

Data Fusion Using Empirical Likelihood

Hsiao-Hsuan Wang, Yuehua Wu, Yuejiao Fu, Xiaogang Wang

Department of Mathematics and Statistics, York University, Toronto, Canada

Email: stevenw@mathstat.yorku.ca

Received September 28, 2012; revised October 30, 2012; accepted November 15, 2012

ABSTRACT

The authors propose a robust semi-parametric empirical likelihood method to integrate all available information from multiple samples with a common center of measurements. Two different sets of estimating equations are used to improve the classical likelihood inference on the measurement center. The proposed method does not require the knowledge of the functional forms of the probability density functions of related populations. The advantages of the proposed method are demonstrated through extensive simulation studies by comparing the mean squared errors, coverage probabilities and average lengths of confidence intervals with those from the classical likelihood method. Simulation results suggest that our approach provides more informative and efficient inference than the conventional maximum likelihood estimator if certain structural relationship exists among the parameters of relevant samples.

Keywords: Data Fusion; Empirical Likelihood; Robust Estimation; Multiple Samples

1. Introduction

A common problem in clinical trials and other medical research is how to accurately and efficiently estimate parameters of interest when the current sample size is too small due to cost and time constraints. Usually there might exist certain surrogate populations with low sampling cost that could provide relevant information for the population of direct inferential interest. In this article, we propose a robust semi-parametric method to integrate related information from different sources to improve the classical likelihood method.

The classical likelihood approach is arguably the most widely used method in statistical inference. It has been routinely applied in almost all the statistical applications. Despite a great success and excellent asymptotic properties, the classical likelihood has known limitations associated with making inference for small sample sizes. Consider a thought experiment as follows. Suppose that a random experiment is to toss a coin twice. The parameter of interest, denoted as θ_1 , is the probability of turning up head for this coin. The maximum likelihood estimator (MLE) of θ_1 is denoted as $\hat{\theta}_1$. If the coin is a fair one, the MLE will obtain the following: $P(\hat{\theta}_1 = 0 \text{ or } 1) = 1/2$ and $P(\hat{\theta}_1 = 1/2) = 1/2$. Thus, the MLE would have 50% chance to make a nonsensical decision by using the MLE when the sample size is only two. In addition, suppose that for some reason we cannot use this coin any more but we can flip another coin instead. In this situation, the classical likelihood approach would not consider the se-

cond experiment since it comes from a different population unless a functional relationship between the two parameters is known. If the second population is related to the first one due to some connection between these two parameters, one should be able to utilize this connection and make better statistical inference.

Different statistical methodologies have been proposed in the literature to integrate information from different sources (or populations) in a very general setting, see [1] and [2], and referees therein. Most of these methods, however, face the challenge of accurately validating or evaluating the relevance of all related information to guard against the possibility of introducing a significant bias or contaminating the current sample. In other words, the magnitude of integration must be controlled carefully and in addition likelihood weights must be chosen judiciously in order to achieve any desired improvement in statistical inference. We propose to tackle this difficult problem using a robust semi-parametric empirical likelihood method to achieve more accurate and robust inferential results.

Empirical likelihood, which was first introduced in [3], is a nonparametric method of inference based on a data-driven likelihood ratio function. It allows the statistician to employ likelihood methods, without specifying a parametric model for the data. It enjoys both the flexibility of nonparametric methods and the efficiency of parametric likelihood. As shown in [4], empirical likelihood is a prominent efficient tool in estimating parameters by incorporating estimating equations into constrained maximization of the empirical likelihood function. In the

problem we consider, the relevant information from different sources could be used by incorporating extra set of estimating equations in the empirical likelihood framework.

To obtain robust estimates, we use median as an estimate of center instead of mean. We propose here using two different kinds of estimating equations; one uses median and the other one uses a smoothed version of median. The smoothing technique is the one proposed in [5] to improve the coverage accuracy. Our method can be easily generalized to multiple samples with relevant information. Without loss of generality, we consider data with two populations in the sequel.

The rest of the paper is organized as follows. The methodology framework, the proposed empirical likelihood approach, and its theoretical properties are presented in Section 2. Results of simulation studies demonstrating the empirical performance are provided in Section 3. Conclusion and some discussion are provided in Section 4.

2. Methodology

Suppose there are two groups of data from different population but sharing the same parameter of interest. Assume that

$$x_1, \dots, x_{n_1} \sim f(x, \theta).$$

The second group of data y_1, \dots, y_{n_2} might be different from the first population, and

$$y_1, \dots, y_{n_2} \sim h(y, \theta).$$

Our goal is to estimate θ by using both samples. Directly using the log-likelihood

$$\sum_{i=1}^{n_1} \log f(x_i, \theta) + \sum_{i=1}^{n_2} \log h(y_i, \theta),$$

we might get a biased estimation due to the difference between the two populations.

We propose a semi-parametric empirical likelihood method which only requires the independence of these two samples. To combine the second sample with the first one, we use the following semi-parametric empirical likelihood

$$\ell = \sum_{i=1}^{n_1} \log f(x_i, \theta) + \sum_{i=1}^{n_2} \log n_2 p_i,$$

where

$$p_i \geq 0, \sum_{i=1}^{n_2} p_i = 1, \sum_{i=1}^{n_2} p_i g(y_i, \theta) = \mathbf{0},$$

and $g(y_i, \theta)$ is an estimating function. From the empirical likelihood theory, we know p_i is maximized by $1/\{n_2 [1 + \lambda^T g(y_i, \theta)]\}$, where λ is the Lagrange multiplier. We can rewrite the log likelihood function as

$$\ell(\theta) = \sum_{i=1}^{n_1} \log f(x_i, \theta) - \sum_{i=1}^{n_2} \log [1 + \lambda^T g(y_i, \theta)], \quad (1)$$

and

$$\hat{\theta} = \arg \max_{\theta} \ell(\theta).$$

We call $\hat{\theta}$ the robust semi-parametric empirical likelihood estimate (RSPELE).

The advantage of the log profile likelihood function is that it does not depend on the likelihood weights which could be difficult to choose. In our propose method, we do not require that the probability density function of the second population is identical to the first population. By using the empirical likelihood method, we do not even need to specify the functional form of the underlying distribution of the second population. Therefore, we can gain robust estimates in the sense that model mis-specification problem is avoided. Consequently, our method can be employed in a relatively wide range of applications when the functional form of the probability density function is not known.

