2	96.1	95.8	96.1

NOTE: (BR-)JEL, (bias-reduced) jackknife empirical likelihood; (BR-)AJEL, (bias-reduced) adjusted jackknife empirical likelihood with $a_n = \max\{0, \log(n)/2\}$; (BR-)AJEL_b, (bias-reduced) adjusted jackknife empirical likelihood with $a_n = b/2$; (BR-)AJEL_b^F, (bias-reduced) AJEL_b with a scaled F -distribution.

The other methods based on bootstrapping and empirical likelihood do not rely on the symmetry of the asymptotic normal distribution of $\hat{\beta}$ or $\hat{\beta}_{br}$ and perform substantially better. They are, however, still affected by finite-sample bias. Their bias-reduced versions, i.e. those that use the variance V_c evaluated at $\beta_c = 0$, on the other hand, perform markedly better for the smaller sample sizes, resulting in coverages that are close to the nominal values. In particular, the bias-reduced Bartlett-corrected adjusted jackknife empirical likelihood method performs overall better than all other empirical likelihood methods. Both the adjusted Bartlett-corrected and the bias-reduced jackknife empirical likelihood methods perform well, but the latter is more computationally intensive than the former.

For three methods, Figure 1 summarizes the results on coverage and length of confidence intervals for several sample sizes. We conclude that lengths of intervals generally reduce with increasing sample size, and that the bias-reduced adjusted jackknife empirical likelihood intervals show overall better coverage (slightly conservative). Particularly for small sample sizes this comes at a price of larger lengths as compared to the other methods, which, however, are overoptimistic in terms of coverage.

4.3. Results for vector α

We now extend the simulation study to a bivariate vector β and compute the empirical coverages after 2,500 Monte Carlo runs. We generate data from (12) with $\mathbf{X}^T = (X_1, X_2)$ and X_1 and X_2 independent from each other, with X_1 taking equally spaced values between 0.1 and 1 and $X_{2i} \sim U(0, 0.8)$, for $i = 1, \dots, n/2$ and $X_{2i} \sim U(0.2, 1)$, for $i = n/2 + 1, \dots, n$. The error term is simulated from a standard normal distribution. The relationship between $\alpha^T = (\alpha_1, \alpha_2)$ and the $\beta^T = (\beta_1, \beta_2)$ parameters of the probabilistic index model (1) still holds. Table 3 shows results for the coverage of β_1 when β_2 is of no interest in its own and is treated as a nuisance parameter. Here we use the profile likelihood discussed in Section 3.3 for inference.

The results are similar to the univariate case. The bias-reduced adjusted jackknife and bias-reduced Bartlett-corrected adjusted jackknife empirical likelihood methods perform generally well. Hence both methods form an attractive and viable approach for making small sample inferences from estimates obtained via the probabilistic index models. However, since the Bartlett-corrected approach makes use of a better motivated choice for the a_n factor in the pseudo-observation, we recommend it for inference for probabilistic index models.

Next we extend our simulation scenario to a vector α of length $p \geq 2$. We take $\alpha = \mathbf{0}_p$ (and thus $\beta = \mathbf{0}_p$), where $\mathbf{0}_p$ is a vector of zeros of length p , and we are now interested in constructing confidence regions for β . To this end, we generate data from the linear model (12), with \mathbf{X} following a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix equal to the $[p \times p]$ identity matrix. The error term follows a standard normal distribution; results for a Gumbel distributed error are available in the Supplementary Material (Table S.4). The empirical coverage based on 2,500 Monte Carlo runs for Wald-based, ANOVA-type and all empirical likelihood methods are displayed in Table 4. Figure 2 summarizes the information from Table 4 by comparing the best Wald-based, ANOVA-type and empirical likelihood based method.

