Extending the two-sample empirical likelihood method

Jānis Valeinis *
University of Latvia

Edmunds Cers
University of Latvia

August 28, 2011

Abstract

In this paper we establish the empirical likelihood method for the two-sample case in a general framework. We show that the result of Qin and Zhao (2000) basically covers the following two-sample models: the differences of two sample means, smooth Huber estimators, distribution and quantile functions, ROC curves, probability-probability (P-P) and quantile-quantile (Q-Q) plots. Finally, the structural relationship models containing the location, the location-scale and the Lehmann alternative models also fit in this framework. The method is illustrated with real data analysis and simulation study. The R code has been developed for all two-sample problems and is available on the corresponding author's homepage.

Keywords: Empirical likelihood, two-sample, two-sample mean difference, two-sample smooth Huber estimator difference, P-P plot, Q-Q plot, ROC curve.

*Corresponding author address: Department of Mathematics, Faculty of Physics and Mathematics, University of Latvia, Zellu 8, LV-1002, Riga, Latvia. E-mail: valeinis@lu.lv

1 Introduction

Since Owen (1988, 1990) introduced the empirical likelihood (EL) method for statistical inference, there have been several attempts to generalize it for the two-sample case. Qin and Zhao (2000) established the empirical likelihood method for differences of two-sample means and cumulative distribution functions at some fixed point. In the PhD thesis of Valeinis (2007) it has been showed that this result can be applied also for other two-sample problems (see also Valeinis, Cers and Cielens, 2010). Moreover, this setup basically generalizes the results of Claeskens, Jing, Peng and Zhou (2003) and Zhou and Jing (2003), where ROC curves and P-P plots in the two-sample case and quantile differences in the one-sample case have been considered, respectively. Although Claeskens, Jing, Peng and Zhou (2003) state that Q-Q plots would need a different theoretical treatment, in our setup we show how to treat them in the same way (see Examples 6 and 7).

Consider the two-sample problem, where i.i.d. random variables $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ are independent and have some unknown distribution functions F_1 and F_2, respectively. Without a loss of generality we assume that $n_2 \geq n_1$. We are interested to make inference for some function $t \to \Delta(t)$ defined on an interval $T \subset \mathbb{R}$. Let $\theta = \theta(t)$ be a function associated with one of the distribution functions F_1 or F_2. For fixed t both $\Delta(t)$ and $\theta(t)$ are univariate parameters in our setup (for simplicity further on we write Δ and θ). We assume that all the information about the unknown true parameters θ_0 and Δ_0 are available in the known form of unbiased estimating functions, i.e.,

$$E_{F_1} w_1(X, \theta_0, \Delta_0, t) = 0, \tag{1}$$

$$E_{F_2} w_2(Y, \theta_0, \Delta_0, t) = 0. \tag{2}$$

If $\Delta_0 = \theta_1 - \theta_0$, where θ_0 and θ_1 are univariate parameters associated with F_1 and F_2, respectively, we have exactly the setup of Qin and Zhao (2000). Now let us list the main two-sample models which fit in our framework.

Example 1 (Qin and Zhao, 2000)**.** The difference of two sample means. This topic using EL has been analyzed by Jing (1995), Qin and Zhao (2000) and recently by Liu, Zou and Zhang (2008), where the Bartlett correction has been established for the difference of two sample means. In Valeinis, Cers and Cielens (2010) the two-sample EL method has been compared by empirical coverage accuracy with t-test and some bootstrap methods. Denote $\theta_0 = \int x dF_1(x)$ and $\Delta_0 = \int y dF_2(y) - \int x dF_1(x)$. We obtain (1) and (2) by taking

$$w_1(X, \theta_0, \Delta_0, t) = X - \theta_0, \quad w_2(Y, \theta_0, \Delta_0, t) = Y - \theta_0 - \Delta_0.$$

Example 2. The difference of smooth M-estimators. Let θ_0 and θ_1 be smooth location M-estimators for samples X and Y, respectively. Then $\Delta_0 = \theta_1 - \theta_0$ and

$$w_1(X, \theta_0, \Delta_0, t) = \tilde{\psi}\left(\frac{X - \theta_0}{\hat{\sigma}_1}\right), \quad w_2(Y, \theta_0, \Delta_0, t) = \tilde{\psi}\left(\frac{Y - \theta_0 - \Delta_0}{\hat{\sigma}_2}\right),$$

where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are scale estimators, and $\tilde{\psi}$-function corresponds to the smooth M-estimator introduced in Hampel, Hennig and Ronchetti (2011).

Example 3 (Qin and Zhao, 2000)**.** The difference of two sample distribution functions. Denote $\theta_0 = F_1(t)$ and $\Delta_0 = F_2(t) - F_1(t)$. In this case we obtain (1) and (2) by taking

$$w_1(X, \theta_0, \Delta_0, t) = I_{\{X \le t\}} - \theta_0, \quad w_2(Y, \theta_0, \Delta_0, t) = I_{\{Y \le t\}} - \theta_0 - \Delta_0.$$

Example 4. The difference of two sample quantile functions. Zhou and Jing (2003) established the EL method in the one-sample case for quantile differences. This can be seen as a special case from our result. Denote $\theta_0 = F_1^{-1}(t)$ and $\Delta_0 = F_2^{-1}(t) - F_1^{-1}(t)$. In this case we obtain (1) and (2) by taking

$$w_1(X, \theta_0, \Delta_0, t) = I_{\{X \le \theta_0\}} - t, \quad w_2(Y, \theta_0, \Delta_0, t) = I_{\{Y \le \theta_0 + \Delta_0\}} - t.$$

Example 5 (Claeskens, Jing, Peng and Zhou, 2003)**.** Receiver operating characteristic (ROC) curve. ROC curve is defined as $\Delta = 1 - F_1(F_2^{-1}(1-t))$, which is important tool used to summarize the performance of a medical diagnostic test for determining whether a patient has a disease or not. Denote $\theta_0 = F_2^{-1}(1-t)$ and $\Delta_0 = 1 - F_1(F_2^{-1}(1-t))$. In this case

$$w_1(X, \theta_0, \Delta_0, t) = I_{\{X \le \theta_0\}} + \Delta_0 - 1, \quad w_2(Y, \theta_0, \Delta_0, t) = I_{\{Y \le \theta_0\}} + t - 1.$$

Example 6 (Claeskens, Jing, Peng and Zhou, 2003)**.** Probability-probability (P-P) plot. ROC curves are closely related to P-P plots, so the results for ROC curves can easily be applied for P-P plots. Denote $\theta_0 = F_2^{-1}(t)$ and $\Delta_0 = F_1(F_2^{-1}(t))$, which is the P-P plot of functions F_1 and F_2. In this case

$$w_1(X, \theta_0, \Delta_0, t) = I_{\{X \le \theta_0\}} - \Delta_0, \quad w_2(Y, \theta_0, \Delta_0, t) = I_{\{Y \le \theta_0\}} - t.$$

Example 7. Quantile-quantile (Q-Q) plot. Denote $\theta_0 = F_2(t)$ and $\Delta_0 = F_1^{-1}(F_2(t))$, which is well known as a Q-Q plot. We have

$$w_1(X, \theta_0, \Delta_0, t) = I_{\{X \le \Delta_0\}} - \theta_0, \quad w_2(Y, \theta_0, \Delta_0, t) = I_{\{Y \le t\}} - \theta_0.$$

Example 8. Structural relationship model. Freitag and Munk (2005) have introduced the notion of *structural relationship model* which generalizes the simple location, the location-scale and the Lehmann alternative models. It has the following form

$$F_1(t) = \phi_2^-(F_2(\phi_1^-(t, h)), h), \quad t \in \mathbb{R}, \tag{3}$$

where ϕ_1, ϕ_2 are some real-valued functions, ϕ_i^- denotes the inverse function with respect to the first argument, and $h \in \mathcal{H} \subseteq \mathbb{R}^l$, where l is some positive constant. In case of the simple location model $F_1(t) = F_2(t - h)$ we have $\phi_1(t, h) = t + h$ and $\phi_2(u, h) = u$. To check whether such relationship models hold for a fixed parameter h one can make inference for the function

$$\Delta := \Delta(t) = F_1(\phi_1(F_2^{-1}(\phi_2(t, h)), h)),$$

which is a generalization of the P-P plot considered in Example 6. In this case the estimating functions are

$$w_1(X, \theta_0, \Delta_0, t) = I_{\{X \le \theta_0\}} - \Delta_0, \quad w_2(Y, \theta_0, \Delta_0, t) = I_{\{Y \le \phi_1^-(\theta_0, h_0)\}} - \phi_2(t, h_0),$$

where $\theta_0 = \phi_1(F_2^{-1}(\phi_2(t, h_0)), h_0)$.