In the following, the theoretical properties of the proposed RSPELE estimator will be presented. For clarity, all proofs are postponed to the Appendix. Theorem 1 below shows that under some regularity conditions, the RSPELE estimator $\hat{\theta}$ is consistent to θ_0 .

Theorem 1 *Let x_1, \dots, x_{n_1} be i.i.d. from $f(x, \theta)$ and y_1, \dots, y_{n_2} be i.i.d. from an unknown distribution. Assume that $f(x, \theta)$ satisfies the regularity conditions given in Shao (2003) on the normality of the maximum likelihood estimator in parametric models. Let θ_0 be the true parameter. We further assume that*

A1) There exists a matrix $\Psi > 0$ such that

$$E[g(y_i, \theta)] = \Psi(\theta - \theta_0) + o(\|\theta - \theta_0\|)$$

as $\theta - \theta_0 \rightarrow \mathbf{0}$.

A2) $\zeta(\theta) = E[\|g(y_i, \theta) - g(y_i, \theta_0)\|^2]$ exists when $\|\theta - \theta_0\|$ is sufficiently small and is continuous at θ_0 .

A3) $E[g(y_i, \theta_0) g^T(y_i, \theta_0)] > 0$.

Then, it follows that $\hat{\theta} \rightarrow \theta_0$ in probability in the neighborhood of θ_0 such that $\{\theta: \|\theta - \theta_0\| \leq n^{-1/2}\}$.

The asymptotic distribution of the $\hat{\theta}$ and $\hat{\lambda}$ is shown in Theorem 2.

Theorem 2 *In addition to the conditions of Theorem 1,*

we assume that $\frac{n_1}{n} \rightarrow b$, where b is a constant. We also

assume that $\frac{\partial g(y, \theta)}{\partial \theta}$ exists with probability one and the set of its discontinuity points has zero probability.

Then

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &\xrightarrow{d} N(\mathbf{0}, S_1), \\ \sqrt{n}\hat{\boldsymbol{\lambda}} &\xrightarrow{d} N(\mathbf{0}, S_2), \end{aligned}$$

where

$$\begin{aligned} S_0 &= bI(\boldsymbol{\theta}_0) + (1-b)E\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^T}\right)^T \Phi^{-1}E\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^T}\right), \\ S_1 &= bS_0^{-1}I(\boldsymbol{\theta}_0)S_0^{-1} \\ &\quad + (1-b)S_0^{-1}E\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^T}\right)^T \Phi^{-1}E\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^T}\right)S_0^{-1}, \\ S_2 &= b\Phi^{-1}E\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^T}\right)S_0^{-1}I(\boldsymbol{\theta}_0)S_0^{-1}E\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^T}\right)^T \Phi^{-1} \quad (2) \\ &\quad + \left[-\Phi^{-1} + (1-b)\mathbf{B}\right]\Phi\left[-\frac{1}{1-b}\Phi^{-1} + \mathbf{B}\right], \\ \Phi &= E\mathbf{g}\mathbf{g}^T, \\ \mathbf{B} &= \Phi^{-1}E\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^T}\right)S_0^{-1}E\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^T}\right)^T \Phi^{-1} \end{aligned}$$

and $I(\boldsymbol{\theta}_0)$ is the Fisher information about $\boldsymbol{\theta}_0$ contained in X .

The asymptotic distribution of $2\left[\ell(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) - \ell(\boldsymbol{\theta}_0, \boldsymbol{\lambda}_0)\right]$ is given in Theorem 3.

Theorem 3 Assume that the assumptions made in Theorem 2 hold. The limiting distribution of $2\ell(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) - 2\ell(\boldsymbol{\theta}_0, \boldsymbol{\lambda}_0)$ is the same as the distribution of

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}^T V^{-1} C V^{-1} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$$

under $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, where U_1 is independent of U_2 ,

$$U_1 \sim N(\mathbf{0}, bI(\boldsymbol{\theta}_0)),$$

$$U_2 \sim N(\mathbf{0}, (1-b)E\left[\mathbf{g}(Y, \boldsymbol{\theta}_0)\mathbf{g}^T(Y, \boldsymbol{\theta}_0)\right]),$$

$$V^{-1} = \begin{pmatrix} S_0^{-1} & S_0^{-1}E\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^T}\right)^T \Phi^{-1} \\ \Phi^{-1}E\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^T}\right)S_0^{-1} & V^{(22)} \end{pmatrix},$$

and

$$C = \begin{pmatrix} S_0 & (1-b)E\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^T}\right)^T \\ (1-b)E\left(\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^T}\right) & 0 \end{pmatrix},$$

where $V^{(22)} = -\frac{1}{1-b}\Phi^{-1} + \mathbf{B}$.

We note that other test statistics, for example, a test statistic based on Theorem 2, may also be used.

Estimating equations provide a very flexible way to specify how the parameters of a statistical model should be estimated. They serve as constraints in maximizing the empirical likelihood. It was shown in [5] that the empirical likelihood method is an efficient tool for point estimation through estimating equations. In this section, we consider two different kinds of estimating equations using the information of median, since median is robust with respect to the outliers, one may use

$$\mathbf{g}_1(\mathbf{y}_i, \boldsymbol{\theta}) = \mathbf{1}(\mathbf{y}_i \leq m(\boldsymbol{\theta})) - 1/2$$

as estimating function based on the second group data, where $m(\boldsymbol{\theta})$ is the median of $f(\mathbf{x}, \boldsymbol{\theta})$ and $\mathbf{1}(\cdot)$ is the usual indicator function. It is easy to verify that $E(\mathbf{g}_1(\mathbf{y}_i, \boldsymbol{\theta})) = 0$.

