An Extension to Empirical Likelihood for Evaluating Probability Weighted Moments

Albert Vexler, Li Zou, Alan D. Hutson

Department of Biostatistics, The State University of New York at Buffalo, Buffalo, NY 14214

ABSTRACT

The scientific literature has addressed widely the theoretical and applied framework based on probability weighted moments (PWMs). PWMs generalize the concept of conventional moments of a probability function. These methods are commonly applied for modeling extremes of natural phenomena. We propose and examine empirical likelihood (EL) inference methods for PWMs. This approach extends the classical EL technique for evaluating usual moments, including the population mean. We provide an asymptotic proposition, extending a well-known nonparametric version of Wilks theorem used to evaluate the Type I error rates of EL ratio tests. This result is applied in order to develop a powerful nonparametric EL ratio test and the corresponding distribution-free confidence interval (CI) estimation of the PWMs. We show that the proposed method can be easily applied towards inference of the Gini index, a widely used measure for assessing distributional inequality. An extensive Monte Carlo (MC) study shows that the proposed technique provides a well-controlled Type I error rate, as well as very accurate CI estimation, that outperforms the CI estimation based on the classical schemes to analyze the PWMs. These results are clearly observed in the cases when underlying data are skewed and/or consist of a relatively small number of data points. A real data example of myocardial infarction disease is used to illustrate the applicability of the proposed method.

Keywords: Confidence interval estimation; Empirical likelihood; Nonparametric testing; Probability weighted moments; Wilks theorem.

*Corresponding author.
E-mail addresses: avexler@buffalo.edu (A. Vexler*), lizou@buffalo.edu (L. Zou), ahutson@buffalo.edu (A. D. Hutson)

1. Introduction

Greenwood et al. (1979) introduced probability weighted moments (PWMs) as a generalization of the conventional moments of a probability distribution. PWMs are widely used for modeling extremes of natural phenomena. For example, applications of statistical methods based on PWMs are of considerable importance in hydrology, since the concept of the PWMs is very efficient for estimating statistical characteristics of the tails of distribution functions pertaining to features such as 50, 100 and 1,000 year floods. Commonly, PWM based techniques provide favorable estimation properties when using samples with relatively small sizes and are computationally straightforward to calculate (Hosking et al. 1985b; Katz et al. 2002). Researchers have proposed using PWM based methods to quantify the uncertainty related to annual maxima of daily stream flows of rivers (Flood Studies Report 1975; Hosking et al. 1985a; Wallis and Wood 1985).

Hosking et al. (1985b) studied PWMs in the form of $\beta_r = E\left[X\left(F\left(X\right)\right)^r\right]$, the expectation of $X\left(F\left(X\right)\right)^r$,where X is random variable with distribution function F and r is an integer. The authors applied the method of PWMs to estimate characteristics of the generalized extreme value distribution, which is related to the limiting distribution of the maximum of a series of independent and identically distributed (i.i.d) observations. Several important properties of data distributions can be estimated and summarized in functional forms depending on β_r's. For example, certain linear combinations of β_r's can be interpreted as measures of the scale and shape of a probability distribution, e.g., multiples of $2\beta_1 - \beta_0$ can be used as assessments of scale parameters of a distribution function (Hosking et al. 1985b) and $6\beta_2 - 6\beta_1 + \beta_0$ denotes a skewness characteristic of a distribution function (Stedinger 1983). More generalized linear combinations of the PWMs can be constructed by employing orthogonal polynomials in a manner of specific PWMs called L-moments (Hosking 1990). Following the approach of Hosking et al. (1985b), in this article we focus on the PWMs in the form of β_r.

We develop an empirical likelihood (EL) based method for inferences about β_r. The statistical literature has shown that the EL methodology is a very powerful inference tool in the nonparametric statistics domain (Qin and Lawless 1994; Lazar and Mykland 1998; Owen 2001; Vexler et al. 2009, 2014). In order to provide background on the EL concept, we consider the commonly used EL ratio test for the null hypothesis that $H_0 : E(X_1) = 0$, based on i.i.d observations $X_1,...,X_n$. In this case, the EL function has the form $EL = \prod_{i=1}^{n} p_i$, where the probability weights p_i' s are derived by maximizing the EL function under empirical constraints $\sum_{i=1}^{n} p_i = 1$ and $\sum_{i=1}^{n} X_i p_i = \beta_0$ corresponding to the null hypothesis. In this manner β_0 represents the population mean. Under alternative hypothesis H_1, the EL function is given by by $EL = n^{-n}$, since $\prod_{i=1}^{n} p_i$ is maximized by $p_i = n^{-1}$ when the only constraint $\sum_{i=1}^{n} p_i = 1$ is in effect. Then the log EL ratio is defined as $\log ELR(\beta_0) = \prod_{i=1}^{n} np_i$. In this framework, values of p_i's can oftentimes be obtained numerically by solving Lagrangian equations regarding the maximization of EL given the constraints. The classical EL ratio test based on the statistic $-2\log ELR(\beta_0)$ has asymptotically a χ^2-distribution according to the nonparametric Wilks theorem (Owen 1990). In this article we propose the EL inference of β_r as an extension to the classical EL technique. We derive the asymptotic distribution of the proposed EL ratio test under the null distribution, developing the appropriate nonparametric version of the Wilks theorem. An extensive Monte Carlo (MC) study confirms the efficiency of the proposed methodology.

We apply the proposed method to develop an EL based test for the Gini index. The Gini index is a widely used measure for assessing distributional inequality. Let X and Y be two independent random variables from the same distribution $F(x) = \Pr(X \le x)$. The Gini mean difference was first defined in Gini (1912) as a distributional scale measure $D = E|X - Y|$ and the Gini index can be regarded as the normalized Gini's mean difference $G = D/(2\mu)$, where $\mu = \int xdF(x)$ is the population mean. The Gini index can also be expressed as the area between the 45-degree line and the Lorenz curve. The Lorenz curve, proposed by Lorenz (1905), is commonly used measure of distributional inequality and the 45-degree line represents perfect equality.

This article is organized as follows. In Section 2, we first review the existing techniques for estimating the PWMs and propose a novel estimator of β_r. This scheme to construct the new estimator motivates us to generalize the existing methods by developing the EL inferences of PWMs. The asymptotic distribution of the proposed EL test statistic is derived. This approximation is useful to construct the EL type CI estimations and to determine corresponding critical values of the proposed EL ratio test. The application of our method to the Gini index is shown in Section 2. In Section 3, an extensive MC study is conducted to investigate the performance of the proposed methodology and the theoretical results. Section 3 shows the numerical comparison between the proposed method and the known procedures. In Section 4, a real data example of myocardial infarction disease is used to illustrate the applicability of the proposed method. This data set is based on the biomarker related to heart disease. In Section 5, some conclusions are drawn.

2. Method

In this section, we outline the known estimation schemes regarding PWMs and construct a new estimator of β_r. Then we extend the classical EL methodology to be suitable for β_r and the Gini index evaluations. The asymptotic distributions of the proposed EL ratio statistics are presented in this section.