Remark 1. *In practice it is more appealing to check whether the structural relationship model* (3) *holds for any h, which correspond to composite hypothesis. In this case one can estimate the structural parameter h using the Mallow's distance and use two-sample plug-in empirical likelihood method introduced in Valeinis (2007).*

When introducing the two-sample empirical likelihood method Qin and Zhao (2000) basically refer to the paper of Qin and Lawless (1994) for most of the proofs. For Examples 1, 2, 3 and 7 this is possible because of the smoothness of the estimating equations (1) and (2) as the functions of the parameter θ. When w_1 and w_2 are nonsmooth (Examples 4, 5, 6 and 8) the proving technique of Qin and Lawless (1994) based on the Taylor series expansions is not valid anymore.

To overcome this problem we will use the smoothed empirical likelihood introduced for quantile inference in the one-sample case by Chen and Hall (1993). This idea also was used by Claeskens, Jing, Peng and Zhou (2003) for ROC curves and P-P plots, and Zhou and Jing (2003) for the difference of two quantiles in the one-sample case. In a recent paper by Lopez, Van Keilegom and Veraverbeke (2009) a different technique has been used not requiring the smoothness of w_1 and w_2. However, among other conditions they need smoothness conditions on the mathematical expectation as a function of w_1 and w_2, which can sometimes require further conditions on the underlying distributions (for example, for quantile inference the underlying distribution function should be twice continuously differentiable in the neighbourhood of the true parameter). We stress that one of the advantages of our approach is that it was possible to produce a rather simple R code for all two-sample problems based on smoothed estimating functions.

The paper is organized as follows. In Section 2 we introduce the general empirical likelihood method in the two-sample case based on functions w_1 and w_2 introduced in (1) and (2). Section 3 introduces the smoothed empirical likelihood method in the two-sample case. The simulation results and a practical data example are shown in Section 4. All proofs are deferred to the Appendix.

2 EL in the two-sample case

Following the ideas of Qin and Zhao (2000) in this section we introduce the empirical likelihood method for two-sample problems. For fixed $t \in \mathcal{T}$ to obtain confidence regions for the function Δ, we define the profile empirical likelihood ratio function

$$R(\Delta, \theta) = \sup_{p,q} \prod_{i=1}^{n_1} (n_1 p_i) \prod_{j=1}^{n_2} (n_2 q_j), \tag{4}$$

where $p = (p_1, \ldots, p_{n_1})$ and $q = (q_1, \ldots, q_{n_2})$ are two probability vectors, that is, consisting of nonnegative numbers adding to one.

A unique solution of (4) exists, provided that 0 is inside the convex hull of the points $w_1(X_i, \theta, \Delta, t)$'s and the convex hull of the $w_2(Y_j, \theta, \Delta, t)$'s. The maximum may be found using the standard Lagrange multipliers method giving

$$p_i = \frac{1}{n_1(1 + \lambda_1(\theta) w_1(X_i, \theta, \Delta, t))}, \quad i = 1, \ldots, n_1.$$
$$q_j = \frac{1}{n_2(1 + \lambda_2(\theta) w_2(Y_j, \theta, \Delta, t))}, \quad j = 1, \ldots, n_2,$$

where the Lagrange multipliers $\lambda_1(\theta)$ and $\lambda_2(\theta)$ can be determined in the terms of θ by the equations

$$\sum_{i=1}^{n_1} \frac{w_1(X_i, \theta, \Delta, t)}{1 + \lambda_1(\theta) w_1(X_i, \theta, \Delta, t)} = 0, \tag{5}$$

$$\sum_{j=1}^{n_2} \frac{w_2(Y_j, \theta, \Delta, t)}{1 + \lambda_2(\theta) w_2(Y_j, \theta, \Delta, t)} = 0. \tag{6}$$

Finally, we define the empirical log-likelihood ratio (multiplied by minus two) as

$$W(\Delta, \theta) = -2 \log R(\Delta, \theta) = 2 \sum_{i=1}^{n_1} \log(1 + \lambda_1(\theta) w_1(X_i, \theta, \Delta, t)) + 2 \sum_{j=1}^{n_2} \log(1 + \lambda_2(\theta) w_2(Y_j, \theta, \Delta, t)).$$

To find an estimator $\hat{\theta} = \hat{\theta}(\Delta)$ for the nuisance parameter θ that maximizes $R(\Delta, \theta)$ for a fixed parameter Δ, we obtain the empirical likelihood equation by setting

$$\frac{\partial W(\Delta, \theta)}{\partial \theta} = \sum_{i=1}^{n_1} \frac{\lambda_1(\theta) \alpha_1(X_i, \theta, \Delta, t)}{1 + \lambda_1(\theta) w_1(X_i, \theta, \Delta, t)} + \sum_{j=1}^{n_2} \frac{\lambda_2(\theta) \alpha_2(Y_j, \theta, \Delta, t)}{1 + \lambda_2(\theta) w_2(Y_j, \theta, \Delta, t)} = 0, \tag{7}$$

where

$$\alpha_1(X_i,\theta,\Delta,t) = \frac{\partial w_1(X_i,\theta,\Delta,t)}{\partial\theta} \quad \text{and} \quad \alpha_2(Y_j,\theta,\Delta,t) = \frac{\partial w_2(Y_j,\theta,\Delta,t)}{\partial\theta}.$$

Theorem 2 (Qin and Zhao, 2000)**.** *Under some standard smoothness assumptions on functions w_1, w_2, α_1 and α_2 (see also Qin and Lawless, 1994), there exists a root $\hat\theta$ of (7) such that $\hat\theta$ is a consistent estimate for θ_0, $R(\Delta,\theta)$ attains its local maximum value at $\hat\theta$, and*

$$\sqrt{n_1}(\hat\theta-\theta_0) \rightarrow_d N\left(0, \frac{\beta_1\beta_2}{\beta_2\beta_{10}^2 + k\beta_1\beta_{20}^2}\right),$$

where $k<\infty$ is a positive constant such that $n_2/n_1 \to k$ as $n_1, n_2 \to \infty$,

$$-2\log R(\Delta_0,\hat\theta) \rightarrow_d \chi_1^2,$$

as $n_1, n_2 \to \infty$, for each fixed $t \in \mathcal{T}$ and $\rightarrow_d$ denotes the convergence in distribution,

$$\beta_1 = E_{F_1} w_1^2(X,\theta_0,\Delta_0,t), \quad \beta_2 = E_{F_2} w_2^2(Y,\theta_0,\Delta_0,t),$$

$$\beta_{10} = E_{F_1}\alpha_1(X,\theta_0,\Delta_0,t), \quad \beta_{20} = E_{F_2}\alpha_2(Y,\theta_0,\Delta_0,t).$$

Theorem 2 holds for Examples 1, 2, 3 and 7, when w_1 and w_2 are smooth functions of θ. In order to apply this result to other Examples we propose to use the smoothed empirical likelihood method (see next Section 3). The pointwise empirical likelihood confidence interval for each fixed $t \in \mathcal{T}$ for Δ has the following form $\Delta : \{R(\Delta,\hat\theta) > c\}$ for the true Δ_0, where $\hat\theta$ is a root of (7). The constant c can be calibrated using Theorem 2.