Due to the discontinuity of \mathbf{g}_1 , we may use the smoothed version of the constraint which was motivated by Shi and Lau (1999). First of all, we define the estimating equation for the smoothed empirical likelihood. In general, let κ be the r^{th} -order kernel (Shi and Lau, 1999), such that

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

where r is a positive integer. Define

$$\psi(u) = \int \kappa(t) I(t < u) dt.$$

For any $h > 0$, let

$\psi_h(u) = \psi(u/h)$ where h is called the smoothing parameter. The kernel κ is a symmetric probability density with bounded and compact support. Let the estimating function $\mathbf{g}_2(\mathbf{y}, \boldsymbol{\theta}, h) = \psi_h(m(\boldsymbol{\theta}) - \mathbf{y}) - 1/2$. Therefore, $\mathbf{g}_2(\mathbf{y}, \boldsymbol{\theta}, h)$ is continuous with respect to \mathbf{y} , but it is not a fixed function as the smoothing parameter varies. See Shi and Lau (1999) for details. In addition, by using the arguments similar to those stated in the Theorems 1-3 and Shi and Lau (1999), we may get the similar asymptotic results.

3. Numerical Experiments

3.1. Data Fusion with Conventional Empirical Likelihood

Simulation studies are carried out by performing data fusion when two samples are available. The first sample $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_{n_1})$ is generated from standard normal distribution and the second sample $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_{n_2})$ is generated from normal, double exponential, or t -distribution respectively. The sample size of first sample, n_1 , is 10 and for the second sample, the sample size, n_2 , varies from 10, 20 to 30.

First of all we use the median constraint, so the log

likelihood function of the simulation model is

$$l(\mu) = -\sum_{i=1}^{n_1} \left\{ \log \sqrt{2\pi s_1^2} + \frac{(x_i - \mu)^2}{2s_1^2} \right\} - \sum_{i=1}^{n_2} \log \{1 + \lambda [I(y_i \leq \mu) - 1/2]\},$$

where s_1^2 is the MLE variance, $s_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2$.

We present the mean square error (MSE) ratio of RSPELE to MLE based on 1000 replications in **Table 1**. The simulation results show that RSPELE performs well except in the situation when the second population is normally distributed with large variation as the first one. When the second sample size is increasing, the RSPELE becomes more accurate. Moreover, we have smaller MSE of RSPELE when the data of the second population is more concentrated around the center, for example, the double exponential distribution.

3.2. Smoothed Empirical Likelihood

In this section we demonstrate the smoothed version of the estimating equations. The kernel we chose is the

same as the one used in [6],

$$\kappa(u) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{1}{5}u^2\right), & \text{if } \|u\| \leq \sqrt{5}, \\ 0, & \text{otherwise} \end{cases}$$

The simulation model is identical to the first experiment. Four values of the smoothed parameter are used which are n_2 to the power of $-1, -3/4, -1/2$ and $-1/4$. The log likelihood function of the simulation model is

$$l(\mu) = -\sum_{i=1}^{n_1} \left\{ \log \sqrt{2\pi s_1^2} + \frac{(x_i - \mu)^2}{2s_1^2} \right\} - \sum_{i=1}^{n_2} \log \{1 + \lambda g_2(y_i, \theta, h)\},$$

where $s_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2$ is the MLE variance.

We provide the MSE ratio of RSPELE to MLE based on 1000 replications in **Table 2**. Results of the smoothed version are slightly better than the results of the median version, no matter which smoothing parameter is chosen. When the underlying distribution of the second population is not the same as the first population, the RSPELE

Table 1. MSE ratio of RSPELE to MLE based on 1000 replications.

Distribution of sample 2										
n_2	$N(0,1)$	$N(0,1.25)$	$N(0,1.5)$	$N(0,2)$	$N(0,3)$	t_3	t_5	$DE(0,0.5)$	$DE(0,1)$	$DE(0,1.5)$
10	0.776	0.857	0.929	1.017	1.028	0.787	0.747	0.431	0.717	0.875
20	0.569	0.693	0.769	0.871	0.956	0.621	0.580	0.209	0.515	0.705
30	0.453	0.559	0.661	0.789	0.926	0.462	0.467	0.137	0.387	0.582

Table 2. MSE Ratio of RSPELE to MLE with the smoothing parameter $h = n_2$ to the power of $-1, -3/4, -1/2$ and $-1/4$.

Distribution of sample 2										
n_2	$N(0,1)$	$N(0,1.25)$	$N(0,1.5)$	$N(0,2)$	$N(0,3)$	t_3	t_5	$DE(0,0.5)$	$DE(0,1)$	$DE(0,1.5)$
$h = n_2^{-1}$										
10	0.739	0.841	0.922	1.019	1.039	0.747	0.719	0.364	0.664	0.847
20	0.539	0.658	0.759	0.863	0.955	0.609	0.560	0.187	0.487	0.697
30	0.432	0.555	0.649	0.770	0.922	0.442	0.437	0.120	0.362	0.572
$h = n_2^{-3/4}$										
10	0.700	0.810	0.878	0.987	1.036	0.730	0.692	0.339	0.645	0.821
20	0.503	0.624	0.734	0.853	0.941	0.588	0.512	0.172	0.466	0.662
30	0.405	0.528	0.624	0.765	0.892	0.418	0.413	0.107	0.333	0.552
$h = n_2^{-1/2}$										
10	0.666	0.780	0.856	0.959	1.038	0.702	0.665	0.326	0.625	0.795
20	0.453	0.579	0.674	0.802	0.916	0.543	0.475	0.162	0.433	0.616
30	0.356	0.476	0.580	0.726	0.855	0.385	0.369	0.101	0.319	0.524
$h = n_2^{-1/4}$										
10	0.616	0.752	0.842	0.943	1.036	0.664	0.630	0.329	0.630	0.779
20	0.403	0.532	0.635	0.773	0.899	0.504	0.438	0.174	0.419	0.600
30	0.320	0.434	0.532	0.682	0.838	0.356	0.338	0.111	0.315	0.506

estimate performs better than the MLE. When the sample size of the second population is increasing, the RSPELE estimate is more accurate.