2.1 Estimators of the PWMs

Given a random sample $X_1,...,X_n$ of size n from an unknown distribution function F, estimation of β_r is most conveniently based on the order statistics $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$. Landwehr et al. (1979) proposed the estimator of β_r as

$$\tilde{b}_r = n^{-1}\sum_{j=1}^{n} k_{j,n} X_{(j)} \text{ with } k_{j,n} = \prod_{l=1}^{r} \frac{j-l}{n-l}.$$

To construct this estimator the authors focused on the relationships between moments of order statistics and β_r. Gelder and Pandy (2005) numerically compared the value of $k_{j,n}$ with $(j/n)^r$ to show that $k_{j,n} \approx (j/n)^r$ for all j ’ s. Hosking et al. (1985b) estimated β_r using the form

$$b_r = n^{-1}\sum_{j=1}^{n} q_{j,n}^{r} X_{(j)} ,$$

where $q_{j,n}$ is a plotting position which is a distribution free estimate of $F\left(X_{(j)}\right)$ and usually it is chosen to be $q_{j,n}=(j-a)/n$, $0<a<1$. Here $F(x)=P(X_1 \le x)$. Hosking et al. (1985b) employed an extensive simulation study to confirm that b_r with the choice of $a=0.35$ was the overall best estimator of β_r among a wide set of estimators. This estimator is asymptotically equivalent to $\tilde{b}_r$, derived by Landwehr et al. (1979). More recently in Section 11.4 of the book presented by David and Nagaraja (2003), the empirical estimator of β_r is shown to be

$$\bar{b}_r = n^{-1}\sum_{i=1}^{n}\left(\frac{i}{n}\right)^r X_{(i)}.$$

The idea of constructing $\bar{b}_r$ and b_r is directly followed by replacing the unknown distribution function F in the definition of β_r with its empirical counterpart $F_n(u)=\sum_{i=1}^{n} I\left(X_{(i)} \le u\right)/n$, where $I(\cdot)$ denotes the indicator function.

We consider the following approximation scheme

$$\begin{aligned}\beta_r &= \int x[F(x)]^r \, dF(x) \cong \sum_{i=1}^{n} \int_{X_{(i-1)}}^{X_{(i)}} x[F(x)]^r \, dF(x) \cong \sum_{i=1}^{n} X_{(i)} \int_{X_{(i-1)}}^{X_{(i)}} [F(x)]^r \, dF(x) \\ &\cong \sum_{i=1}^{n} X_{(i)} \left[F_n^{r+1}\left(X_{(i)}\right) - F_n^{r+1}\left(X_{(i-1)}\right)\right]\frac{1}{r+1}, \quad X_{(0)} = -\infty.\end{aligned}$$

This simple technique implies the new formula to estimate β_r using

$$\hat{b}_r = \sum_{i=1}^{n} X_{(i)} \left\{\left(\frac{i}{n}\right)^{r+1} - \left(\frac{i-1}{n}\right)^{r+1}\right\}\frac{1}{r+1}.$$

It turns out that asymptotically $\hat{b}_r$ behaves similarly to the well-known estimator $\bar{b}_r$ (for details, see Remark A1 presented in the supplementary materials).

To compare the estimators $\tilde{b}_r$, $\bar{b}_r$ and $\hat{b}_r$ based on fixed-size samples, we employed 1,000,000 MC generations of the i.i.d. sample $\{X_1,...,X_n\}$ at each sample size n and each value of $r=1,2,3,4$. In this limited numerical study we calculated the MC variances based on normally $(X \sim N(0,1))$, exponentially $(X \sim Exp(1))$ and lognormally $(X \sim LogN(0,1))$ distributed observations. Table 1 demonstrates that the proposed estimator $\hat{b}_r$ provides the smallest variances for relatively small sample sizes in the considered scenarios. The variances of the estimators are asymptotically equivalent. (Table SM1 in the supplementary materials displays values of the Monte Carlo approximations to the corresponding mean square errors of the estimators.)

Table 1. The MC variances multiplied by the sample size: $V_1 = nVar(\bar{b}_r)$, $V_2 = nVar(\hat{b}_r)$, $V_3 = nVar(\tilde{b}_r)$. The parameters' values are $\beta_r = 0.283$, 0.282, 0.257, 0.233 for $X \sim N(0,1)$; $\beta_r =$ 0.750, 0.611, 0.521, 0.457 for $X \sim Exp(1)$; $\beta_r =$ 1.253, 1.045, 0.910, 0.814 for $X \sim LogN(0,1)$ and $r = 1,\ldots,4$, respectively.

		$X \sim N(0,1)$			$X \sim Exp(1)$			$X \sim LogN(0,1)$		
n	r	V_1	V_2	V_3	V_1	V_2	V_3	V_1	V_2	V_3
15	1	0.3230	0.2886	0.2943	0.6084	0.5584	0.5862	3.6216	3.3556	3.5585
	2	0.1801	0.1510	0.1574	0.4460	0.3815	0.4160	3.1059	2.6865	3.0043
	3	0.1238	0.0974	0.1040	0.3605	0.2884	0.3261	2.8005	2.2660	2.6676
	4	0.0940	0.0694	0.0760	0.3038	0.2277	0.2669	2.5615	1.9412	2.4030
25	1	0.3098	0.2894	0.2927	0.5978	0.5678	0.5844	3.5907	3.4306	3.5528
	2	0.1692	0.1520	0.1558	0.4326	0.3937	0.4146	3.0747	2.8186	3.0131
	3	0.1136	0.0982	0.1021	0.3411	0.2981	0.3207	2.7384	2.4111	2.6584
	4	0.0848	0.0706	0.0744	0.2860	0.2402	0.2639	2.5067	2.1211	2.4106
50	1	0.3003	0.2902	0.2918	0.5911	0.5761	0.5844	3.5839	3.5033	3.5649
	2	0.1610	0.1526	0.1544	0.4222	0.4028	0.4132	3.0264	2.8977	2.9956
	3	0.1066	0.0991	0.1010	0.3303	0.3087	0.3201	2.6759	2.5106	2.6356
	4	0.0784	0.0714	0.0733	0.2717	0.2489	0.2607	2.4301	2.2347	2.3818

100	1	0.2953	0.2903	0.2911	0.5870	0.5795	0.5837	3.5703	3.5300	3.5607
	2	0.1573	0.1531	0.1541	0.4169	0.4072	0.4124	3.0239	2.9589	3.0084
	3	0.1033	0.0995	0.1005	0.3233	0.3125	0.3182	2.6670	2.5833	2.6468
	4	0.0754	0.0720	0.0729	0.2642	0.2528	0.2588	2.4092	2.3102	2.3849
250	1	0.2930	0.2910	0.2913	0.5831	0.5801	0.5817	3.5478	3.5317	3.5440
	2	0.1552	0.1536	0.1539	0.4135	0.4097	0.4117	3.0061	2.9801	2.9999
	3	0.1011	0.0996	0.1000	0.3188	0.3145	0.3168	2.6477	2.6141	2.6396
	4	0.0734	0.0721	0.0724	0.2608	0.2563	0.2586	2.3840	2.3442	2.3742
1200	1	0.2914	0.2910	0.2911	0.5836	0.5829	0.5833	3.5675	3.5642	3.5667
	2	0.1536	0.1533	0.1533	0.4107	0.4099	0.4103	3.0000	2.9945	2.9987
	3	0.1000	0.0997	0.0997	0.3182	0.3173	0.3178	2.6357	2.6287	2.6340
	4	0.0723	0.0720	0.0721	0.2587	0.2578	0.2583	2.3774	2.3691	2.3754

One can extend the estimation scheme based on $F_n(u)$ by using the general form of the empirical distribution function $\tilde{F}_n(u)=\sum_{i=1}^{n} w_i I\left(X_{(i)} \le u\right)$, where the weights w_i's, $0 < w_1,...,w_n < 1$, satisfy the assumption $\sum_{i=1}^{n} w_i = 1$. This approach implies more general forms of the estimators of β_r. For example, we can rewrite the estimators $\bar{b}_r$ and $\hat{b}_r$ as

$$\bar{b}_r = \sum_{i=1}^{n} w_i \left(\sum_{j=1}^{i} w_j \right)^r X_{(i)} \text{ and } \hat{b}_r = \sum_{i=1}^{n} X_{(i)} \left\{ \left(\sum_{j=1}^{i} w_j \right)^{r+1} - \left(\sum_{j=1}^{i-1} w_j \right)^{r+1} \right\} \frac{1}{r+1}.$$

Associating the weights w_i's and the probability weights p_i's in the EL framework allows us to develop EL inference about PWMs. For clarity and simplicity of explanation we will focus on $\hat{b}_r$ type constructions. The proposed EL type procedure can be modified using b_r and $\bar{b}_r$ type algorithms. Note that, in the context of a $\hat{b}_r$-based EL technique, we propose an exact algorithm to compute values of $p_i, i=1,...,n$, whereas, e.g., $\bar{b}_r$-based-EL's methods require very complicated schemes to estimate the corresponding probability weights (for details, see Remark A2 in the supplementary materials). In a subsequent paper, we plan to address this issue in the aspect of general EL L-estimate-type constructions, including linear combinations of PWMs. Towards this end sequential linearization of EL constraints and data-driven methods will be proposed to approximate probability weights related to $\bar{b}_r$-based-EL type procedures. Further theoretical and Monte Carlo studies are needed to complete this research.