3 Smoothed EL in the two-sample case

By appropriate smoothing of the empirical likelihood method Chen and Hall (1993) showed that the coverage accuracy may be improved from order $n^{-1/2}$ to n^{-1}. Moreover, the smoothed empirical likelihood appears to be Bartlett-correctable. Thus, an empirical correction for scale can reduce the size of coverage error from order n^{-1} to n^{-2}. For $j = 1, 2$ let H_j denote a smoothed version of the degenerate distribution function H_0 defined by $H_0(x) = 1$ for $x \geq 0, 0$ otherwise. Define $H_j(t) = \int_{u \leq t} K_j(u)du$, where K_j is a compactly supported r-th order kernel which is commonly used in nonparametric density estimation. That is, for some integer $r \geq 2$ and constant $\kappa \neq 0$, K_j is a function satisfying

$$\int u^k K_j(u)du = \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{if } 1 \leq k \leq r-1, \\ \kappa, & \text{if } k = r. \end{cases}$$

We also define $H_{b_j}(t) = H_j(t/b_j)$, where $b_1 = b_1(n_1)$ and $b_2 = b_2(n_2)$ are bandwidth sequences, converging to zero as n_1, n_2 grows to infinity.
Let $p = (p_1, \ldots, p_{n_1})$ and $q = (q_1, \ldots, q_{n_2})$ be two vectors consisting of non-negative numbers adding to one. Define further the estimators

$$\hat{F}_{b_1,p}(x) = \sum_{i=1}^{n_1} p_i H_{b_1}(x - X_i) \quad \text{and} \quad \hat{F}_{b_2,q}(y) = \sum_{j=1}^{n_2} q_j H_{b_2}(y - Y_j).$$

For this setting we define the profile two-sample smoothed empirical likelihood ratio function for Δ as

$$R^{(sm)}(\Delta, \theta) = \sup_{p,q} \prod_{i=1}^{n_1} (n_1 p_i) \prod_{j=1}^{n_2} (n_2 q_j), \tag{8}$$

where the latter supremum is a subject to the constraints which are listed for Examples 3-8 in Table 1. For all these examples the respective smoothed estimating equations can be found in Table 2.

Table 1: Constraints for the smoothed empirical likelihood method.

Example	Constraint 1	Constraint 2
3	$\hat{F}_{b_1,p}(t) = \theta_0$	$\hat{F}_{b_2,q}(t) = \theta_0 + \Delta_0$
4	$\hat{F}_{b_1,p}(\theta_0) = t$	$\hat{F}_{b_2,q}(\theta_0 + \Delta_0) = t$
5	$\hat{F}_{b_1,p}(\theta_0) = 1 - \Delta_0$	$\hat{F}_{b_2,q}(\theta_0) = 1 - t$
6	$\hat{F}_{b_1,p}(\theta_0) = \Delta_0$	$\hat{F}_{b_2,q}(\theta_0) = t$
7	$\hat{F}_{b_1,p}(\Delta_0) = \theta_0$	$\hat{F}_{b_2,q}(t) = \theta_0$
8	$\hat{F}_{b_1,p}(\theta_0) = \Delta_0$	$\hat{F}_{b_2,q}(\phi_1^{-}(\theta_0, h_0)) = \phi_2(t, h_0)$

Table 2: Estimating equations for the smoothed empirical likelihood method.

Example	w_1	w_2
3	$H_{b_1}(t - X_i) - \theta_0$	$H_{b_2}(t - Y_j) - \theta_0 - \Delta_0$
4	$H_{b_1}(\theta_0 - X_i) - t$	$H_{b_2}(\theta_0 + \Delta_0 - Y_j) - t$
5	$H_{b_1}(\theta_0 - X_i) - 1 + \Delta_0$	$H_{b_2}(\theta_0 - Y_j) - 1 + t$
6	$H_{b_1}(\theta_0 - X_i) - \Delta_0$	$H_{b_2}(\theta_0 - Y_j) - t$
7	$H_{b_1}(\Delta_0 - X_i) - \theta_0$	$H_{b_2}(t - Y_j) - \theta_0$
8	$H_{b_1}(\theta_0 - X_i) - \Delta_0$	$H_{b_2}(\phi_1^{-}(\theta_0, h_0) - Y_j) - \phi_2(t, h_0)$

From Table 2 it is obvious that all estimating equations have a similar form. To unify all these cases let us rewrite the estimating equations as follows

$$w_1(X_i, \theta, \Delta, t) = H_{b_1}(\xi_1(\theta) - X_i) - \xi_2(\theta), \quad w_2(Y_j, \theta, \Delta, t) = H_{b_2}(\psi_1(\theta) - Y_j) - \psi_2(\theta),$$

where

$$\mathbb{E}_{F_1} w_1(X_i, \theta_0, \Delta_0, t) = 0, \quad \mathbb{E}_{F_2} w_2(Y_j, \theta_0, \Delta_0, t) = 0. \tag{9}$$

First we define the smoothed empirical likelihood estimator $\hat{\Delta}$ as follows

$$\hat{\Delta} = \arg\min_{\Delta}\{-2 \log R(\Delta, \hat{\theta})\}. \tag{10}$$

The smoothness conditions of Theorem 2 are fulfilled for the smoothed empirical likelihood function $R^{(sm)}(\Delta, \theta)$. However, in order to find out the asymptotic rates for the bandwidths b_1 and b_2 one should review the proof of Theorem 2. For ROC curves (see Example 5) the rates have been derived by Claeskens, Jing, Peng and Zhou (2003). Concerning the generalized P-P plot of structural relationship models (see Example 8) the rates can be found in Valeinis (2007).

Following Qin and Lawless (1994) and Qin and Zhao (2000) first we show that there exists a root $\hat{\theta}$ of (7) when the function $R^{(sm)}(\Delta, \hat{\theta})$ attains its local maximum value in the neighborhood of the true parameter θ_0 such as $|\theta - \theta_0| \leq n_1^{-\eta}$, where $1/3 < \eta < 1/2$. Concerning the proof of maximization problem we follow closely the proving tehnique of Qin and Lawless (1994), where they use the technique introduced by Owen (1990).

Throughout, we assume that the following conditions hold.
(A) Assume that $f_1(= F_1')$ and $f_1^{(r-1)}$ exist in a neighborhood of $\xi_1(\theta_0)$ and are continuous at $\xi_1(\theta_0)$. Assume the same for the density function f_2 in a neighborhood of $\psi_1(\theta_0)$.
(B) As $n_1, n_2 \to \infty$, $n_2/n_1 \to k$, where $0 < k < \infty$.

The condition (A) is a standard one (see also Chen and Hall, 1993 or Claeskens, Jing, Peng and Zhou, 2003), which requires that the respective density functions are smooth enough in a neighborhood of $\xi_1(\theta_0)$ and $\psi_1(\theta_0)$. Furthermore the condition (B) assures the same growth rate for both samples.

Lemma 3. *Assume conditions (A) and (B) hold. Then with the probability converging to 1 as $n_1, n_2 \to \infty$, $R^{(sm)}(\Delta, \theta)$ attains its maximum value at some point $\hat{\theta}$ in the neighbourhood of θ_0 for all the cases considered in Examples 3-8. More specifically there exists a root $\hat{\theta}$ of equation (7), such that $|\hat{\theta} - \theta_0| \le n_1^{-\eta}$, where $1/3 < \eta < 1/2$.*

Next we derive the limiting distribution of the statistic $-2\log R^{(sm)}(\Delta_0, \hat{\theta})$, which as expected is the chi-squared distribution with the degree of freedom one.