3.3. Confidence Intervals

In this subsection, we construct the confidence interval for the median by bootstrapping. In this simulation study, the first sample X is generated from standard normal distribution and the second sample Y is generated from normal, double exponential, or t -distribution. The sample size of X is 10 and of Y varies from 10, 20, to 30. The size of the bootstrapped sample is 200 and the number of iterations is set to be 1000. First of all, we use the median estimating equation and record the coverage probabilities and the simulated average confidence interval lengths (AL) in **Table 3** for nominal levels of 80, 90, 95, and 99 percent. The coverage probabilities and the AL of using the smoothed version of the estimating equation

are recorded in **Tables 4** and **5** with different smoothing parameters which are n_2 to the power of -1 and $-1/2$. We report the results of MLE in **Table 6**. Since the results of MLE do not depend on the second population, we further compare the coverage probabilities and AL as in **Tables 3-5** with **Table 6**.

The results of smoothed version are better than median version in terms of the coverage probabilities. The coverage probabilities of RSPELE and MLE are very close but the confidence intervals of RSPELE are about 10% narrower than of MLE. The results of RSPELE when the underlying distribution of the second population is either t or double exponential distribution are better than the results of RSPELE when underlying distribution is normal distribution. That is because normal distribution is flatter than t and double exponential. Consequently, if the second population provides a good information about the center we can use it to get better estimates.

Table 3. Coverage probability of RSPELE with different distributions of second population by using median estimating equation, numbers show in the brackets are AL, $\alpha =$ nominal level.

α	0.800	0.900	0.950	0.990	0.800	0.900	0.950	0.990	0.800	0.900	0.950	0.990
n_2	$N(0,1)$				$N(0,2)$				$DE(0,0.5)$			
10	0.754 (0.725)	0.846 (0.930)	0.902 (1.102)	0.963 (1.435)	0.746 (0.744)	0.839 (0.949)	0.898 (1.126)	0.960 (1.471)	0.759 (0.716)	0.854 (0.919)	0.909 (1.094)	0.951 (1.441)
20	0.771 (0.708)	0.868 (0.910)	0.913 (1.085)	0.964 (1.415)	0.756 (0.754)	0.855 (0.961)	0.908 (1.138)	0.958 (1.482)	0.780 (0.678)	0.872 (0.878)	0.918 (1.055)	0.966 (1.401)
30	0.805 (0.679)	0.888 (0.874)	0.935 (1.047)	0.976 (1.389)	0.791 (0.754)	0.873 (0.957)	0.925 (1.135)	0.970 (1.483)	0.780 (0.643)	0.883 (0.843)	0.920 (1.018)	0.971 (1.368)
	$N(0,1.5)$				t_3				$DE(0,1)$			
10	0.748 (0.735)	0.845 (0.945)	0.899 (1.116)	0.962 (1.457)	0.758 (0.733)	0.848 (0.938)	0.908 (1.112)	0.964 (1.460)	0.758 (0.736)	0.848 (0.941)	0.907 (1.119)	0.951 (1.460)
20	0.761 (0.731)	0.864 (0.938)	0.911 (1.111)	0.961 (1.447)	0.766 (0.711)	0.851 (0.911)	0.892 (1.080)	0.951 (1.411)	0.770 (0.711)	0.859 (0.911)	0.912 (1.086)	0.962 (1.425)
30	0.799 (0.715)	0.882 (0.912)	0.936 (1.084)	0.973 (1.422)	0.771 (0.700)	0.869 (0.897)	0.916 (1.068)	0.966 (1.402)	0.772 (0.685)	0.866 (0.882)	0.915 (1.053)	0.965 (1.392)
	$N(0,2)$				t_5				$DE(0,2)$			
10	0.748 (0.741)	0.842 (0.948)	0.899 (1.120)	0.960 (1.464)	0.734 (0.744)	0.840 (0.949)	0.897 (1.124)	0.963 (1.467)	0.755 (0.743)	0.847 (0.952)	0.904 (1.132)	0.951 (1.470)
20	0.753 (0.743)	0.858 (0.951)	0.909 (1.127)	0.958 (1.462)	0.761 (0.716)	0.858 (0.916)	0.906 (1.096)	0.974 (1.441)	0.762 (0.731)	0.859 (0.935)	0.908 (1.109)	0.963 (1.445)
30	0.798 (0.736)	0.879 (0.936)	0.931 (1.111)	0.972 (1.456)	0.766 (0.693)	0.877 (0.892)	0.929 (1.065)	0.971 (1.403)	0.770 (0.712)	0.866 (0.913)	0.906 (1.086)	0.961 (1.416)

Table 4. Coverage probability of RSPELE with different distributions of second population, and the smoothing parameter $h = n_2^{-1}$, numbers show in the brackets are AL, $\alpha =$ nominal level.

α	0.800	0.900	0.950	0.990	0.800	0.900	0.950	0.990	0.800	0.900	0.950	0.990
n_2	$N(0,1)$				$N(0,2)$				$DE(0,0.5)$			
10	0.763 (0.687)	0.861 (0.886)	0.920 (1.058)	0.973 (1.394)	0.752 (0.754)	0.840 (0.963)	0.902 (1.139)	0.966 (1.483)	0.787 (0.626)	0.889 (0.829)	0.934 (1.009)	0.979 (1.368)
20	0.797 (0.649)	0.875 (0.840)	0.915 (1.006)	0.968 (1.344)	0.765 (0.756)	0.863 (0.963)	0.907 (1.138)	0.956 (1.477)	0.813 (0.568)	0.903 (0.766)	0.941 (0.945)	0.979 (1.311)
30	0.816 (0.610)	0.912 (0.796)	0.941 (0.965)	0.980 (1.305)	0.802 (0.751)	0.885 (0.956)	0.926 (1.130)	0.968 (1.478)	0.821 (0.523)	0.901 (0.718)	0.946 (0.892)	0.979 (1.263)
	$N(0,1.5)$				t_3				$DE(0,1)$			
10	0.756 (0.726)	0.855 (0.930)	0.914 (1.104)	0.972 (1.448)	0.782 (0.709)	0.873 (0.910)	0.923 (1.082)	0.968 (1.441)	0.777 (0.703)	0.869 (0.909)	0.920 (1.084)	0.965 (1.438)
20	0.786 (0.707)	0.873 (0.908)	0.913 (1.077)	0.966 (1.415)	0.774 (0.668)	0.869 (0.861)	0.911 (1.028)	0.958 (1.361)	0.794 (0.660)	0.875 (0.856)	0.927 (1.029)	0.970 (1.370)
30	0.803 (0.684)	0.899 (0.874)	0.935 (1.041)	0.974 (1.377)	0.793 (0.648)	0.880 (0.836)	0.924 (1.001)	0.971 (1.341)	0.785 (0.622)	0.878 (0.811)	0.926 (0.981)	0.971 (1.320)
	$N(0,2)$				t_5				$DE(0,2)$			
10	0.752 (0.743)	0.846 (0.948)	0.910 (1.123)	0.970 (1.473)	0.765 (0.716)	0.865 (0.914)	0.911 (1.091)	0.964 (1.436)	0.774 (0.732)	0.862 (0.943)	0.915 (1.121)	0.958 (1.467)
20	0.771 (0.732)	0.871 (0.936)	0.909 (1.112)	0.961 (1.453)	0.785 (0.666)	0.881 (0.855)	0.922 (1.028)	0.976 (1.376)	0.780 (0.707)	0.868 (0.907)	0.919 (1.078)	0.966 (1.417)
30	0.798 (0.720)	0.893 (0.917)	0.934 (1.087)	0.973 (1.433)	0.799 (0.637)	0.893 (0.822)	0.939 (0.990)	0.974 (1.334)	0.780 (0.677)	0.874 (0.873)	0.917 (1.043)	0.965 (1.381)