2.2 EL Inference for β_r

Following the materials mentioned in the previous section, one can define the EL function for β_r as

$$L(\beta_r) = \max_{0<p_1,\ldots,p_n<1}\left[\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} X_{(i)} \left\{(S_i)^{r+1} - (S_{i-1})^{r+1}\right\}\frac{1}{r+1} = \beta_r\right],$$

where the probability weights $0 < p_1,\ldots,p_n < 1$, $S_i = \sum_{j=1}^{i} p_j$ and we let that $\sum_{i=n+1}^{n} p_i = 0$ and $S_0 = 0$.

In order to find the expressions of p_i's, $i = 1,...,n$, in $L(\beta_r)$, we denote the corresponding Lagrangian function

$$U = \sum_{i=1}^{n}\log(p_i) + \lambda_1\left(1 - \sum_{i=1}^{n} p_i\right) + \lambda_2\left(\beta_r - \sum_{i=1}^{n} X_{(i)}\left\{(S_i)^{r+1} - (S_{i-1})^{r+1}\right\}\frac{1}{r+1}\right),$$

where λ_1 and λ_2 are the Lagrange multipliers. Calculating p_k as roots of $\partial U/\partial p_k = 0$, $k = 1,...,n,$ one can show that

$$\frac{1}{p_k} - \lambda_1 - \lambda_2\left(X_{(k)}(S_k)^r + \sum_{i=k+1}^{n} X_{(i)}\left\{(S_i)^r - (S_{i-1})^r\right\}\right) = 0 \ .$$

Thus, taking into account the constraint $S_n = 1$ and summing up $\sum_{k=1}^{n} p_k \partial U/\partial p_k = 0$, we have

$$n-\lambda_1-\lambda_2\sum_{k=1}^{n}\left(X_{(k)}\left(S_k\right)^r+\sum_{i=k+1}^{n} X_{(i)}\left\{\left(S_i\right)^r-\left(S_{i-1}\right)^r\right\}\right)p_k=0. \tag{1}$$

To simplify Equation (1), we use the following well-known result that will be oftentimes applied throughout this article.

Lemma 1. Given two sequences a_i and b_j of real numbers, $i, j=1,\ldots,n$, we have

$$\sum_{j=1}^{n-1}\left(a_j\sum_{i=k+1}^{n} b_i\right)=\sum_{i=2}^{n}\left(b_i\sum_{j=1}^{i-1} a_j\right). \tag{2}$$

Applying Lemma 1 to Equation (1), we obtain

$$\begin{aligned}\sum_{k=1}^{n} p_k\left(\sum_{i=k+1}^{n} X_{(i)}\left\{\left(S_i\right)^r-\left(S_{i-1}\right)^r\right\}\right) &= \sum_{i=2}^{n} X_{(i)}\left\{\left(S_i\right)^r-\left(S_{i-1}\right)^r\right\}\sum_{j=1}^{i-1} p_j=\sum_{i=2}^{n} X_{(i)}\left\{\left(S_i\right)^r-\left(S_{i-1}\right)^r\right\}S_{i-1} \\ &= \sum_{i=2}^{n} X_{(i)}\left(S_i\right)^r\left(S_i-p_i\right)-\sum_{i=2}^{n} X_{(i)}\left(S_{i-1}\right)^{r+1}\end{aligned} \tag{3}$$

$$=\sum_{i=2}^{n} X_{(i)}\left\{\left(S_i\right)^{r+1}-\left(S_{i-1}\right)^{r+1}\right\}-\sum_{i=2}^{n} X_{(i)}\left(S_i\right)^r p_i=(r+1)\beta_r-X_{(1)}\left(S_1\right)^{r+1}-\sum_{i=2}^{n} X_{(i)}\left(S_i\right)^r p_i$$

$$=(r+1)\beta_r-\sum_{i=1}^{n} X_{(i)}\left(S_i\right)^r p_i,$$

where the constraint $\sum_{i=1}^{n} X_{(i)}\left\{\left(S_i\right)^{r+1}-\left(S_{i-1}\right)^{r+1}\right\}/(r+1)=\beta_r$ is used.

That is, Equation (1) can be rewritten as $n-\lambda_1-\lambda_2(r+1)\beta_r=0.$ This shows that $\lambda_1=n-\lambda_2(r+1)\beta_r$ and then the equation $\partial U/\partial p_k=0$ yields the expressions for $p_1,\ldots,p_n$ in the form

$$p_k=\left[n+\lambda_2\left(J_k-(r+1)\beta_r\right)\right]^{-1},\ \ k=1,\ldots,n, \tag{4}$$

where $J_k=X_{(k)}\left(S_k\right)^r+\sum_{i=k+1}^{n} X_{(i)}\left\{\left(S_i\right)^r-\left(S_{i-1}\right)^r\right\}$ and $S_k=1-\sum_{i=k+1}^{n} p_i$, for $k=1,\ldots,n$ and $S_0=0$. We have $S_n=1,$ and λ_2 is a numerical solution of the equation

$$\sum_{\mathrm{i=1}}^{\mathrm{n}} X_{(i)} \left\{ (S_i)^{r+1} - (S_{i-1})^{r+1} \right\} / (r+1) = \beta_r .$$

Note that values of λ_2 can be found by applying a standard numerical zero root finding algorithm, e.g., the Newton-Raphson scheme, which has been well equipped in the standard statistical software, can be readily employed for this purpose. We use the R (R Development Core Team 2012) functions *uniroot* and *optimize* to calculate the values of λ_2 . The corresponding R-code is shown in the supplementary materials. We illustrate the computing scheme in detail in the next section. In the case with $r = 0$ we have the regular expressions for $p_1, \ldots, p_n$ that are components of the classical EL methodology for the population mean estimation (Owen 2001).

Now we define the log EL ratio for the parameter β_r as $\log ELR(\beta_r) = \log\left(L(\beta_r) / n^{-n} \right)$. In a similar manner to the classical EL ratio definition, we denote $\log ELR(\beta_r) = -\infty$ if there are no such p_i 's that satisfy the constraints $\sum_{i=1}^{n} p_i = 1$ and $\sum_{i=1}^{n} X_{(i)} \left\{ (S_i)^{r+1} - (S_{i-1})^{r+1} \right\} / (r+1) = \beta_r$. Such cases can arise when the value of β_r is strongly not appropriate for underlying data distributions.

In the case of $r = 0$, the $-2\log ELR(\beta_0)$ has an asymptotic χ_1^2 distribution under the null hypothesis (Owen 1988). The next proposition extends the nonparametric version of Wilks theorem in the context of the proposed EL inference of $\beta_r = EX_1(F(X_1))^r$.

Proposition 1: Assume that i.i.d data points $X_1,..., X_n$ are from an unknown distribution function F , where its inverse function F^{-1} is continuous almost everywhere. If $E|X_1|^3 < \infty$ then

$$-2\log ELR(\beta_r) \xrightarrow{d} \chi_1^2 \text{ as } n \to \infty .$$

Remark: According to Proposition 1, one can derive the EL ratio test for the hypothesis that says $EX(F(X))^r = \beta_r$ for a specific value of β_r. We reject the hypothesis when $-2\log ELR(\beta_r) \geq \chi_1^2(1-\alpha)$, where $\chi_1^2(1-\alpha)$ is the $100(1-\alpha)\%$ percentile of the chi-square distribution with the degree of freedom one, and α is the significance level. By virtue of the relation between the testing and CI estimation, we can obtain the CI estimator of $EX(F(X))^r$ in the form of

$$\mathrm{CI}_{1\text{-}\alpha} = \left\{ \beta_r : \ -2\log ELR\left(\beta_r\right) \leq \chi_1^2(1\text{-}\alpha) \right\} ,$$

assuming that the nominal coverage probability is specified as $1-\alpha$.