Lemma 4. *In addition to the conditions of Lemma 3 assume that for $i = 1, 2$,*

$$n_i b_i^{4r} \to 0 \tag{11}$$

as $n_1, n_2 \to \infty$. Then for all cases considered in Examples 3-8 it follows that

$$\sqrt{n_1}(\hat{\theta} - \theta_0) \to_d N\left(0, \frac{\beta_1\beta_2}{\beta_2\beta_{10}^2 + k\beta_1\beta_{20}^2}\right),$$

where $\beta_1 = \xi_2(\theta_0)(1 - \xi_2(\theta_0)), \beta_2 = \psi_2(\theta_0)(1 - \psi_2(\theta_0)), \beta_{10} = f_1(\xi_1(\theta_0))\xi_1'(\theta_0)$, $\beta_{20} = f_2(\psi_1(\theta_0))\psi_1'(\theta_0)$. We also have

$$\lambda_1(\hat{\theta}) = -k\frac{\beta_{20}}{\beta_{10}}\lambda_2(\hat{\theta}) + o_p(n_1^{-1/2}),$$

$$\sqrt{n_1}\lambda_2(\hat{\theta}) \to_d N\left(0, \frac{\beta_{10}^2}{k\beta_2\beta_{10}^2 + k^2\beta_1\beta_{20}^2}\right).$$

Theorem 5. *In addition to the conditions of Lemma 3 assume that for* $i = 1, 2$ *we have*

$$n_i b_i^{3r} \to 0 \tag{12}$$

as $n_1, n_2 \to \infty$. *Then*

$$-2 \log R^{(sm)}(\Delta_0, \hat{\theta}) \to_d \chi_1^2.$$

4 Simulation study and a data example

To illustrate the method we have performed a coverage accuracy simulation study for Examples 3, 4, 6 and 7 from Section 1 and compared it to the coverage accuracies obtained by several bootstrap methods. Examples 5 and 8 were omitted because they can essentially be transformed to Example 6, while a coverage accuracy study for Example 1 is given in Valeinis, Cers and Cielens (2010). The results on Example 2 have been recently discussed in the International Conference on Robust statistics 2011 and will be reported elsewhere. Finally, the method is also demonstrated on a real data example.

We implemented the empirical likelihood method in **R**. The program was written to be as general as possible, so that implementing each example from Section 1 required only minimal additional programming. Besides the estimating functions (1) and (2), the search intervals for the Δ (when searching for the optimal value) and θ must be calculated differently for each example. The intervals are calculated from the requirement of 0 being inside the convex hulls of the points $w_1(X_i, \theta, \Delta, t)$ for $i = 1, 2, \ldots, n_1$ and $w_2(Y_j, \theta, \Delta, t)$ for $j = 1, 2, \ldots, n_2$. This criterion is sufficient to find intervals for θ and Δ where all equations (5), (6) and (7) have solutions and it is possible to calculate the log-likelihood ratio.

The program calculates the (smoothed) log-likelihood ratio (4) defined also in (8) for a given value of Δ by solving (7) for θ, implicitly finding λ_1 and λ_2 from (5) and (6). To estimate the true parameter Δ_0 we minimize the log-likelihood over the possible values of Δ. To construct pointwise confidence intervals we solve for values of Δ where the log-likelihood ratio statistic equals the critical values from the χ_1^2 distribution. We solve for the solutions using the built in function `uniroot`, and minimize using the built in function `optimize`. The program can be found on the corresponding author homepage.

We constructed 95% coverage accuracies for the difference of two sample distribution and quantile functions, P-P plots and Q-Q plots (Examples 3, 4, 6 and 7 from Section 1) shown in Tables 3 and 4. Each coverage accuracy was constructed by taking 10,000 pseudo-random samples from the $N(0,1)$ distribution. For smoothing with the empirical likelihood method the standard normal kernel was used with three sample size dependent values of the bandwidth parameter $b_i = \{b_{i1} = n_i^{-0.1}, b_{i1} = n_i^{-0.2}, b_{i1} = n_i^{-0.3}\}$ for $i = 1, 2$. For the comparison, the coverage accuracies by the normal and the percentile bootstrap were also constructed and are shown in the Tables along with the empirical likelihood coverage accuracies.

It can be seen from Tables 3 and 4, that the coverage accuracies are generally converging towards 0.95 as the sample size increases. However, there are significant differences depending on the statistic and the point at which the coverage accuracies are calculated. For all statistics, except the difference of two sample distribution functions, the method performs significantly better at the center point ($t = 0.5$ for P-P plots and quantile differences; $t = 0$ for Q-Q plots), for which it is very near to 0.95 already from the smallest sample sizes. For all statistics the empirical likelihood converges to 0.95 fairly quickly, and at sample size $n_1 = n_2 = 50$ all coverage accuracies are already very near to 0.95.

The empirical likelihood outperforms the bootstrap for P-P plots, Q-Q plots and sample distribution function differences, while the bootstrap was better for quantile differences. The choice of bandwidth affects the overall performance of the empirical likelihood only moderately and the effect produced by changes in sample size is greater.

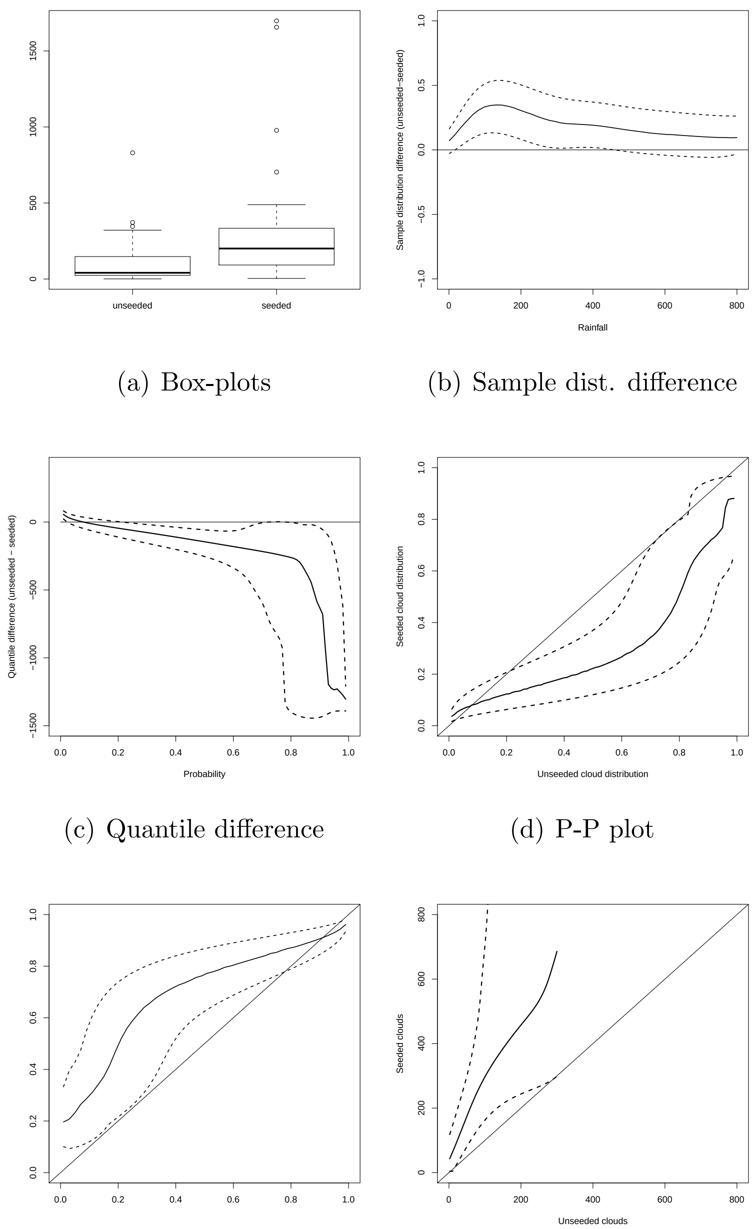
We demonstrate the practical application of the method on the data set analyzed in Simpson, Olsen and Eden (1975). The data shows the results of a number of weather modification experiments conducted between 1968 and 1972 in Florida. In the experiments some clouds were seeded with silver iodide which should result in a higher growth of cumulus clouds and, therefore, increased precipitation. The data shows the rainfall from seeded and unseeded clouds measured in acre-feet — there are 26 observations for each

Table 3: 95% coverage accuracy of pointwise confidence intervals for $\Delta_1 = F_2(t) - F_1(t)$ and $\Delta_2 = F_1^{-1}(F_2(t))$ constructed by bootstrap (B.N. — normal bootstrap, B.P. — percentile bootstrap) and by EL using different bandwidths $b_i = \{b_{i1} = n_i^{-0.1}, b_{i2} = n_i^{-0.2}, b_{i3} = n_i^{-0.3}\}$ for $i = 1, 2$ and $F_1 = F_2 = N(0, 1)$ at three different values of $t = \{-1, 0, 1\}$.