Table 5. Coverage probability of RSPELE with different distributions of second population, and the smoothing parameter $h = n_2^{-1}$, numbers show in the brackets are AL, $\alpha =$ nominal level.

α	0.800	0.900	0.950	0.990	0.800	0.900	0.950	0.990	0.800	0.900	0.950	0.990
n_2	$N(0,1)$				$N(0,2)$				$DE(0,0.5)$			
10	0.770 (0.630)	0.855 (0.814)	0.918 (0.977)	0.974 (1.310)	0.750 (0.756)	0.846 (0.965)	0.908 (1.143)	0.970 (1.491)	0.773 (0.508)	0.877 (0.685)	0.933 (0.861)	0.980 (1.240)
20	0.777 (0.551)	0.870 (0.713)	0.921 (0.862)	0.973 (1.176)	0.773 (0.745)	0.865 (0.953)	0.902 (1.127)	0.954 (1.459)	0.797 (0.391)	0.899 (0.531)	0.945 (0.676)	0.982 (1.051)
30	0.805 (0.491)	0.904 (0.636)	0.936 (0.770)	0.980 (1.079)	0.800 (0.732)	0.887 (0.933)	0.933 (1.106)	0.972 (1.442)	0.795 (0.313)	0.893 (0.431)	0.943 (0.568)	0.987 (0.910)
	$N(0,1.5)$				t_3				$DE(0,1)$			
10	0.763 (0.702)	0.849 (0.903)	0.911 (1.074)	0.970 (1.416)	0.782 (0.662)	0.874 (0.853)	0.923 (1.021)	0.966 (1.375)	0.758 (0.649)	0.876 (0.845)	0.925 (1.019)	0.969 (1.377)
20	0.780 (0.657)	0.864 (0.844)	0.911 (1.003)	0.964 (1.327)	0.782 (0.589)	0.862 (0.763)	0.916 (0.918)	0.967 (1.241)	0.791 (0.566)	0.889 (0.740)	0.923 (0.897)	0.975 (1.235)
30	0.816 (0.611)	0.900 (0.787)	0.934 (0.939)	0.976 (1.258)	0.797 (0.538)	0.881 (0.698)	0.928 (0.845)	0.978 (1.152)	0.777 (0.493)	0.885 (0.649)	0.928 (0.795)	0.978 (1.123)
	$N(0,2)$				t_5				$DE(0,2)$			
10	0.753 (0.733)	0.852 (0.940)	0.905 (1.113)	0.970 (1.461)	0.766 (0.662)	0.862 (0.854)	0.920 (1.021)	0.967 (1.365)	0.755 (0.708)	0.864 (0.910)	0.908 (1.087)	0.967 (1.446)
20	0.774 (0.706)	0.865 (0.906)	0.906 (1.073)	0.959 (1.402)	0.786 (0.578)	0.880 (0.747)	0.936 (0.900)	0.972 (1.226)	0.786 (0.649)	0.866 (0.836)	0.922 (1.002)	0.963 (1.342)
30	0.809 (0.677)	0.899 (0.862)	0.935 (1.026)	0.973 (1.359)	0.803 (0.520)	0.901 (0.679)	0.940 (0.823)	0.975 (1.137)	0.785 (0.596)	0.875 (0.774)	0.919 (0.934)	0.968 (1.264)

Table 6. Coverage probability of MLE with the distribution of second population normal, t_3 , t_5 and double exponential, numbers show in the brackets are AL, $\alpha =$ nominal level.

α	0.800	0.900	0.950	0.990	0.800	0.900	0.950	0.990
n_2	Normal				t_3			
10	0.774 (0.774)	0.864 (0.994)	0.913 (1.184)	0.972 (1.557)	0.774 (0.777)	0.854 (0.998)	0.915 (1.189)	0.973 (1.563)
20	0.773 (0.784)	0.871 (1.007)	0.913 (1.200)	0.976 (1.577)	0.769 (0.777)	0.849 (0.997)	0.901 (1.188)	0.964 (1.562)
30	0.810 (0.788)	0.888 (1.011)	0.932 (1.204)	0.983 (1.583)	0.760 (0.786)	0.858 (1.009)	0.912 (1.203)	0.972 (1.581)
	t_5				Double Exponential			
10	0.737 (0.789)	0.858 (1.013)	0.913 (1.207)	0.966 (1.586)	0.771 (0.781)	0.863 (1.002)	0.913 (1.194)	0.962 (1.569)
20	0.759 (0.795)	0.862 (1.020)	0.915 (1.216)	0.975 (1.598)	0.771 (0.781)	0.863 (1.002)	0.913 (1.194)	0.962 (1.569)
30	0.769 (0.790)	0.880 (1.015)	0.928 (1.209)	0.977 (1.589)	0.771 (0.781)	0.863 (1.002)	0.913 (1.194)	0.962 (1.569)

4. Discussion

In this paper, we propose a robust semi-parametric empirical likelihood in a multiple-sample model with common measurement of center. We use two different kind of estimating equations of information about the median. Simulation studies have shown that the second population could provide very useful information on the parameter of interest by comparing the performance of various commonly used measures for evaluations.