2.3 A scheme to implement the proposed EL ratio Technique

To calculate values of $L(\beta_r) = \prod_{i=1}^{n} p_i$, we consider the following algorithm. Begin with defining $p_n = 1/\left[n + \lambda_2\left(X_{(n)} - (r+1)\beta_r\right)\right]$ and $S_{n-1} = 1 - p_n$, as a function of λ_2. Recursively given a value of λ_2 we have $p_{n-1} = 1/\left[n + \lambda_2\left\{X_{(n-1)}\left(S_{n-1}\right)^r + X_{(n)}\left(1-\left(S_{n-1}\right)^r\right) - (r+1)\beta_r\right\}\right]$ and $S_{n-2} = 1 - p_n - p_{n-1}$. Sequentially one can obtain values of $p_1, \ldots, p_n$ depending on λ_2. The appropriate λ_2 can be calculated by a zero root finding algorithm, e.g., the Newton-Raphson technique, employed to solve the equation $\sum_{i=1}^{n} X_{(i)}\left\{\left(S_i\right)^{r+1} - \left(S_{i-1}\right)^{r+1}\right\}/(r+1) - \beta_r = 0$ with respect to λ_2. Since $\sum_{k=1}^{n} p_k = 1$, we rewrite the equation above in the form of $C(\lambda_2) = 0$, where the function $C(\lambda_2) = \sum_{k=1}^{n} p_k\left(J_k - (1+r)\beta_r\right)$ and J_k is denoted in (4). Note that, in the case of $r = 0$, this equation is widely used in the standard EL methodology to find values of p_i's. Plugging the appropriate value for λ_2 to the expressions of p_i's, we obtain a value of

$-2\log ELR(\beta_r)$. The corresponding R code of the above algorithm is provided in the supplementary materials.

2.4 An application of the proposed method to make inference regarding the Gini Index.

The Gini index, *G*, is given as

$$G = \frac{E\,|\,X - Y\,|}{2EX} = \frac{1}{\beta_0}\int_0^{\infty}(2F(x)-1)x\,dF(x) = \frac{2\beta_1 - \beta_0}{\beta_0},$$

where β_1 and β_0 are the PWMs defined in Section 1 and *X*, *Y* are two independent random variables with non-negative values following the same distribution $F(x)$.

Following the method proposed in Section 2.2, one can define the EL function with respect to β_1 and β_0 as

$$L(\beta_1,\beta_0) = \max_{0<p_1,\ldots,p_n<1}\left[\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} X_{(i)}\left\{(S_i)^2 - (S_{i-1})^2\right\}/2 = \beta_1\,, \sum_{i=1}^{n} X_{(i)}p_i = \beta_0\right],$$

where the probability weights $0 < p_1,\ldots,p_n < 1,\;\; S_i = \sum_{j=1}^{i} p_j$ and $S_0 = 0$.

In a similar manner to the computing scheme in Section 2.2, one can derive the expressions of p_k's, $k = 1,\ldots,n$, in $L(\beta_1,\beta_0)$ by solving the corresponding Lagrangian function as

$$p_k = \left[n + \lambda_1(J_k - 2\beta_1) + \lambda_2(X_{(k)} - \beta_0)\right]^{-1},$$

where $J_k = X_{(k)}S_k + \sum_{i=k+1}^{n} X_{(i)}p_k$ and $S_k = 1 - \sum_{i=k+1}^{n} p_i$, for $k = 1,\ldots,n.$ We let $S_0 = 0$. In this case, $S_n = 1,$ and λ_1, λ_2 are numerical solutions of the equations

$$\sum_{i=1}^{n} X_{(i)}\left\{(S_i)^2 - (S_{i-1})^2\right\}/2 = \beta_1 \text{ and } \sum_{i=1}^{n} X_{(i)}p_i = \beta_0.$$

Note that values of λ_1 and λ_2 can be found by applying one of numerical zero root finding algorithms.

Since $\beta_0 = 2\beta_1/(1+G)$, we can rewrite the EL function in the form $L_G(G,\beta_1) = L(\beta_1, 2\beta_1/(1+G))$. Following the Qin and Lawless (1994) inference method pertaining to *G*, we propose the maximum EL ratio $\max_{\beta_1} L_G(G,\beta_1)/n^{-n}$. The next proposition defines the asymptotic distribution of the maximum EL ratio statistic.

Proposition 2: Under the assumptions of Proposition 1, we have

$$-2\log\left(\max_{\beta_1} L_G(G,\beta_1)/n^{-n}\right) \xrightarrow{d} \chi_1^2, \text{ as } n \to \infty,$$

when $E|X_2 - X_1|/EX_1$ is known to be equal to 2*G*.

The proof of Proposition 2 is directly based on a technical combination of the proof scheme of Proposition 1 and the results of Qin and Lawless (1994), and then the proof is omitted.

Remark: According to Proposition 2, one can derive the EL ratio test for the hypothesis $E|X-Y|/(2EX) = G$. We reject the null hypothesis when $-2\log ELR\left(G, \beta_1^{\mathrm{M}}\right) \geq \chi_1^2(1-\alpha)$, where β_1^M is the value of β_1 that maximizes $L_G(G,\beta_1)$ and $\chi_1^2(1-\alpha)$ is the $100(1-\alpha)\%$ percentile of the chi-square distribution with the degree of freedom one, and α is the significance level. By the virtue of the relation between the testing and CI estimation, we can obtain the CI estimator of *G* in the form of

$$\mathrm{CI}_{1\text{-}\alpha} = \left\{G:\ -2\log ELR\left(G, \beta_1^{\mathrm{M}}\right) \leq \chi_1^2(1\text{-}\alpha)\right\},$$

assuming that the nominal coverage probability is specified as $1-\alpha$. In the supplementary material we provide results of a limited MC study to compare the proposed method with the EL method of testing the Gini index provided by Qin et al. (2010). The method of Qin et al. (2010)

suggests using the empirical distribution function $F_n(x)=\sum_{i=1}^{n} I(X_i < x)/n$, while normalizing the Gini index in the EL manner. The MC study demonstrated the Type I error control related to the proposed method is significantly more accurate than that of the method by Qin et al. (2010).

3. Monte Carlo Evaluations

In this section, we demonstrate results of a MC study related to comparisons of properties of the following tests: (1) the proposed EL ratio test; (2) the classical EL ratio test; and (3) a test based on asymptotic properties of the estimator $\overline{b}_r$. In order to apply the classical EL technique, we pretend that data in the form of $Z_{(i)} = X_{(i)}\left(F(X_{(i)})\right)^r$, $i=1,..,n,$ can be observed in order to consider the classical EL ratio test for β_1 and β_2 based on $Z_{(1)},...,Z_{(n)}$. In this context, we hypothetically assume the underlying data distribution is known. In practice, this EL method cannot be performed in the nonparametric setting. According to David and Nagaraja (2003), the estimator $\overline{b}_r$ has an asymptotic normal distribution. This implies constructing a $\overline{b}_r$-based testing procedure. By virtue of the test construction's scheme for $\overline{b}_r$, we can expect the $\overline{b}_r$-based test will provide good operating characteristics when normally distributed observations are used. In this MC experiments, the underlying data distributions were chosen to be in Normal, Exponential, Chi-square and Lognormal forms, since the statistical literature (Vexler et al. 2009) showed EL type tests provide good properties when data are generated from a normally distributed population and the Type I error control is not robust to the scenarios when skewness of data distributions is in effect. For each baseline distribution, we repeated 50,000 samples of observations with sizes n=20, 25, 50, 150 and 300. We set up the expected significant level of the considered tests to be $\alpha = 0.05$. Tables 2-4 present the results of the actual Type I error rates, the power comparisons and the CI estimations, respectively.