			Δ_1					Δ_2				
t	n_1	n_2	b_{i1}	b_{i2}	b_{i3}	B.N.	B.P.	b_{i1}	b_{i2}	b_{i3}	B.N.	B.P.
-1	15	15	0.920	0.920	0.907	0.726	0.602	0.874	0.868	0.869	0.844	0.687
-1	20	20	0.929	0.928	0.924	0.760	0.689	0.913	0.911	0.911	0.887	0.740
-1	30	30	0.939	0.938	0.938	0.799	0.760	0.942	0.938	0.936	0.925	0.805
-1	50	50	0.947	0.945	0.947	0.835	0.808	0.946	0.946	0.948	0.938	0.861
-1	20	30	0.933	0.933	0.933	0.868	0.845	0.934	0.933	0.934	0.905	0.764
-1	30	20	0.938	0.930	0.933	0.872	0.848	0.920	0.920	0.921	0.904	0.783
0	15	15	0.939	0.940	0.943	0.784	0.759	0.939	0.939	0.939	0.941	0.867
0	20	20	0.946	0.944	0.945	0.804	0.792	0.942	0.942	0.945	0.945	0.874
0	30	30	0.947	0.947	0.947	0.846	0.830	0.948	0.945	0.946	0.950	0.887
0	50	50	0.947	0.947	0.946	0.863	0.850	0.946	0.948	0.949	0.942	0.892
0	20	30	0.945	0.941	0.944	0.896	0.890	0.947	0.949	0.946	0.942	0.864
0	30	20	0.945	0.950	0.945	0.894	0.890	0.942	0.944	0.945	0.947	0.901
1	15	15	0.922	0.918	0.907	0.735	0.607	0.877	0.872	0.859	0.849	0.688
1	20	20	0.934	0.930	0.923	0.762	0.691	0.911	0.915	0.909	0.887	0.740
1	30	30	0.940	0.936	0.939	0.799	0.760	0.940	0.936	0.941	0.919	0.816
1	50	50	0.945	0.944	0.947	0.822	0.796	0.947	0.948	0.948	0.942	0.864
1	20	30	0.930	0.936	0.929	0.866	0.842	0.933	0.931	0.921	0.905	0.772
1	30	20	0.932	0.930	0.931	0.869	0.841	0.916	0.919	0.916	0.901	0.790

Table 4: 95% coverage accuracy of pointwise confidence intervals for $\Delta_1 = F_2^{-1}(t) - F_1^{-1}(t)$ and $\Delta_2 = F_1(F_2^{-1}(t))$ constructed by bootstrap (B.N. — normal bootstrap, B.P. — percentile bootstrap) and by EL using different bandwidths $b_i = \{b_{i1} = n_i^{-0.1}, b_{i2} = n_i^{-0.2}, b_{i3} = n_i^{-0.3}\}$ for $i = 1, 2$ and $F_1 = F_2 = N(0, 1)$ at three different values of $t = \{0.1, 0.5, 0.9\}$.

			Δ_1					Δ_2				
t	n_1	n_2	b_{i1}	b_{i2}	b_{i3}	B.N.	B.P.	b_{i1}	b_{i2}	b_{i3}	B.N.	B.P.
0.1	15	15	0.879	0.833	0.797	0.934	0.949	0.895	0.896	0.897	0.787	0.781
0.1	20	20	0.911	0.892	0.875	0.950	0.966	0.912	0.916	0.919	0.851	0.830
0.1	30	30	0.928	0.929	0.926	0.954	0.965	0.931	0.933	0.941	0.912	0.898
0.1	50	50	0.942	0.947	0.951	0.950	0.966	0.945	0.948	0.948	0.947	0.952
0.1	20	30	0.914	0.912	0.894	0.943	0.960	0.924	0.922	0.929	0.855	0.842
0.1	30	20	0.919	0.914	0.896	0.951	0.962	0.918	0.924	0.932	0.901	0.883
0.5	15	15	0.949	0.952	0.952	0.953	0.966	0.946	0.949	0.950	0.928	0.967
0.5	20	20	0.943	0.948	0.952	0.944	0.962	0.947	0.953	0.950	0.930	0.950
0.5	30	30	0.950	0.954	0.951	0.948	0.964	0.949	0.948	0.949	0.935	0.951
0.5	50	50	0.950	0.949	0.950	0.948	0.962	0.953	0.951	0.950	0.945	0.956
0.5	20	30	0.948	0.947	0.951	0.948	0.962	0.950	0.947	0.945	0.940	0.950
0.5	30	20	0.949	0.945	0.954	0.948	0.962	0.948	0.947	0.950	0.933	0.952
0.9	15	15	0.894	0.891	0.880	0.929	0.943	0.903	0.897	0.898	0.935	0.945
0.9	20	20	0.915	0.909	0.911	0.950	0.963	0.913	0.914	0.914	0.940	0.932
0.9	30	30	0.930	0.935	0.937	0.953	0.965	0.929	0.935	0.938	0.945	0.940
0.9	50	50	0.940	0.946	0.953	0.949	0.965	0.942	0.945	0.947	0.944	0.947
0.9	20	30	0.919	0.920	0.924	0.947	0.961	0.923	0.926	0.925	0.934	0.939
0.9	30	20	0.914	0.920	0.920	0.947	0.958	0.922	0.919	0.931	0.947	0.935



(a) Box-plots

(b) Sample dist. difference

(c) Quantile difference

(d) P-P plot

(e) ROC curve

(f) Q-Q plot

Figure 1: Different two-sample plots for the cloud seeding example, constructed using smoothed empirical likelihood. All plots show also the 95% pointwise confidence bands. A box-plot (a) was added for visual comparison of both samples.

case.

When comparing means, the t-test assigns a p-value of 0.029 to the hypothesis of the two means being equal, while the empirical likelihood method (Example 1 from Section 1) assigns a p-value of only 0.007. The difference can be explained by the non-normality of the data coupled with the small data sizes.

Figure 1 shows Examples 3, 4, 5, 6 and 7 from Section 1 with the smoothed empirical likelihood estimator (10) and 95% pointwise confidence intervals. For each sample we used the function `bw.nrd` in **R**, which implements a rule-of-thumb for choosing the bandwidth of a Gaussian kernel density estimator. All the results seem to indicate that the seeding has had an effect on the rainfall. Finally, note that simultaneous confidence bands can be constructed using the empirical likelihood pointwise intervals combined by the bootstrap method as first shown in Hall and Owen (1993). For application examples for this method see also Claeskens, Jing, Peng and Zhou (2003) and Valeinis, Cers and Cielens (2010).

From our simulation study we find that for normally distributed data the empirical likelihood method is comparable with the t-test (see Valeinis, Cers and Cielens, 2010) and bootstrap methods. Clearly it is advantageous due to asymmetric confidence intervals, which can reflect some interesting features in data sets. Moreover, it is a nonparametric procedure, which does not require the normality of the data. Therefore, in special cases it can be advantageous over classical parametric procedures.