5. Acknowledgements

We would like to thank Dr. Jing Qin for his generous and valuable comments and suggestions for this article.

REFERENCES

[1] X. Wang, C. van Eeden and J. V. Zidek, "Asymptotic Properties of Maximum Weighted Likelihood Estimators," *Journal of Statistical Planning and Inference*, Vol. 119, No. 1, 2004, pp. 37-54.

[2] Y. Fu, X. Wang and C. Wu, "Weighted Empirical Likelihood Inference for Multiple Samples," *Journal of Statistical Planning and Inference*, Vol. 139, No. 4, 2009, pp. 1462-1473.

[3] A. B. Owen, "Empirical Likelihood Ratio Confidence Interval for a Single Functional," *Biometrika*, Vol. 75, No. 2, 1988, pp. 237-249.

[4] J. Qin and J. Lawless, "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, Vol. 22, No. 1, 1994, pp. 300-325.

[5] J. Shi and T. S. Lau, "Robust Empirical Likelihood for Linear Models under Median Constraints," *Communications in Statistics*, Vol. 28, No. 10, 1999, pp. 2465-2479.

[6] J. Shao, "Mathematical Statistics," Springer, New York, 2003. [doi:10.1007/b97553](https://doi.org/10.1007/b97553)

[7] A. B. Owen, "Empirical Likelihood," Chapman & Hall/CRC, New York, 2001. [doi:10.1201/9781420036152](https://doi.org/10.1201/9781420036152)

[8] Z. D. Bai, C. R. Rao and Y. Wu, "M-Estimation of Multivariate Linear Regression Parameters under a Convex Discrepancy Function," *Statistica Sinica*, Vol. 2, 1992, pp. 237-254.

Appendix: Proofs

Proof of Theorem 1

We rewrite the Equation (1) as $\ell(\boldsymbol{\theta}) = \ell_1(\boldsymbol{\theta}) + \ell_2(\boldsymbol{\theta})$,

where $\ell_1(\boldsymbol{\theta}) = \sum_{i=1}^{n_1} \log f(\mathbf{x}_i, \boldsymbol{\theta})$ and

$\ell_2(\boldsymbol{\theta}) = -\sum_{i=1}^{n_2} \log[1 + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta})]$. We denote $N(\boldsymbol{\theta}_0)$

be the neighborhood of $\boldsymbol{\theta}_0$ such that

$\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq n^{-1/2}\}$, $\partial N(\boldsymbol{\theta}_0)$ be the boundary of

$N(\boldsymbol{\theta}_0)$, i.e. for all $\boldsymbol{\theta}_*$ such that

$\|\boldsymbol{\theta}_* - \boldsymbol{\theta}_0\| = n^{-1/2}$, $N_{\text{int}}(\boldsymbol{\theta}_0)$ be the neighborhood of $\boldsymbol{\theta}_0$

such that $\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < n^{-1/2}\}$, and $n = n_1 + n_2$.

Case 1. $0 < b_1 < \frac{n_1}{n_2} < b_2 < \infty$, where b_1 and b_2 are

two constants.

In view of the proof of Theorem 4.17 of [6], it follows that for any $\varepsilon > 0$,

$$P[\ell_1(\boldsymbol{\theta}_*) - \ell_1(\boldsymbol{\theta}_0) < 0 \text{ for all } \boldsymbol{\theta}_* \in \partial N(\boldsymbol{\theta}_0)] \geq 1 - \varepsilon, \tag{3}$$

for large n_1 .

By Assumptions (A1)-(A3), applying a similar approach as in [7], it can be shown that $\hat{\boldsymbol{\lambda}} = O(n^{1/2})$. In light of the arguments in [4,8], and under the assumptions (A1)-(A3), it follows that for any $\varepsilon > 0$,

$$P[\ell_2(\boldsymbol{\theta}_*) - \ell_2(\boldsymbol{\theta}_0) < 0 \text{ for all } \boldsymbol{\theta}_* \in \partial N(\boldsymbol{\theta}_0)] \geq 1 - \varepsilon, \tag{4}$$

for large n_2 .

Combining (3) with (4), we have for any $\varepsilon > 0$,

$$P[\ell(\boldsymbol{\theta}_*) - \ell(\boldsymbol{\theta}_0) < 0 \text{ for all } \boldsymbol{\theta}_* \in \partial N(\boldsymbol{\theta}_0)] \geq 1 - \varepsilon,$$

for large n . Therefore, there exists $\hat{\boldsymbol{\theta}} \in N_{\text{int}}(\boldsymbol{\theta}_0)$ such that

$$\frac{\partial \ell(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \mathbf{0} \text{ and } \hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in N_{\text{int}}(\boldsymbol{\theta}_0)} \ell(\hat{\boldsymbol{\theta}}).$$

By the definition of $N(\boldsymbol{\theta}_0)$, it follows that $\hat{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}_0$ in probability.

$$\text{Case 2. } \frac{n_1}{n_2} \rightarrow \infty \text{ or } \frac{n_1}{n_2} \rightarrow 0.$$

The consistency of $\hat{\boldsymbol{\theta}}$ can be shown similarly. The details are omitted. \square

Proof of Theorem 2

We denote

$$w_{n_1}(\boldsymbol{\theta}) = \frac{\partial \ell_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n_1} \frac{\partial \log f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}};$$

$$q_{1n_2}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \frac{\partial \ell_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\sum_{i=1}^{n_2} \frac{1}{1 + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta})} \frac{\partial \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \boldsymbol{\lambda};$$

$$q_{2n_2}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \sum_{i=1}^{n_2} \frac{1}{1 + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta})} \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}).$$