Table 2 shows the MC Type I error rates of the tests for $H_0: E\left[XF^r(X)\right] = \beta_r$, given different sample sizes *n*. The proposed test and the classical EL ratio test outperform the $\overline{b_r}$-based test in the context of the Type I error control for most of the considered scenarios. The MC Type I error rates of the proposed test are closer to the expected 0.05 than those of the classical EL ratio tests based on the normally distributed data with relatively small sample sizes 20, 25 and 50. The proposed test performs well in control of the Type I error rates for the exponential distribution based on relatively small samples with sizes 25 and 50, and has a fairly good control regarding the Type I error rates for a mildly skewed chi-square distribution, when the sample size is increased to 50. For the heavily skewed lognormal distribution, the proposed test has better Type I error rate control than the $\overline{b_r}$-based test, and the actual Type I error rates of the proposed test are better than the classical EL test for relatively small samples sizes n=20, 25 and *r=2*, the two behave similarly as sample sizes increase to 50, 150 and 300.

Table 3 shows the results of the MC power comparisons of the considered tests when the null hypothesis is associated with the three scenarios: (1) $\beta_1 = 0.75$ and $\beta_2 = 0.61$, (2) $\beta_1 = 0.2821$ and $\beta_2 = 0.2820$, and (3) $\beta_1 = 1.2534$ and $\beta_2 = 1.0448$. Corresponding to these scenarios, the alternative data distributions were: (a) $X \sim Exp(\text{rate})$ with *rate*=0.9 ($\beta_1 = 0.83$, $\beta_2 = 0.68$); *rate*=0.8 ($\beta_1 = 0.94$, $\beta_2 = 0.76$); *rate*=0.7 ($\beta_1 = 1.07$, $\beta_2 = 0.97$); and that *rate*=1 is related to scenario (1); (b) $X \sim Normal(0, \sigma^2)$ with $\sigma^2 = 4$ ($\beta_1 = 0.5642$, $\beta_2 = 0.5641$); $\sigma^2 = 9$ ($\beta_1 = 0.8469$, $\beta_2 = 0.8467$); and that $\sigma^2 = 1$ corresponds to scenario (2); (c)

$X \sim Lognormal(0, \sigma^2)$ with $\sigma^2 = 2.25$ ($\beta_1 = 2.6354$, $\beta_2 = 2.3589$); $\sigma^2 = 4$ ($\beta_1 = 6.8079$, $\beta_2 = 6.3972$); and that $\sigma^2 = 1$ corresponds to scenario (3). In the normal cases, $\overline{b_r}$ -based test is expected to have good performance due to its construction mentioned in Sec. 3. When $r = 1$, the proposed test is more powerful than the $\overline{b_r}$ -based test, and the EL test is a bit more powerful than the proposed test. Note that the EL test based on $Z_{(i)}$'s is used here for the purpose of illustrating how the proposed test performs by comparing with a test that uses some unobserved information as noted in the beginning of this section. For the mildly skewed exponential cases, the proposed test is the most powerful one among the considered tests for β_1. When $r = 2$, the $\overline{b_r}$ -based test has the largest actual power for sample sizes n =20, 25. As the sample size increases up to 50, the proposed test has the largest actual power. For the heavily skewed lognormal cases, the proposed test is the best of all the three test procedures for most considered cases. And the proposed test is significantly more powerful than the other two tests especially for relatively small sample sizes n =20, 25 (*r=1*).

In Table 4, the 95% CI's for β_1 and β_2 are investigated via CI mechanisms based on the considered on inverting the test statistics. In the normal cases, the proposed test provides the actual coverage probabilities (CP) that are closer to the target 0.95 than those of the $\overline{b_r}$ -based CI estimation. Specially, the proposed CI estimation performs reasonably well and is more competitive than the $\overline{b_r}$ -based CI estimation in cases of relatively small sample sizes *n*=20, 25. Similar results hold for data generated from skewed distributions of exponential, χ_3^2 and lognormal. Note that the classical EL CI estimation is hypothetically used to compare with the proposed CI estimation. The two have similar results for most considered cases. In practice the EL CI estimation however cannot be used in nonparametric setting.

Table 2. The MC Type I Error rates of the considered tests based on samples followed the null distributions: Normal(0,1), Exp(1), χ_3^2, Lognormal(0,1) distributions. The sample sizes n= 20, 25, 50, 150 and 300 and r= 1,2 at $E\left[XF^r(X)\right]$. The notations: pro.test, EL.test, $\bar{b}_r$ .test correspond to the proposed EL ratio test, the classical EL ratio test, and the test based on the asymptotic properties of $\bar{b}_r$, respectively (The expected Type I error is 0.05).

$X \sim Normal(0,1)$		Sample sizes				
r	Test	20	25	50	150	300
r=1	pro.test	0.077	0.067	0.058	0.054	0.050
	EL.test	0.083	0.072	0.064	0.054	0.047
	$\bar{b}_r$ .test	0.098	0.087	0.072	0.056	0.056
r=2	pro.test	0.055	0.051	0.053	0.057	0.051
	EL.test	0.089	0.080	0.056	0.057	0.051
	$\bar{b}_r$ .test	0.107	0.100	0.070	0.062	0.056
$X \sim Exp(1)$		20	25	50	150	300
r=1	pro.test	0.115	0.084	0.076	0.055	0.051
	EL.test	0.110	0.083	0.073	0.054	0.052
	$\bar{b}_r$ .test	0.136	0.108	0.086	0.063	0.055
r=2	pro.test	0.034	0.050	0.078	0.059	0.058
	EL.test	0.113	0.095	0.074	0.061	0.056
	$\bar{b}_r$ .test	0.146	0.132	0.093	0.071	0.061
$X \sim \chi_3^2$		20	25	50	150	300
r=1	pro.test	0.097	0.092	0.069	0.059	0.051
	EL.test	0.080	0.082	0.065	0.056	0.058
	$\bar{b}_r$ .test	0.103	0.103	0.075	0.060	0.058
r=2	pro.test	0.057	0.051	0.076	0.059	0.051
	EL.test	0.098	0.084	0.063	0.056	0.049
	$\bar{b}_r$ .test	0.131	0.114	0.077	0.066	0.062
$X \sim Lognormal(0,1)$		20	25	50	150	300
r=1	pro.test	0.153	0.135	0.106	0.082	0.058
	EL.test	0.141	0.119	0.092	0.072	0.065
	$\bar{b}_r$ .test	0.189	0.165	0.123	0.094	0.077
r=2	pro.test	0.059	0.065	0.108	0.075	0.059
	EL.test	0.157	0.146	0.100	0.074	0.059
	$\bar{b}_r$ .test	0.209	0.197	0.143	0.098	0.082

Table 3. The MC power comparisons of the tests for $E\left[X\left(F\left(X\right)\right)^r\right]=\beta_r$, $r=1,2$ based on $X_1,...,X_n \sim$ Exponential , Normal and Lognormal .