5 Appendix

First we need a technical lemma, which will be used proving the main theorem.

i) Technical lemma

Lemma 6. *Assume that condition (i) is satisfied. For some fixed θ, such that $|\theta - \theta_0| \leq n_1^{-\eta}$ we have*

$$\begin{aligned}
\mathbb{E}\{w_1(X,\theta,\Delta,t)\} &= F_1(\xi_1(\theta)) - F_1(\xi_1(\theta_0)) + \zeta n_1^{-\eta} + O(b_1^r), && (13)\\
\mathbb{E}\{w_2(Y,\theta,\Delta,t)\} &= F_2(\psi_1(\theta)) - F_2(\psi_2(\theta_0)) + \zeta n_2^{-\eta} + O(b_2^r), && (14)\\
var\{w_1(X,\theta,\Delta,t)\} &= F_1(\xi_1(\theta))(1 - F_1(\xi_1(\theta))) + O(b_1), && (15)\\
var\{w_2(Y,\theta,\Delta,t)\} &= F_1(\psi_1(\theta))(1 - F_1(\psi_1(\theta))) + O(b_2), && (16)\\
\mathbb{E}\{\alpha_1(X,\theta,\Delta,t)\} &= f_1(\xi_1(\theta))\xi_1'(\theta) + \zeta + O(b_1^r), && (17)\\
\mathbb{E}\{\alpha_2(Y,\theta,\Delta,t)\} &= f_2(\psi_1(\theta))\psi_1'(\theta) + \zeta + O(b_2^r). && (18)
\end{aligned}$$

Proof. Let us do some simple calculations involving Taylor expansions. We will do it only for the function w_1. Integration by parts and variable transformation gives

$$\begin{aligned}
&\mathbb{E}\left\{H_1\left(\frac{\xi_1(\theta) - X}{b_1}\right)\right\} - F_1(\xi_1(\theta))\\
&= \int_{-\infty}^{+\infty} H_1\left(\frac{\xi_1(\theta) - x}{b_1}\right) dF_1(x) - F_1(\xi_1(\theta))\\
&= \int_{-\infty}^{\infty} \{F_1(\xi_1(\theta) - b_1 u) - F_1(\xi_1(\theta))\} K_1(u) du\\
&= \int_{-\infty}^{\infty} \left\{f_1(\xi_1(\theta))(-b_1 u) + \ldots + \frac{1}{r!} f_1^{(r-1)}(\xi_1(\theta))(-b_1 u)^r + o(b_1^r)\right\} K_1(u) du\\
&= O(b_1^r).
\end{aligned}$$

It follows

$$\begin{aligned}
\mathbb{E}\{w_1(X,\theta,\Delta,t) &= \mathbb{E}\left\{H_1\left(\frac{\xi_1(\theta) - X}{b_1}\right)\right\} - \xi_2(\theta)\\
&= F_1(\xi_1(\theta)) - \xi_2(\theta) + O(b_1^r)\\
&= F_1(\xi_1(\theta)) - \xi_2(\theta_0) - \xi_2'(\theta^*) n_1^{-\eta} + O(b_1^r)\\
&= F_1(\xi_1(\theta)) - F_1(\xi_1(\theta_0)) + \zeta n_1^{-\eta} + O(b_1^r)
\end{aligned}$$

where $\theta^* \in [\theta_0, \theta]$ and the last assertion follows from (9) with $\zeta = -1$ for Examples 3 and 7 and $\zeta = 0$ for Examples 4, 5, 6, 8 (see Table 2). Thus we have shown (13) and (14).

$$
\begin{aligned}
\mathbb{E}\left\{H_1^2\left(\frac{\xi_1(\theta) - X}{b_1}\right)\right\} &= \int_{-\infty}^{+\infty} H_1^2\left(\frac{\xi_1(\theta) - y}{b_1}\right) dF_1(y) \\
&= 2\int_{-\infty}^{\infty} \{F_1(\xi_1(\theta) - b_1 u)\} H_1(u) dH_1(u) \\
&= F_1(\xi_1(\theta) \int_{-\infty}^{\infty} dH_1^2(u) - 2\int_{-\infty}^{\infty} b_1 u f_1(\mu^*) H_1(u) dH_1(u) \\
&= F_1(\xi_1(\theta)) + O(b_1),
\end{aligned}
$$

where $\mu^* \in [\xi_1(\theta) - b_1 u, \xi_1(\theta)]$. Therefore we have

$$
var\{w_1(X, \theta, \Delta, t)\} = F_1(\xi(\theta)) - F_1^2(\xi(\theta)) + O(b_1)
$$

and both (15) and (16) follow. At last

$$
\begin{aligned}
\mathbb{E}\{\alpha_1(X, \theta, \Delta, t)\} &= \int_{-\infty}^{+\infty} \left(\frac{\partial H_{b_1}(\xi_1(\theta) - x)}{\partial \theta} - \xi_2'(\theta)\right) dF(x) \\
&= \frac{1}{b_1}\int_{-\infty}^{+\infty} K_1\left(\frac{\xi_1(\theta) - x}{b_1}\right) \xi_1'(\theta) dF(x) + \zeta \\
&= \xi_1'(\theta)\int_{-\infty}^{+\infty} f_1(\xi_1(\theta) - b_1 y) K_1(y) dy + \zeta \\
&= f_1(\xi_1(\theta))\xi_1'(\theta) + \zeta + O(b_1^r).
\end{aligned}
$$

Therefore the equations (17) and (18) hold.

ii) *Proof of Lemma 3.*

First for some $k \in \mathbb{N}$ denote by

$$
\bar{w}_1^k(X, \theta, \Delta, t) = \frac{1}{n_1}\sum_{i=1}^{n_1} w_1^k(X_i, \theta, \Delta, t), \;\; \bar{\alpha}_1^k(X, \theta, \Delta, t) = \frac{1}{n_1}\sum_{i=1}^{n_1} \alpha_1^k(X_i, \theta, \Delta, t).
$$

Then for fixed θ such that $|\theta - \theta_0| \le n^{-\eta}$ the following is true

$$
\bar{w}_1(X, \theta, \Delta_0, t) = \bar{w}_1(X, \theta_0, \Delta_0, t) + \bar{\alpha}_1(X, \theta^*, \Delta_0, t)(\theta - \theta_0), \qquad (19)
$$

for some $\theta^* \in [\theta_0, \theta]$. From strong law of large numbers (see, for example, Serfling, 1980) and Lemma 6

$$\begin{aligned}\bar{w}_1(X, \theta_0, \Delta_0, t) &= E(w_1(X, \theta_0, \Delta_0, t)) + O(n_1^{-1/2} \ln(n_1)^{1/2}) \\ &= O(b_1^r) + O(n_1^{-1/2} \ln(n_1)^{1/2}) \;\; a.s.\end{aligned}$$

and

$$\bar{\alpha}_1(X, \theta^*, \Delta_0, t) = f_1(\xi_1(\theta^*))\xi_1'(\theta^*) + \zeta + O(b_1^r) + O(n_1^{-1/2} \ln(n_1)^{1/2}) = O(1) \;\; a.s.$$

Therefore from (19) we have $\bar{w}_1(X, \theta, \Delta_0, t) = O(b_1^r) + O(n^{-\eta})$ almost surely. From definition of λ_1, the following inequality holds, where the first term is equal to zero,

$$\begin{aligned}&n_1^{-1} \left| \sum_{i=1}^{n_1} \{\lambda_1(\theta) w_1^2(X_i, \theta, \Delta_0, t)(1 + \lambda_1(\theta) w_1(X_i, \theta, \Delta_0, t))^{-1} - w_1(X_i, \theta, \Delta_0, t)\} \right| \\ &\geq |\lambda_1(\theta)|(1 + |\lambda_1(\theta)| \max_{1 \leq i \leq n_1} |w_1(X_i, \theta, \Delta_0, t)|)^{-1} \bar{w}_1^2(X, \theta, \Delta_0, t) - \bar{w}_1(X, \theta, \Delta_0, t).\end{aligned}$$

Similarly as in Qin and Lawless (1994) and Owen (1990) we conclude that

$$\begin{aligned}&|\lambda_1(\theta)|(1 + |\lambda_1(\theta)| \max_{1 \leq i \leq n_1} |w_1(X_i, \theta, \Delta_0, t)|)^{-1} \bar{w}_1^2(X, \theta, \Delta_0, t) \\ &\qquad = O(b_1^r) + O(n^{-\eta}). \quad (20)\end{aligned}$$

Clearly for all Examples 3-8 (see Table 2) we have $\max_{1 \leq i \leq n_1} |w_1(X_i, \theta, \Delta_0, t)| \leq c$ for some constant c. Again by the strong law of large numbers and Lemma 6,

$$\begin{aligned}&\bar{w}_1^2(X, \theta, \Delta_0, t) \\ &\quad = F_1(\xi_1(\theta))(1 - F_1(\xi_1(\theta))) + O(b_1) + O(n_1^{-1/2} \ln(n_1)^{1/2}) = O(1). \quad (21)\end{aligned}$$