Table 4: Empirical coverages (in %) for β , based on 2,500 Monte Carlo simulations for Model (12) with $\alpha = \mathbf{0}_p$ and a normal error distribution. Results are presented for nominal coverages of 95% and different sample sizes.

p	2	3	4	5	2	3	4	5	2	3	4	5
	$n = 25$				$n = 50$				$n = 100$			
PIM _w	85.5	79.2	70.6	61.2	90.8	89.2	86.3	81.8	92.8	91.7	91.1	90.0
BR-PIM _w	86.2	80.8	73.6	64.6	91.2	89.3	87.0	83.4	93.0	91.8	91.6	90.4
ATS ₁	88.4	86.2	84.3	81.5	91.4	91.3	89.8	88.6	93.5	92.9	92.6	91.4
BR-ATS ₁	89.8	89.0	89.3	89.3	91.9	91.9	91.7	91.0	93.4	93.2	93.0	92.5
ATS ₂	89.4	87.4	85.5	82.7	91.7	91.8	90.5	89.0	93.6	93.2	92.7	91.7
BR-ATS ₂	90.8	90.6	90.4	90.0	92.4	92.3	92.3	91.3	93.8	93.6	93.1	92.7
JEL	90.6	87.0	81.8	73.0	93.2	91.6	89.8	86.8	94.1	94.0	92.9	92.4
BR-JEL	90.6	87.0	81.8	73.0	93.2	91.6	89.8	86.8	94.1	94.0	92.9	92.4
AJEL	92.6	89.6	86.0	79.6	93.8	92.9	91.5	88.6	94.5	94.7	93.4	93.1
BR-AJEL	92.6	89.6	86.0	79.6	93.8	92.9	91.5	88.6	94.5	94.7	93.4	93.1
AJEL _b	91.7	88.4	84.1	78.1	93.6	92.3	90.8	88.1	94.4	94.3	93.1	92.8
BR-AJEL _b	91.7	88.4	84.1	78.1	93.6	92.3	90.8	88.1	94.4	94.3	93.1	92.8
AJEL _b ^F	94.7	93.2	92.1	91.3	95.0	95.0	93.6	92.8	95.0	95.2	94.4	94.3
BR-AJEL _b ^F	94.7	93.2	92.1	91.3	95.0	95.0	93.6	92.8	95.0	95.2	94.4	94.3

NOTE: (BR-)PIM_w, standard Wald intervals based on $(\hat{\beta}_{br}) \hat{\beta}$; (BR-)ATS₁, (bias-reduced) anova-type-statistic of type 1; (BR-)ATS₂ anova-type-statistic of type 2; (BR-)EF_t, (bias-reduced) bootstrapping U -statistics; (BR-)JEL, (bias-reduced) jackknife empirical likelihood; (BR-)AJEL, (bias-reduced) adjusted jackknife empirical likelihood with $a_n = \max\{0, \log(n)/2\}$; (BR-)AJEL_b, (bias-reduced) adjusted jackknife empirical likelihood with $a_n = b/2$; (BR-)AJEL_b^F, (bias-reduced) AJEL_b with a scaled F -distribution.