Since

$$\frac{\partial \ell(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \mathbf{0} \text{ and } q_{2n_2}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) = \mathbf{0},$$

by applying Taylor's expansion, it follows that

$$\mathbf{0} = w_{n_1}(\hat{\boldsymbol{\theta}}) + q_{1n_2}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) = w_{n_1}(\boldsymbol{\theta}_0) + \frac{\partial w_{n_1}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \frac{\partial q_{1n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}^T} \hat{\boldsymbol{\lambda}} + o_p(\delta_n),$$

and

$$\mathbf{0} = q_{2n_2}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) = q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) + \frac{\partial q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\theta}^T} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \frac{\partial q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}^T} \hat{\boldsymbol{\lambda}} + o_p(\delta_n),$$

where $\delta_n = \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| + \|\hat{\boldsymbol{\lambda}}\|$. It is noted that

$$\frac{\partial w_{n_1}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T} = \sum_{i=1}^{n_1} \frac{\partial^2 \log f(\mathbf{x}_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T};$$

$$\frac{\partial q_{1n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}^T} = -\sum_{i=1}^{n_2} \frac{\partial \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T};$$

$$\frac{\partial q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\theta}^T} = \sum_{i=1}^{n_2} \frac{\partial \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T};$$

$$\frac{\partial q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}^T} = -\sum_{i=1}^{n_2} \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0) \mathbf{g}^T(\mathbf{y}_i, \boldsymbol{\theta}_0).$$

Hence we have

$$\begin{pmatrix} -\frac{\partial w_{n_1}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T} & -\frac{\partial q_{1n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}^T} \\ \frac{\partial q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\theta}^T} & \frac{\partial q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}^T} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \hat{\boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} w_{n_1}(\boldsymbol{\theta}_0) + o_p(\delta_n) \\ -q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) + o_p(\delta_n) \end{pmatrix},$$

which can be written as

$$\begin{pmatrix} \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \sqrt{n}\hat{\boldsymbol{\lambda}} \end{pmatrix} = V_n^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}} w_{n_1}(\boldsymbol{\theta}_0) + o_p\left(\frac{\delta_n}{\sqrt{n}}\right) \\ -\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) + o_p\left(\frac{\delta_n}{\sqrt{n}}\right) \end{pmatrix},$$

where

$$V_n = \begin{pmatrix} \frac{1}{n} \frac{\partial w_{n_1}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T} & \frac{1}{n} \frac{\partial q_{1n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}^T} \\ \frac{1}{n} \frac{\partial q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\theta}^T} & \frac{1}{n} \frac{\partial q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}^T} \end{pmatrix}.$$

It can be shown that

$$\begin{aligned} \delta_n &= O_p(n^{-1/2}); \\ n^{-1/2} w_{n_1}(\boldsymbol{\theta}_0) &\xrightarrow{d} N(\mathbf{0}, bI(\boldsymbol{\theta}_0)); \\ -n^{-1/2} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) &\xrightarrow{d} N(\mathbf{0}, (1-b)E[\mathbf{g}(\mathbf{Y}, \boldsymbol{\theta}_0)\mathbf{g}^T(\mathbf{Y}, \boldsymbol{\theta}_0)]); \\ \frac{1}{n} \frac{\partial w_{n_1}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T} &\xrightarrow{a.s.} -bI(\boldsymbol{\theta}_0); \\ \frac{1}{n} \frac{\partial q_{1n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}^T} &\xrightarrow{a.s.} -(1-b)E\left[\frac{\partial \mathbf{g}(\mathbf{Y}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T}\right]; \\ \frac{1}{n} \frac{\partial q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\theta}^T} &\xrightarrow{a.s.} (1-b)E\left[\frac{\partial \mathbf{g}(\mathbf{Y}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T}\right]; \\ \frac{1}{n} \frac{\partial q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}^T} &\xrightarrow{a.s.} -(1-b)E(\mathbf{g}(\mathbf{Y}, \boldsymbol{\theta}_0)\mathbf{g}^T(\mathbf{Y}, \boldsymbol{\theta}_0)). \end{aligned}$$

Therefore, it follows that

$$\begin{pmatrix} \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \sqrt{n}\hat{\boldsymbol{\lambda}} \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, W),$$

where W is the covariance matrix of

$$V^{-1} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$$

with

$$\begin{aligned} V &= \begin{pmatrix} bI(\boldsymbol{\theta}_0) & (1-b)E\left[\frac{\partial \mathbf{g}(\mathbf{Y}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T}\right]^T \\ (1-b)E\left[\frac{\partial \mathbf{g}(\mathbf{Y}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T}\right] & -(1-b)E[\mathbf{g}(\mathbf{Y}, \boldsymbol{\theta}_0)\mathbf{g}^T(\mathbf{Y}, \boldsymbol{\theta}_0)] \end{pmatrix} \\ &= \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}, \end{aligned}$$

U_1 is independent of U_2 and

$$\begin{aligned} U_1 &\sim N(\mathbf{0}, bI(\boldsymbol{\theta}_0)), \\ U_2 &\sim N(\mathbf{0}, (1-b)E[\mathbf{g}(\mathbf{Y}, \boldsymbol{\theta}_0)\mathbf{g}^T(\mathbf{Y}, \boldsymbol{\theta}_0)]). \end{aligned}$$

Since

$$\begin{aligned} V^{-1} &= \begin{pmatrix} V_{11}^{-1} + V_{11}^{-1}V_{12}V_{22}^{-1}V_{21}V_{11}^{-1} & -V_{22}^{-1}V_{21}V_{11}^{-1} \\ -V_{11}^{-1}V_{12}V_{22}^{-1} & V_{22}^{-1} \end{pmatrix} \\ &= \begin{pmatrix} V_{11.2}^{-1} & -V_{11.2}^{-1}V_{12}V_{22}^{-1} \\ -V_{22}^{-1}V_{21}V_{11.2}^{-1} & V_{22}^{-1} + V_{22}^{-1}V_{21}V_{11.2}^{-1}V_{12}V_{22}^{-1} \end{pmatrix} = \begin{pmatrix} V^{(11)} & V^{(12)} \\ V^{(21)} & V^{(22)} \end{pmatrix} \end{aligned}$$

with $V_{11.2} = V_{11} - V_{12}V_{22}^{-1}V_{21}$ and $V_{22.1} = V_{22} - V_{21}V_{11}^{-1}V_{12}$, it follows that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, S_1), \sqrt{n}\hat{\boldsymbol{\lambda}} \xrightarrow{d} N(\mathbf{0}, S_2),$$

where

$$\begin{aligned} S_1 &= V^{(11)} \text{cov}(U_1)(V^{(11)})^T + V^{(12)} \text{cov}(U_2)(V^{(12)})^T, \\ S_2 &= V^{(21)} \text{cov}(U_1)(V^{(21)})^T + V^{(22)} \text{cov}(U_2)(V^{(22)})^T. \end{aligned} \quad \square$$