Finally from equations (20) and (21) we conclude that almost surely $\lambda_1(\theta) = O(b_1^r) + O(n_1^{-\eta})$ when θ such as $|\theta - \theta_0| \leq n_1^{-\eta}$. Also note that $\lambda_1(\theta_0) = O(b_1^r) + O(n_1^{-1/2} \ln(n_1)^{1/2})$ almost surely. From (5) we have

$$\begin{aligned}0 = \frac{1}{n_1} \sum_{i=1}^{n_1} & w_1(X_i, \theta, \Delta_0, t) \\ & \left(1 - \lambda_1(\theta) w_1(X_i, \theta, \Delta_0, t) + \frac{\lambda_1^2(\theta) w_1^2(X_i, \theta, \Delta_0, t)}{1 + \lambda_1(\theta) w_1(X_i, \theta, \Delta_0, t)}\right) \quad (22)\end{aligned}$$

which leads to

$$0 = \bar{w}_1(X, \theta, \Delta, t) - \lambda_1(\theta)\bar{w}_1^2(X, \theta, \Delta_0, t) + \frac{1}{n_1}\sum_{i=1}^{n_1} \frac{\lambda_1^2(\theta) w_1^3(X_i, \theta, \Delta_0, t)}{1 + \lambda_1(\theta) w_1(X_i, \theta, \Delta_0, t)}.$$

The modulo from the last term is bounded by

$$\begin{aligned} &\frac{1}{n_1}\sum_{i=1}^{n_1} |w_1^3(X_i, \theta, \Delta_0, t)||\lambda_1(\theta)|^2 |1 + \lambda_1(\theta) w_1(X_i, \theta, \Delta_0, t)|^{-1} \\ &\qquad\qquad = O(1)O(b_1^r + n_1^{-\eta})^2 O(1) = O(b_1^r + n_1^{-\eta})^2 \quad a.s. \end{aligned}$$

Therefore

$$\lambda_1(\theta) = \frac{\bar{w}_1(X, \theta, \Delta_0, t)}{\bar{w}_1^2(X, \theta, \Delta_0, t)} + O(b_1^r + n_1^{-\eta})^2 = O(b_1^r + n^{-\eta}). \tag{23}$$

uniformly for $\theta \in \{\theta : |\theta - \theta_0| \leq n_1^{-\eta}\}$. Put

$$\begin{aligned} H_1(\theta, \Delta) &= \sum_{i=1}^{n_1} \log(1 + \lambda_1(\theta) w_1(X_i, \theta, \Delta, t)), \\ H_2(\theta, \Delta) &= \sum_{j=1}^{n_2} \log(1 + \lambda_2(\theta) w_2(Y_j, \theta, \Delta, t)). \end{aligned}$$

Hence the test statistic $-2\log R^{sm}(\theta, \Delta) = 2H_1(\theta, \Delta) + 2H_2(\theta, \Delta)$. Note that

$$\log(1 + x) = x - \frac{1}{2}x^2 + O(x^3).$$

By (23) and Taylor expansion we have almost surely

$$\begin{aligned} &H_1(\theta, \Delta_0) \\ &= \sum_{i=1}^{n_1} \lambda_1(\theta) w_1(X_i, \theta, \Delta_0, t) - \frac{1}{2}\sum_{i=1}^{n_1} \lambda_1^2(\theta) w_1^2(X_i, \theta, \Delta_0, t) + O(n_1\lambda_1^3(\theta))O(1) \\ &= \frac{n_1}{2}\left(\bar{w}_1(X, \theta, \Delta_0, t)\right)^2 \left(\bar{w}_1^2(X, \theta, \Delta_0, t)\right)^{-1} + O(n_1)O(b_1^r + n_1^{-\eta})^3 \\ &= \frac{n_1}{2}\left(\bar{w}_1(X, \theta_0, \Delta_0, t) + \bar{\alpha}_1(X, \theta_0, \Delta_0, t)(\theta - \theta_0)\right)^2 O(1) + O(n_1)O(b_1^r + n_1^{-\eta})^3 \\ &= \frac{n_1}{2}\left(O(b_1^r) + O(n_1^{-1/2}\ln(n_1)^{1/2}) + O(1)O(n_1^{-\eta})\right)^2 + O(n_1)O(b_1^r + n_1^{-\eta})^3 \\ &= O(n_1)O(b_1^r + n_1^{-\eta})^2. \end{aligned}$$

Similarly,

$$
\begin{aligned}
&H_1(\theta_0, \Delta_0) \\
&= \frac{n_1}{2}\left(\bar{w}_1(X_i, \theta_0, \Delta_0, t)\right)^2 \left(\bar{w}_1^2(X_i, \theta_0, \Delta_0, t)\right)^{-1} + O(n_1)O(b_1^r + n_1^{-1/2}\ln(n_1)^{1/2})^3. \\
&= \frac{n_1}{2}(O(b_1^r) + O(n_1^{-1/2}\ln(n_1)^{1/2}))^2 + O(n_1)O(b_1^r + n_1^{-1/2}\ln(n_1)^{1/2})^3. \\
&= O(n_1)(O(b_1^r + n_1^{-1/2}\ln(n_1)^{1/2})^2.
\end{aligned}
$$

Since $H(\theta, \Delta)$ is continuous with respect to the parameter θ as $|\theta - \theta_0| \leq n^{-\eta}$, it has the minimum value in the interior of the interval. A similar proof can be carried out for the function $H_2(\theta, \Delta)$. Therefore also $-2\log R^{sm}(\theta, \Delta)$ has the minimum value in the interior of the interval $|\theta - \theta_0| \leq n_1^{-\eta}$. The value $\hat{\theta}$ obtained at this minimum point satisfies (7).

iii) *Proof of Lemma 4*

For $j = 1, 2$ denote $\hat{\lambda}_j = \lambda_j(\hat{\theta})$ and

$$
Q_1(\theta, \lambda_1, \lambda_2) = \frac{1}{n_1}\sum_{i=1}^{n_1} \frac{w_1(X_i, \theta, \Delta, t)}{1 + \lambda_1(\theta) w_1(X_i, \theta, \Delta, t)},
$$

$$
Q_2(\theta, \lambda_1, \lambda_2) = \frac{1}{n_2}\sum_{j=1}^{n_2} \frac{w_2(Y_j, \theta, \Delta, t)}{1 + \lambda_2(\theta) w_2(Y_j, \theta, \Delta, t)},
$$

$$
\begin{aligned}
Q_3(\theta, \lambda_1, \lambda_2) = \lambda_1(\theta) \sum_{i=1}^{n_1} \frac{\alpha_1(X_i, \theta, \Delta, t)}{1 + \lambda_1(\theta) w_1(X_i, \theta, \Delta, t)} \\
+ \lambda_2(\theta) \sum_{j=1}^{n_2} \frac{\alpha_2(Y_j, \theta, \Delta, t)}{1 + \lambda_2(\theta) w_2(Y_j, \theta, \Delta, t)}.
\end{aligned}
$$

From Lemma 3, we have

$$
Q_i(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) = 0 \text{ for } i = 1, 2, 3.
$$

By Taylor expansion we have

$$
\begin{aligned}
0 = Q_i(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) = Q_i(\theta_0, 0, 0) + \frac{\partial Q_i(\theta_0, 0, 0)}{\partial \theta}(\hat{\theta} - \theta_0) \\
+ \frac{\partial Q_i(\theta_0, 0, 0)}{\partial \lambda_1}\hat{\lambda}_1 + \frac{\partial Q_i(\theta_0, 0, 0)}{\partial \lambda_2}\hat{\lambda}_2 + O_p(b_1^r + n_1^{-\eta})^2, \quad i = 1, 2, 3.
\end{aligned}
$$