	$X \sim Normal(0,\sigma^2)$	*n*			
r=1		20	25	50	150
σ^2=4	pro.test	0.21	0.29	0.53	0.90
	EL.test	0.24	0.29	0.52	0.91
	$\overline{b}_r$ .test	0.21	0.25	0.46	0.87
σ^2=9	pro.test	0.36	0.44	0.72	0.98
	EL.test	0.43	0.51	0.78	0.99
	$\overline{b}_r$ .test	0.35	0.41	0.68	0.95
r=2		20	25	50	150
σ^2=4	pro.test	0.33	0.41	0.69	0.99
	EL.test	0.33	0.40	0.63	0.98
	$\overline{b}_r$ .test	0.42	0.49	0.74	0.99
σ^2=9	pro.test	0.61	0.69	0.95	1.00
	EL.test	0.60	0.67	0.95	1.00
	$\overline{b}_r$ .test	0.65	0.70	0.94	1.00
	$X \sim Exp(\text{rate})$	*n*			
r=1		20	25	50	150
rate =0.9	pro.test	0.11	0.11	0.14	0.26
	EL.test	0.10	0.10	0.13	0.25
	$\overline{b}_r$ .test	0.10	0.10	0.11	0.20
rate =0.8	pro.test	0.19	0.21	0.36	0.78
	EL.test	0.18	0.15	0.36	0.78
	$\overline{b}_r$ .test	0.15	0.16	0.28	0.73
r=2			25	50	150
rate =0.9	pro.test	0.08	0.09	0.12	0.27
	EL.test	0.11	0.10	0.12	0.26
	$\overline{b}_r$ .test	0.12	0.13	0.11	0.20
rate =0.8	pro.test	0.18	0.21	0.35	0.77
	EL.test	0.19	0.21	0.36	0.77
	$\overline{b}_r$ .test	0.20	0.21	0.29	0.71
	$X \sim Lognormal(0,\sigma^2)$	*n*			
r=1		20	25	50	150
σ^2=2.25	pro.test	0.39	0.46	0.69	0.98
	EL.test	0.32	0.38	0.59	0.95
	$\overline{b}_r$ .test	0.12	0.15	0.31	0.86
σ^2=4	pro.test	0.73	0.81	0.99	1.00
	EL.test	0.65	0.74	0.93	1.00

	$\bar{b}_r$ .test	0.21	0.28	0.53	0.88
r=2		20	25	50	150
σ^2 =2.25	pro.test	0.33	0.42	0.74	0.99
	EL.test	0.33	0.36	0.56	0.93
	$\bar{b}_r$ .test	0.19	0.20	0.36	0.87
σ^2 =4	pro.test	0.59	0.65	0.86	0.96
	EL.test	0.64	0.72	0.92	0.99
	$\bar{b}_r$ .test	0.30	0.35	0.59	0.87

Table 4. The MC CI estimations of β_r, $r=1,2$, when $X_1,...,X_n$ ~ Normal(0,1), Exp(1), χ_3^2 and Lognormal(0,1) with n= 25, 50, 150 and 300. Notations: The Pro.ci, EL.ci, $\bar{b}_r$ .ci represent for the proposed EL ratio CI estimation, the classical EL ratio CI method, and the empirical estimation CI respectively (The expected coverage probability (CP) is 95% and LG denotes the MC average length of CI estimation)

Sample sizes		25		50		150		300	
$X \sim Normal(0,1)$		CP	LG	CP	LG	CP	LG	CP	LG
	Pro.ci	0.94	0.41	0.94	0.29	0.95	0.17	0.95	0.12
r=1	EL.ci	0.93	0.45	0.93	0.32	0.96	0.18	0.94	0.13
	$\bar{b}_r$ .ci	0.91	0.41	0.93	0.29	0.95	0.17	0.93	0.12
	Pro.ci	0.89	0.33	0.92	0.26	0.94	0.15	0.95	0.10
r=2	EL.ci	0.89	0.40	0.91	0.29	0.94	0.17	0.95	0.11
	$\bar{b}_r$ .ci	0.87	0.29	0.91	0.21	0.96	0.13	0.94	0.09
$X \sim Exp(1)$		CP	LG	CP	LG	CP	LG	CP	LG
	Pro.ci	0.91	0.54	0.93	0.41	0.95	0.24	0.95	0.16
r=1	EL.ci	0.93	0.75	0.94	0.55	0.94	0.32	0.95	0.22
	$\bar{b}_r$ .ci	0.88	0.54	0.90	0.40	0.94	0.24	0.96	0.17
	Pro.ci	0.86	0.74	0.91	0.71	0.97	0.69	0.96	0.67
r=2	EL.ci	0.88	0.73	0.92	0.53	0.98	0.31	0.96	0.21
	$\bar{b}_r$ .ci	0.84	0.44	0.91	0.32	0.98	0.21	0.95	0.14
$X \sim \chi_3^2$		CP	LG	CP	LG	CP	LG	CP	LG
	Pro.ci	0.90	2.11	0.92	1.90	0.94	0.86	0.95	0.41
r=1	EL.ci	0.91	1.93	0.92	1.37	0.93	0.80	0.96	0.56
	$\bar{b}_r$ .ci	0.90	1.29	0.92	0.93	0.93	0.56	0.93	0.40
	Pro.ci	0.86	1.37	0.93	1.36	0.93	0.89	0.93	0.75
r=2	EL.ci	0.89	1.81	0.93	1.36	0.94	0.77	0.95	0.55
	$\bar{b}_r$ .ci	0.87	0.97	0.91	0.77	0.91	0.46	0.94	0.34
$X \sim Lognormal(0,1)$		CP	LG	CP	LG	CP	LG	CP	LG
	Pro.ci	0.87	2.32	0.90	2.16	0.92	1.56	0.95	1.54
r=1	EL.ci	0.86	1.50	0.92	1.16	0.93	0.67	0.93	0.50
	$\bar{b}_r$ .ci	0.82	1.51	0.88	0.91	0.89	0.54	0.92	0.42
	Pro.ci	0.82	1.37	0.85	1.10	0.93	0.69	0.94	0.51

r=2	EL.ci	0.88	1.56	0.88	1.11	0.94	0.69	0.95	0.48
	$\bar{b}_r$ .ci	0.83	1.06	0.83	0.78	0.90	0.52	0.92	0.37

4. Data Example

In this section, a real data example is presented to illustrate the applicability of the proposed method. The example is based on data from a study that evaluated biomarkers related to the myocardial infarction (MI). The study was focused on the residents of Erie and Niagara counties, 35-79 years of age (Schisterman et al. 2001). The New York State department of Motor Vehicles drivers' license rolls was used as the sampling frame for adults between the age of 35 and 65 years, while the elderly sample (age 65-79) was randomly chosen from the Health Care Financing Administration database. We consider the biomarker "Vitamin E" supplement that is often used to quantify antioxidant status of an individual and scientific literature showed that it could prevent heart disease (Rimm et al. 1993). A total of 2390 measurements of Vitamin E were evaluated by the study. 547 of them were collected on cases who survived on MI and the other 1843 on controls who had no previous MI.

To illustrate and examine the proposed method based on the Vitamin E data we employ the following technique: The strategy for the case group was that a sample with size n was randomly selected from the Vitamin E data to estimate the CIs at the 95% level of the PWMs β_1, say [a, b] using the proposed CI method and the $\bar{b}_r$-based CI method that utilizes asymptotic normality approximations. Note that the EL ratio test cannot be used in this nonparametric setting since the distribution of the Vitamin E data is unknown. The rest of the data (547-n observations) are used to compute the β_1 using the existing estimator $\bar{b}_1$ in Section 2.1. The value of (547-n) was chosen to be relatively large so that the calculated $\bar{b}_1$ estimator are close to the theoretical value of β_1. We repeated this strategy 3,000 times observing the frequencies of the event of $\bar{b}_1 \in [a, b]$. We repeat the same process for the control group. Table 5 presents these results for the different sample sizes of n=25 and 50. The outcomes in Table 5 show that the proposed method provides the coverage probabilities that are closer to the expected 0.95 level than those of the $\bar{b}_r$-based method.