Proof of Theorem 3

Assume that the assumptions made in Theorem 3 hold. A statistic for testing $H_0: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ is given by

$$\begin{aligned} &2\left[\ell(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) - \ell(\boldsymbol{\theta}_0, \boldsymbol{\lambda}_0)\right] \\ &= 2\left\{\sum_{i=1}^{n_1} \log f(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) - \sum_{i=1}^{n_2} \log [1 + \hat{\boldsymbol{\lambda}}^T \mathbf{g}(\mathbf{y}_i, \hat{\boldsymbol{\theta}})]\right\} \\ &\quad - 2\left\{\sum_{i=1}^{n_2} \log f(\mathbf{x}_i, \boldsymbol{\theta}_0) - \sum_{i=1}^{n_2} \log [1 + \boldsymbol{\lambda}_0^T \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0)]\right\}. \end{aligned} \quad (5)$$

Expanding $\ell(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})$ around $(\boldsymbol{\theta}_0, \mathbf{0})$ by Taylor's expansion, we have

$$\begin{aligned} \ell(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) &= \sum_{i=1}^{n_1} \log f(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) - \sum_{i=1}^{n_2} \log [1 + \hat{\boldsymbol{\lambda}}^T \mathbf{g}(\mathbf{y}_i, \hat{\boldsymbol{\theta}})] \\ &= \sum_{i=1}^{n_1} \log f(\mathbf{x}_i, \boldsymbol{\theta}_0) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T w_{n_1}(\boldsymbol{\theta}_0) \\ &\quad + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \frac{\partial w_{n_1}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^T} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &\quad - \hat{\boldsymbol{\lambda}}^T q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) - \frac{1}{2} \hat{\boldsymbol{\lambda}}^T \frac{\partial q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}^T} \hat{\boldsymbol{\lambda}} + o_p(1) \\ &= \sum_{i=1}^{n_1} \log f(\mathbf{x}_i, \boldsymbol{\theta}_0) \\ &\quad + \begin{pmatrix} \frac{1}{\sqrt{n}} w_{n_1}(\boldsymbol{\theta}_0) \\ -\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) \end{pmatrix}^T V^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}} w_{n_1}(\boldsymbol{\theta}_0) \\ -\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) \end{pmatrix} \\ &\quad - \frac{1}{2} \begin{pmatrix} \frac{1}{\sqrt{n}} w_{n_1}(\boldsymbol{\theta}_0) \\ -\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) \end{pmatrix}^T V^{-1} \text{diag}(V_{11}, V_{22}) V^{-1} \\ &\quad \times \begin{pmatrix} \frac{1}{\sqrt{n}} w_{n_1}(\boldsymbol{\theta}_0) \\ -\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) \end{pmatrix} + o_p(1). \end{aligned} \quad (6)$$

Under H_0 , we have

$$\frac{1}{n_2} \sum_{i=1}^{n_2} \frac{\mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0)}{[1 + \boldsymbol{\lambda}_0^T \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0)]} = \mathbf{0},$$

which implies that

$$\begin{aligned} \boldsymbol{\lambda}_0 &= \left[\frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0) \mathbf{g}^T(\mathbf{y}_i, \boldsymbol{\theta}_0) \right]^{-1} \left[\frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0) \right] \\ &+ o_p(1) = -(1-b)V_{22}^{-1} \frac{1}{n_2} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) + o_p(1). \end{aligned}$$

Hence,

$$\begin{aligned} &\sum_{i=1}^{n_2} \log[1 + \boldsymbol{\lambda}_0^T \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0)] \\ &= \boldsymbol{\lambda}_0^T \sum_{i=1}^{n_2} \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0) - \frac{1}{2} \boldsymbol{\lambda}_0^T \sum_{i=1}^{n_2} \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0) \mathbf{g}^T(\mathbf{y}_i, \boldsymbol{\theta}_0) \boldsymbol{\lambda}_0 + o_p(1) \\ &= -\frac{n_2(1-b)}{2} \left[\frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0) \right]^T \times V_{22}^{-1} \left[\frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{g}(\mathbf{y}_i, \boldsymbol{\theta}_0) \right] \\ &+ o_p(1) \\ &= \frac{1}{2} \left[\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) \right]^T [-V_{22}]^{-1} \left[\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) \right] + o_p(1) \end{aligned} \tag{7}$$

Combining Equations (5)-(7), we have

$$\begin{aligned} &2\ell(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) - 2\ell(\boldsymbol{\theta}_0, \boldsymbol{\lambda}_0) \\ &= 2 \begin{pmatrix} \frac{1}{\sqrt{n}} w_{n_1}(\boldsymbol{\theta}_0) \\ -\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) \end{pmatrix}^T V^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}} w_{n_1}(\boldsymbol{\theta}_0) \\ -\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) \end{pmatrix} \\ &- \begin{pmatrix} \frac{1}{\sqrt{n}} w_{n_1}(\boldsymbol{\theta}_0) \\ -\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) \end{pmatrix}^T V^{-1} \\ &\text{diag}(V_{11}, V_{22}) V^{-1} \times \begin{pmatrix} \frac{1}{\sqrt{n}} w_{n_1}(\boldsymbol{\theta}_0) \\ -\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) \end{pmatrix} \\ &+ \left[\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) \right]^T V_{22}^{-1} \left[\frac{1}{\sqrt{n}} q_{2n_2}(\boldsymbol{\theta}_0, \mathbf{0}) \right] + o_p(1). \end{aligned}$$

Hence the limiting distribution of $2[\ell(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}}) - \ell(\boldsymbol{\theta}_0, \boldsymbol{\lambda}_0)]$ is the same as the distribution of $\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}^T V^{-1} \begin{pmatrix} V_{11.2} & V_{12} \\ V_{21} & \mathbf{0} \end{pmatrix} V^{-1} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$ under H_0 .