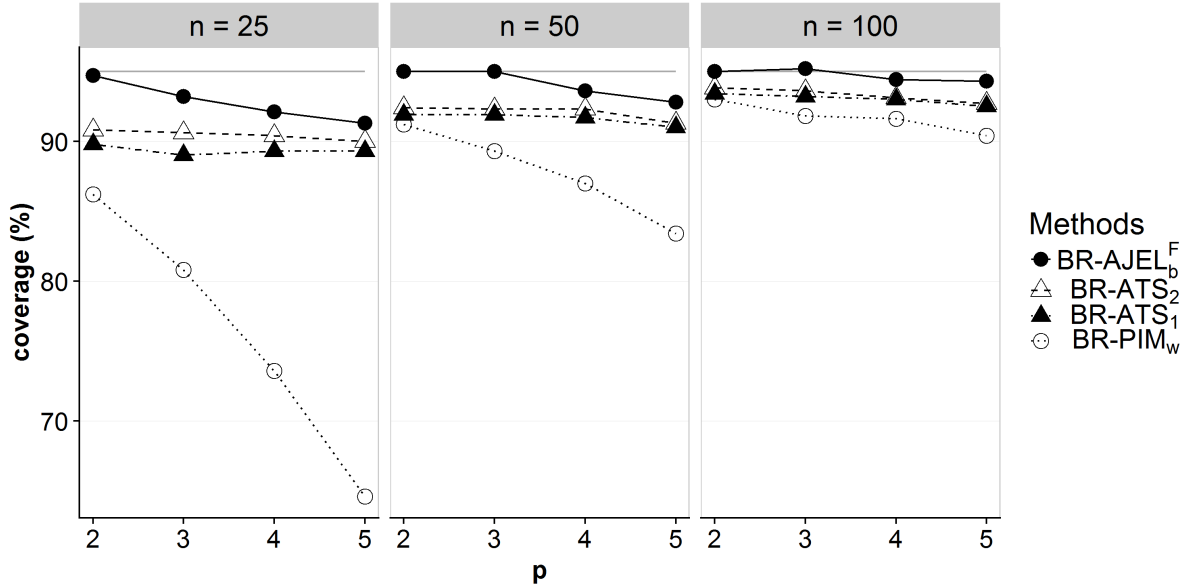


Figure 2: Empirical coverages of the Wald-based confidence interval for $\hat{\beta}_{br}$, the two ANOVA-type-statistics ATS₁ and ATS₂ based on the bias-reduced PIM estimate $\hat{\beta}_{br}$, and bias-reduced adjusted jackknife empirical likelihood with the scaled F -distribution approximation BR-AJEL_b^F. Results are average values of 2,500 Monte Carlo runs, for data generated from Model (12) with $\alpha_p = \mathbf{0}_p$, normally distributed error and for samples of sizes $n = 25, 50$ and 100 (left to right).

Wald-based confidence intervals show very poor performance, with empirical coverage close to 60% when $\beta = \mathbf{0}_5$ and $n = 25$. Its bias-reduced version leads to better empirical coverage, but it is still far below the nominal value. Both ANOVA-type statistics show large improvements over Wald-based confidence regions and their bias-reduced versions,

which use the bias-reduced estimate $\hat{\beta}_{br}$ in the computation of the test statistic, show substantial improvements. For example, when $\alpha = \mathbf{0}_5$ and $n = 25$, the unadjusted ANOVA-type statistic of type 2 has an empirical coverage of 82.7%, while its bias-reduced version lead to 90.0%. Their empirical coverages, however, are still below those obtained from the Bartlett-corrected adjusted jackknife empirical likelihood with the F -distribution approximation. This is clearly seen in Figure 2. Finally, as the ANOVA-type statistic is based on $\hat{\beta}^T T \hat{\beta}$, where T , in this case, is the identity matrix with dimensions $p \times p$, it does not take into account the correlation between the estimated parameters and will lead, in the special case when $p = 2$, to circular confidence regions. Empirical likelihood based methods, on the other hand, lead to confidence regions that resemble an ellipse, as the correlation between the outcomes is now taken into account. This, together with its better performance in our simulation studies, suggests that the bias-reduced Bartlett-corrected adjusted jackknife empirical likelihood with the F -distribution approximation is to be recommended for small sample inference in the probabilistic index model framework.

5. Illustration

The PASSAGE (Training Program of Efficient Self-Management Strategies, in French) program is a group intervention study for the self-management of Fibromyalgia syndrome (FMS), which is a chronic disorder of uncertain origin and causes widespread musculoskeletal pain in association with fatigue, poor sleep quality, cognitive dysfunction, mood disturbances, and many other variable somatic symptoms (Bourgault et al., 2015). The intervention consists of providing psycho-educational tools and physical exercises with the purpose of reducing FMS symptoms and improving quality of life (QOL).

The primary outcome was reduction in pain intensity (numerical rating scale, where 0 indicates “no pain” and 10 indicates “worst possible pain”) and secondary outcomes included whether the intervention improved pain (binary outcome), quality of life (QOL) (“remained unchanged”, “considerably deteriorated” or “considerably improved”) or pain relief (dichotomized variable, based on 0 – 100% perception of pain relief, with cut-off value at 50%), among others. The study was performed in two centers located in Quebec, Canada, and the intervention was randomized within study site. A total of 43 individuals (23 in the control group) with complete data measured at the beginning and three months after the intervention had started were later used in the analysis.