Table 5. The coverage probabilities related to the Vitamin E evaluation. (The expected coverage probability is 0.95)

	Case group (N=547)		Control Group (N=1843)	
Sample size n	Proposed CI	$\bar{b}_r$-CI	Proposed CI	$\bar{b}_r$-CI
25	93.8%	90.3%	92.5%	89.4%
50	94.8%	91.0%	93.9%	91.4%

Thus the proposed method can be recommended to construct CI estimator of β_1. The results for 95% CI estimations of β_1 and β_2 with respect to the Vitamin E biomarker for both the case and control groups are shown in Table 6. The results show that the proposed CI estimation is very similar to the $\bar{b}_r$-based estimation proposed by David and Nagaraja (2003). Both methods show evidence that the levels of β_1 and β_2 of the Vitamin E Biomarker are significantly different among the case and control groups. We notice that the proposed method may have some skewness-correction for the CI estimation with respect to the skewed underlying data distribution based on the advantage of the empirical likelihood methodology (Vexler et al. 2009).

Table 6. The 95% CI estimation of β_1 and β_2 related to the Vitamin E biomarker with respect to the case and control groups.

	Case group (N=547)		Control Group (N=1843)	
	β_1	β_2	β_1	β_2
Proposed CI	(7.692, 8.345)	(5.839, 6.240)	(8.374, 8.717)	(6.227, 6.506)
$\bar{b}_r$-CI	(7.671, 8.344)	(5.696, 6.240)	(8.360, 8.770)	(6.217, 6.556)

5. Concluding Remarks

In this article, we proposed a general scheme to construct EL inference of the PWMs. The asymptotic evaluation of the proposed method can be considered as an extension of the nonparametric version of Wilks theorem. The statistical test and CI estimation of the PWMs are derived based on the asymptotic proposition. An extensive MC study confirmed the efficiency of the proposed method across a wide variety of underlying data distributions especially in the cases of relatively small sample sizes and/or data generated from skewed distributions. We showed that the proposed method can be easily applied to make inference of the Gini index. The real-life example showed excellent applicability of the proposed method.

SUPPLEMENTARY MATERIALS

Remarks A1 and A2, the proof of Proposition 1, R code to implement the proposed method and additional Monte Carlo results are available in the supplementary materials.

Acknowledgements

We are grateful to the Editor, Associate Editor and reviewer for their helpful comments that led to a substantial improvement in this paper.

REFERENCES

David, H. A., and Nagaraja, H. N., 2003. Order statistics. 3rd ed. Wiley Interscience (John Wiley and Sons), Hoboken, NJ.

Gelder, P., and Pandey, M. D., 2005. An analysis of drag forces based on L-moments. Advances in Safety and Reliability, 1, 661-668.

Gini, C., 1912. Variabilità e Mutabilità: contributo allo Studio delle distribuzioni e delle relazioni statistiche. Facoltá di Giurisprudenza della R. Universitá dei Cagliari, anno III, parte 2ª.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R., 1979. Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form. Water Resource Research, 15(5), 1049-1054.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F., 1985a. An appraisal of the regional flood frequency procedure in the UK Flood Studies Report. Hydrological Sciences Journal, 30(1), 85-109.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F., 1985b. Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. Technometrics, 27, 251-261.

Hosking, J. R. M., 1990. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. J. R. Stat. Soc. Ser. B 52(1), 105-124.

Katz, R. W., Parlange, M. B., and Naveau, P., 2002. Statistics of extremes in hydrology. Advances in Water Resources, 25, 1287-1304.

Lawless, J. F., 1982. Statistical Models and Methods for Lifetime Data. John Wiley & Sons, New York.

Lazar, N. A., and Mykland, P. A., 1998. An Evaluation of the Power and Conditionality Properties of Empirical Likelihood. Biometrika, 85, 523-524.

Landwehr, J. M., and Matalas, N. C., 1979. Probability Weighted Moments Compared With Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles. Water Resource Research, 15(5), 1055-1064.

Lorenz, M. O., 1905. Methods for measuring concentration of wealth. J. Amer. Statist. Assoc. 9, 209-219.

Nature Environment Research Council, 1975, Flood Studies Report, I, Hydrological Studies, London.

Owen, A. B., 1990. Empirical Likelihood Ratio Confidence Regions. Ann. Stat. 18(1), 90-120.

Owen, A. B., 2001. Empirical Likelihood. Chapman and Hall/CRC, New York.

Qin, J., and Lawless, J., 1994. Empirical Likelihood and General Estimating Equations. Ann. Stat. 22, 300-325

Qin, Y. S., Rao, J. N. K., and Wu, C. B., 2010. Empirical likelihood confidence intervals for the Gini measure of income inequality. Economic Modelling, 27(6), 1429-1435

R Development Core Team, 2012. R: A language and environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. *http://www.R-project.org*

Stedinger, J. R., 1983. Estimating a Regional Flood Frequency Distribution. Water Resource Research, 19(2), 503-510.

Rimm E.B., Stampfer M.J., Ascherio A., Giovannucci E., Colditz G.A., and Willett W.C., 1993. Vitamin E Consumption and the Risk of Coronary Heart Disease in Men. N. Engl. J. Med. 328, 1450-1456.

Schisterman, E. F., Faraggi, D., Browne, R., Freudenheim, J., Dorn, J., Muti, P., Armstrong, D., Reiser, N., and Trevisan, M., 2001. TBARS and Cardiovascular Disease in a Population-Based Sample. Journal of Cardiovascular Risk, 8, 219-225.

Vexler, A., Liu, S., Kang, L., and Hutson, A.D., 2009. Modifications of the empirical likelihood interval estimation with improved coverage probabilities. Commu. Stat. Theor. M. B38, 2173-83.

Vexler, A., Tao, G., and Hutson, A. D., 2014. Posterior expectation based on empirical likelihoods. Biometrika, 101, 711-7118.

Wallis, J.R., and Wood, E. F., 1985. Relative Accuracy of Log Pearson III Procedures, A.S.C.E. Journal of hydraulic Engineering, 3(7), 1043-1056.

Supplementary Materials for "An Extension to Empirical Likelihood for Evaluating Probability Weighted Moments"

Albert Vexler, Li Zou and Alan D. Hutson

Remark A1.

The Taylor theorem shows that

$$(u - s)^{r+1} = (u)^{r+1} - s\left(r+1\right)(u)^{r} + 0.5s^{2}\left(r+1\right)r\left(u-\theta\right)^{r-1},$$

for $\theta \in (0,1)$. Then, defining $u = i/n,\ s = 1/n$, we obtain

$$\hat{b}_r = \sum_{i=1}^{n} X_{(i)} \left\{ \left(\frac{i}{n}\right)^{r+1} - \left(\frac{i-1}{n}\right)^{r+1} \right\} \frac{1}{r+1} = \frac{1}{n} \sum_{i=1}^{n} X_{(i)} \left(\frac{i}{n}\right)^{r} + \frac{r}{2n^2} \sum_{i=1}^{n} X_{(i)} \left(\frac{i}{n} - \theta_{i,n}\right)^{r-1}$$

$$= \bar{b}_r + \frac{r}{2n^2} \sum_{i=1}^{n} X_{(i)} \left(\frac{i}{n} - \theta_{i,n}\right)^{r-1}, \quad \theta_{i,n} \in (0,1).$$

Since $0 < i/n < 1$, it is clear there exists a nonrandom constant $c_r \geq \max_{1 \leq i \leq n} \left| (i/n - \theta_{i,n})^{r-1} \right|$.

Thus the remainder term

$$\left| \frac{r}{2n^2} \sum_{i=1}^{n} X_{(i)} \left(\frac{i}{n} - \theta_{i,n}\right)^{r-1} \right| \leq \frac{rc_r}{2n^2} \sum_{i=1}^{n} \left| X_{(i)} \right| = \frac{rc_r}{2n^2} \sum_{i=1}^{n} \left| X_i \right| \ .$$

This implies, e.g., if $E\left|X_1\right|^{1+\varepsilon} < \infty$, for $\varepsilon > 0$, we have $\hat{b}_r = \bar{b}_r + O\left(n^{-1}\right)$, as $n \to \infty$. Note that asymptotic properties of $\bar{b}_r$ can be found in the literature dealt with L-estimates (e.g., David and Nagaraja, 2003; Serfling, 1980).