From the assumption (11), we have $O_p(b_1^r + n_1^{-\eta})^2 = o_p(n_1^{-1/2})$. From the strong law of large numbers it follows almost surely

$$\frac{\partial Q_1(\theta_0, 0, 0)}{\partial \theta} = \frac{\partial Q_1(\theta_0, 0, 0)}{\partial \lambda_1} \to f_1(\xi_1(\theta_0))\xi_1'(\theta_0),$$

$$\frac{\partial Q_2(\theta_0, 0, 0)}{\partial \theta} = \frac{\partial Q_2(\theta_0, 0, 0)}{\partial \lambda_2} \to f_2(\psi_1(\theta_0))\psi_1'(\theta_0),$$

$$\frac{\partial Q_1(\theta_0, 0, 0)}{\partial \lambda_1} \to -F_1(\xi_1(\theta_0))(1 - F_1(\xi_1(\theta_0))) = -\xi_2(\theta_0)(1 - \xi_2(\theta_0)),$$

$$\frac{\partial Q_2(\theta_0, 0, 0)}{\partial \lambda_2} \to -F_2(\psi_1(\theta))(1 - F_2(\psi_1(\theta_0))) = -\psi_2(\theta_0)(1 - \psi_2(\theta_0)).$$

Other partial derivatives evaluated at point $(\theta_0, 0, 0)$ are zero. So,

$$\begin{pmatrix} \hat{\theta} - \theta_0 \\ \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} = -S^{-1} \begin{pmatrix} Q_1(\theta_0, 0, 0) \\ Q_2(\theta_0, 0, 0) \\ 0 \end{pmatrix} + o_p(n_1^{-1/2}),$$

where the limiting matrix

$$S = \begin{pmatrix} f_1(\xi_1(\theta_0))\xi_1'(\theta_0) & -\xi_2(\theta_0)(1 - \xi_2(\theta_0)) & 0 \\ f_2(\psi_1(\theta_0))\psi_1'(\theta_0) & 0 & -\psi_2(\theta_0)(1 - \psi_2(\theta_0)) \\ 0 & f_1(\xi_1(\theta_0))\xi_1'(\theta_0) & k f_2(\psi_1(\theta_0))\psi_1'(\theta_0) \end{pmatrix}.$$

The asymptotic behaviour of $\hat{\theta}$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ now follows from

$$\sqrt{n} \begin{pmatrix} Q_1(\theta_0, 0, 0) \\ Q_2(\theta_0, 0, 0) \end{pmatrix}$$
$$\to_d N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \xi_2(\theta_0)(1 - \xi_2(\theta_0)) & 0 \\ 0 & k^{-1}\psi_2(\theta_0)(1 - \psi_2(\theta_0)) \end{pmatrix} \right)$$

and easy calculations (see also Qin and Zhao, 2000).

IV) *Proof of Theorem 5.*

First note that from assumption (12) we have almost surely

$$O(n_1)O(\lambda_1^3(\hat{\theta}))O\left(\frac{1}{n_1}\sum_{i=1}^{n_1} w_1^3(X_i, \hat{\theta}, \Delta, t)\right) = O(n_1)O(b_1^r + n_1^{-\eta})^3 O(1) = o(1).$$

Therefore using Taylor expansion,

$$\log R(\hat{\theta}, \Delta_0) = -n_1\lambda_1(\hat{\theta})\bar{w}_1(X, \hat{\theta}, \Delta_0, t) + \frac{n_1}{2}\lambda_1^2(\hat{\theta})\bar{w}_1^2(X, \hat{\theta}, \Delta_0, t)$$
$$- n_2\lambda_2(\hat{\theta})\bar{w}_1^2(Y, \hat{\theta}, \Delta_0, t) + \frac{n_2}{2}\lambda_2^2(\hat{\theta})\bar{w}_2^2(Y, \hat{\theta}, \Delta_0, t) + o_p(1),$$

Now from (5) and (6) similarly as in (22), we have

$$\bar{w}_1(X, \hat{\theta}, \Delta_0, t) = \lambda_1(\hat{\theta})\bar{w}_1^2(X, \hat{\theta}, \Delta_0, t) + O_p(b_1^r + n_1^{-\eta})^2.$$

Similarly

$$\bar{w}_2(Y, \hat{\theta}, \Delta_0, t) = \lambda_2(\hat{\theta})\bar{w}_2^2(Y, \hat{\theta}, \Delta_0, t) + O_p(b_2^r + n_2^{-\eta})^2.$$

Using

$$\bar{w}_1^2(X, \hat{\theta}, \Delta_0, t) = \xi_2(\theta_0)(1 - \xi_2(\theta_0)) + o_p(1),$$
$$\bar{w}_2^2(Y, \hat{\theta}, \Delta_0, t) = \psi_2(\theta_0)(1 - \psi_2(\theta_0)) + o_p(1),$$

and from Lemma 3 follows

$$-2\log R^{(sm)}(\Delta_0, \hat{\theta}) = n_1\lambda_1^2(\hat{\theta})\bar{w}_1^2(X, \hat{\theta}, \Delta_0, t) + n_2\lambda_2^2(\hat{\theta})\bar{w}_2^2(Y, \hat{\theta}, \Delta_0, t) + o_p(1)$$
$$= n_1 k^2 \left(\frac{f_2(\psi_1(\theta_0))\psi_1'(\theta_0)}{f_1(\xi_1(\theta_0))\xi_1'(\theta_0)}\right)^2 \lambda_2^2(\hat{\theta})\xi_2(\theta_0)(1-\xi_2(\theta_0)) + n_2\lambda_2^2(\hat{\theta})\psi_2(\theta_0)(1-\psi_2(\theta_0)) + o_p(1)$$
$$= (\sqrt{n_1}\lambda_2(\hat{\theta}))^2 \left(k^2\frac{\beta_{20}^2}{\beta_{10}^2}\beta_1 + \frac{n_2}{n_1}\beta_2\right) + o_p(1) \to_d \chi_1^2,$$

where $\beta_{10}, \beta_{20}, \beta_1$ and β_2 are defined in Lemma 4.

Acknowledgment Valeinis acknowledges a partial support of the LU project Nr. B-ZB81-100 and project 2009/0223/1DP/ 1.1.1.2.0/09/APIA/VIAA/008 of the European Social Fund.

References

Chen, S. X. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Stat.* **21**, 1166-1181.

Claeskens, G., Jing, B. Y., Peng, L. and Zhou, W. (2003). Empirical likelihood confidence regions for comparison distributions and ROC curves. *Can. J. Stat.* **31**, 173-190.

Freitag, G. and Munk, A. (2005). On Hadamard differentiability in k-sample semiparametric models–with applications to the assessment of structural relationships. *J. Multivariate Anal.* **94**, 123-158.

Hall, P. and Owen, A. B. (1993). Empirical likelihood confidence bands in density estimation. *J. Comput. Graph. Stat.* **2**, 273-289.

Hampel, F., Hennig C. and Ronchetti, E. (2011). A smoothing principle for the Huber and other location M-estimators. *Comput. Stat. Data Anal.* **55**, 324-337.

Jing, B. Y. (1995). Two-sample empirical likelihood method. *Stat. Probab. Lett.* **24**, 315-319.

Liu, Y., Zou, C. and Zhang, R. (2008). Empirical likelihood for the two-sample mean problem. *Stat. Probab. Lett.* **78**, 548-556.

Lopez, E. M. M., Van Keilegom, I. and Veraverbeke, N. (2009). Empirical likelihood for non-smooth criterion functions. *Scand. J. Stat.* **36**, 413-432.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika.* **75**, 237-249

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Stat.* **18**, 90-120.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Stat.* **22**, 300-325.

Qin, Y. and Zhao, L. (2000). Empirical likelihood ratio confidence intervals for various differences of two populations. *Syst. Sci. Math. Sci.* **13**, 23-30.

Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. Wiley, New York.

Simpson, J., Olsen, A. and Eden, J. C. (1975). A Bayesian analysis of a multiplicative treatment effect in weather modification. *Technometrics.* **17**, 161-166.

Valeinis, J. (2007). Confidence bands for structural relationship models. Ph.D. dissertation, University of Göttingen.

Valeinis, J., Cers, E. and Cielens, J. (2010). Two-sample problems in statistical data modelling. *Math. Model. Anal.* **15**, 137-151.

Zhou, W. and Jing, B. Y. (2003). Smoothed empirical likelihood confidence intervals for the difference of quantiles. *Stat. Sinica.* **13**, 83-96.