As the nature of the outcomes is mostly ordinal, the probabilistic index model is a natural choice to make inference regarding the efficacy of the intervention conditioned on study site. We analyzed the impact of the intervention on improving pain, QOL and pain relief, while adjusting for study site with a probabilistic index model with logit link. We also analyzed the primary outcome with a probit link, with average pain intensity at baseline, treatment and study site as regressors. Confidence intervals, after 999 bootstrap replicates, for the estimated probabilistic index for two randomly selected persons from the same study site, but from different treatment groups, are shown in Figure 3. Table S.5 in Supplementary Material shows the results numerically. In all analyses, study site was not significant at the 5% level, but it was kept in all models.

Our results agree with the original paper in their main conclusions. In particular, patients reported a significant improvement in their pain levels and in their quality of life, while their pain relief and their average pain were not significantly improved by the PASSAGE program, at a 5% significance level. For instance, the bias-reduced Bartlett-corrected adjusted jackknife empirical likelihood method with the scaled F -distribution, which often performed better than its competitors in the simulation studies, shows that for two randomly selected persons from the same study site, the person that attended the PASSAGE program reported improvements in their pain levels with probability of 81.0% and with a 95% confidence interval of [66.4%; 91.7%].

Acknowledgment

The authors would like to thank the Flemish Research Council (FWO) for financial support (Grant G.0202.14N).

Aitchison, J. and S. Silvey (1958). Maximum-likelihood estimation of parameters subject to restraints. *The Annals of Mathematical Statistics*, 813–828.

Basu, S. (2001). Improved small sample inference procedures for epidemiological parameters under cross-sectional sampling. *Journal of the Royal Statistical Society: Series D (The Statistician)* 50(3), 309–319.

Bergsma, W., M. A. Croon, and J. A. Hagenaars (2009). *Marginal models for Dependent, Clustered, and Longitudinal Categorical Data*. New York: Springer.

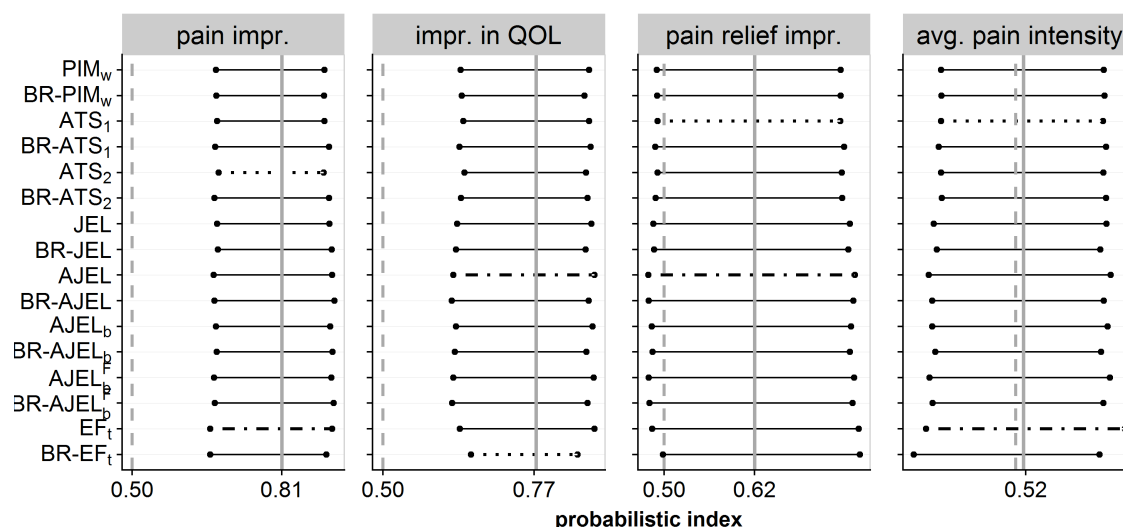


Figure 3: 95% confidence intervals for the treatment effect, on the scale of the probabilistic index, on pain improvement, on improvement in quality of life (QOL), on pain relief improvement and on average pain intensity. The horizontal dotted and dot-dashed lines denote the narrowest and widest confidence intervals, respectively. The vertical long-dashed and solid lines correspond to a probabilistic index of 0.5 (no effect) and to the estimate probabilistic index, respectively. All methods are abbreviated as in the footnote below Table 1.

- Bergsma, W., M. A. Croon, J. A. Hagenaars, and A. van der Ark (2012). Discussion of “Probabilistic Index Models” by O.Thas, J. De Neve, L. Clement and J.P. Ottoy. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74, 655–656.
- Bourgault, P., A. Lacasse, S. Marchand, R. Courtemanche-Harel, J. Charest, I. Gaumond, J. B. de Souza, and M. Choinière (2015). Multicomponent interdisciplinary group intervention for self-management of fibromyalgia: A mixed-methods randomized controlled trial. *PLoS one* 10(5), e0126324.
- Brunner, E., H. Dette, and A. Munk (1997). Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association* 92, 1494–1502.
- Brunner, E., F. Konietzschke, M. Pauly, and M. L. Puri (2016). Rank-based procedures in factorial designs: hypotheses about nonparametric treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.
- Chen, J., A. M. Variyath, and B. Abraham (2008). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics* 17(2), 426–443.
- Cordeiro, G. M. and P. McCullagh (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 629–643.
- De Neve, J. (2013). *Probabilistic index models*. Ph. D. thesis, Ghent University, Faculty of Bioscience Engineering, Ghent, Belgium.
- De Neve, J. and O. Thas (2015). A regression framework for rank tests based on the probabilistic index model. *Journal of the American Statistical Association* 110(511), 1276–1283.
- Hu, F. and J. D. Kalbfleisch (2000). The estimating function bootstrap. *Canadian Journal of Statistics* 28(3), 449–481.
- Jiang, W. and J. D. Kalbfleisch (2012). Bootstrapping u-statistics: applications in least squares and robust regression. *Sankhya B* 74(1), 56–76.
- Jing, B.-Y., J. Yuan, and W. Zhou (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association* 104(487), 1224–1232.
- Lang, J. B. (1996). On the comparison of multinomial and poisson log-linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 253–266.
- Lang, J. B. and A. Agresti (1994). Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association* 89(426), 625–632.
- Lee, A. J. (1990). *U-statistics: Theory and Practice*. Marcel Dekker Inc., New York.
- Li, M., L. Peng, and Y. Qi (2011). Reduce computation in profile empirical likelihood method. *Canadian Journal of Statistics* 39(2), 370–384.
- Lin, Q. (2013). *A jackknife empirical likelihood approach to goodness of fit U-statistic testing with side information*. Ph. D. thesis, Purdue University.
- Liu, Y. and J. Chen (2010). Adjusted empirical likelihood with high-order precision. *The Annals of Statistics* 38(3), 1341–1362.
- Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75(2), 237–249.
- Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics* 18(1), 90–120.
- Parzen, M., L. Wei, and Z. Ying (1994). A resampling method based on pivotal estimating functions. *Biometrika* 81(2), 341–350.
- Paul, S. and X. Zhang (2014). Small sample gee estimation of regression parameters for longitudinal data. *Statistics in Medicine* 33(22), 3869–3881.
- Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics* 1(22), 300–325.
- Thas, O., J. De Neve, L. Clement, and J.-P. Ottoy (2012). Probabilistic index models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74(4), 623–671.

- Vermeulen, K., O. Thas, and S. Vansteelandt (2015). Increasing the power of the mann-whitney test in randomized experiments through flexible covariate adjustment. *Statistics in Medicine* 34(6), 1012–1030.
- Wang, D., Y. Zhao, and D. W. Gilmore (2015). Jackknife empirical likelihood confidence interval for the gini index. *Statistics & Probability Letters* 110, 289–295.
- Wood, A. T., K.-A. Do, and B. Broom (1996). Sequential linearization of empirical likelihood constraints with application to u-statistics. *Journal of Computational and Graphical Statistics* 5(4), 365–385.
- Zhao, Y., X. Meng, and H. Yang (2015). Jackknife empirical likelihood inference for the mean absolute deviation. *Computational Statistics & Data Analysis* 91, 92–101.
- Zhou, W. (2012). Discussion of “Probabilistic Index Models” by O.Thas, J. De Neve, L. Clement and J.P. Ottoy. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74, 666.