Remark A2: ***$\overline{b_r}$ -based EL Inference for β_r.***

Define the EL function for β_r as

$$L(\beta_r) = \max_{0 < p_1, \ldots, p_n < 1} \left[\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i X_{(i)} (S_i)^r = \beta_r \right],$$

where the probability weights $0 < p_1, \ldots, p_n < 1,\ \ S_i = \sum_{j=1}^{i} p_j$.

In order to express p_i, $i = 1, \ldots, n$, we denote the corresponding Lagrangian function

$$Q=\sum_{i=1}^{n}\log(p_i)+\lambda_1\left(1-\sum_{i=1}^{n}p_i\right)+\lambda_2\left(\beta_r-\sum_{i=1}^{n}p_iX_{(i)}\left(S_i\right)^r\right),$$

where λ_1 and λ_2 are the Lagrange multipliers. Calculating p_k as the roots of $\partial Q/\partial p_k=0$, $k=1,...,n,$ we have

$$\frac{1}{p_k}-\lambda_1-\lambda_2\left(X_{(k)}\left(S_k\right)^r+r\sum_{i=k}^{n}p_iX_{(i)}\left(S_i\right)^{r-1}\right)=0\ .$$

Thus, considering $\sum_{k=1}^{n}p_k\partial Q/\partial p_k=0$ with $S_n=1$, we have

$$n-\lambda_1-\lambda_2\sum_{k=1}^{n}\left(X_{(k)}\left(S_k\right)^r+r\sum_{i=k}^{n}p_iX_{(i)}\left(S_i\right)^{r-1}\right)p_k=0.$$

To simplify this equation, we use Lemma 1 presented in Section 2.2, obtaining that

$$r\sum_{k=1}^{n}p_k\sum_{i=k}^{n}p_iX_{(i)}\left(S_i\right)^{r-1}=r\sum_{j=1}^{n}p_j\left(S_j\right)^rX_{(j)}=r\beta_r,$$

where the constraint $\sum_{i=1}^{n}p_iX_{(i)}\left(S_i\right)^r=\beta_r$ is used. That is, $\sum_{k=1}^{n}p_k\partial Q/\partial p_k=0$ can be represented in the form $n-\lambda_1-\lambda_2(r+1)\beta_r=0.$ Then $\lambda_1=n-\lambda_2(r+1)\beta_r$ and the equation $\partial Q/\partial p_k=0$ yield $p_k=\left[n+\lambda_2\left(D_k-(r+1)\beta_r\right)\right]^{-1},\quad k=1,\dots,n$, where $D_k=X_{(k)}\left(S_k\right)^r+r\sum_{i=k}^{n}p_i\left(S_i\right)^{r-1}X_{(i)}$, $k=1,\dots,n$. This provides non-explicit functions to find values of p_k, $k=1,...,n,$ depending on λ_2 via the scheme: p_n can be derived using $p_n=\left[n+\lambda_2\left(X_{(n)}+rp_nX_{(n)}-(r+1)\beta_r\right)\right]^{-1}$ then p_{n-1} can be derived using $p_{n-1}=\left[n+\lambda_2\left(X_{(n-1)}(1-p_n)^r+rp_{n-1}X_{(n-1)}(1-p_n)^{r-1}+rp_nX_{(n)}-(r+1)\beta_r\right)\right]^{-1}$ and so on. Now, λ_2 can be computed applying the equation $\sum_{i=1}^{n}p_i\left[X_{(i)}\left(S_i\right)^r-\beta_r\right]=0$. Note that, in general, the corresponding expression of p_k gives two root values of p_k, that we reconsider satisfying $0 < p_1, \ldots, p_n < 1$ and these root values depend on λ_2. This strong complexity may lead the direct use of the $\bar{b}_r$ -based EL Inference for β_r to be inapplicable.

Table SM1. The Monte Carlo (MC) mean square errors (MSE) multiplied by the sample size: $M_1 = nMSE\left(\bar{b}_r\right)$, $M_2 = nMSE\left(\hat{b}_r\right)$, $M_3 = nMSE\left(\tilde{b}_r\right)$. The parameters' values are $\beta_r = 0.283$, 0.282, 0.257, 0.233 for $X \sim N(0,1)$; $\beta_r = 0.750$, 0.611, 0.521, 0.457 for $X \sim Exp(1)$; $\beta_r = 1.253$, 1.045, 0.910, 0.814 for $X \sim LogN(0,1)$ and $r = 1, \ldots, 4$, respectively. The results are based on 1,000,000 MC generations of samples $\{X_1, ..., X_n\}$ at each sample size n and each value of $r = 1,2,3,4$.

		$X \sim N(0,1)$			$X \sim Exp(1)$			$X \sim LogN(0,1)$		
n	r	M_1	M_2	M_3	M_1	M_2	M_3	M_1	M_2	M_3
15	1	0.3284	0.2939	0.2943	0.6126	0.5626	0.5862	3.6319	3.3679	3.5585
	2	0.1802	0.1563	0.1574	0.4574	0.3889	0.4160	3.1321	2.7119	3.0043
	3	0.1250	0.1024	0.1040	0.3801	0.2973	0.3261	2.8441	2.3023	2.6676
	4	0.0975	0.0741	0.0760	0.3312	0.2379	0.2669	2.6229	1.9870	2.4030
25	1	0.3130	0.2926	0.2927	0.6003	0.5703	0.5844	3.5970	3.4379	3.5528
	2	0.1692	0.1553	0.1558	0.4396	0.3981	0.4146	3.0905	2.8339	3.0131
	3	0.1144	0.1013	0.1021	0.3528	0.3036	0.3207	2.7649	2.4330	2.6584
	4	0.0871	0.0734	0.0744	0.3028	0.2463	0.2639	2.5441	2.1489	2.4106
50	1	0.3019	0.2918	0.2918	0.5924	0.5773	0.5844	3.5873	3.5068	3.5649
	2	0.1610	0.1541	0.1544	0.4257	0.4050	0.4132	3.0342	2.9055	2.9956
	3	0.1071	0.1006	0.1010	0.3362	0.3114	0.3201	2.6893	2.5216	2.6356
	4	0.0795	0.0729	0.0733	0.2800	0.2520	0.2607	2.4484	2.2493	2.3818
100	1	0.2961	0.2911	0.2911	0.5877	0.5801	0.5837	3.5716	3.5321	3.5608
	2	0.1573	0.1539	0.1541	0.4186	0.4083	0.4124	3.0278	2.9628	3.0084
	3	0.1035	0.1003	0.1005	0.3262	0.3140	0.3182	2.6739	2.5887	2.6468
	4	0.0760	0.0727	0.0729	0.2682	0.2545	0.2588	2.4187	2.3172	2.3849
250	1	0.2933	0.2913	0.2913	0.5833	0.5804	0.5817	3.5483	3.5325	3.5440
	2	0.1552	0.1539	0.1539	0.4142	0.4101	0.4117	3.0077	2.9816	2.9999
	3	0.1012	0.0999	0.1000	0.3200	0.3151	0.3168	2.6501	2.6165	2.6396
	4	0.0736	0.0723	0.0724	0.2625	0.2569	0.2586	2.3878	2.3471	2.3742
1200	1	0.2915	0.2911	0.2911	0.5836	0.5830	0.5833	3.5677	3.5643	3.5667
	2	0.1536	0.1533	0.1533	0.4108	0.4100	0.4103	3.0002	2.9949	2.9987
	3	0.1000	0.0997	0.0997	0.3185	0.3175	0.3178	2.6362	2.6291	2.6340
	4	0.0723	0.0721	0.0721	0.2591	0.2579	0.2583	2.3782	2.3697	2.3754

PROOFS

Proof of Proposition 1.

To outline the proof of Proposition 1, we use an algorithm that is similar to that applied to analyze the classical log EL ratio (Vexler et al. 2014). We first note that by virtue of the Taylor theorem we can expand $\log(ELR(\beta_r))$ around $\hat{b}_r$ as

$$\log(ELR(\beta_r)) = \log\left(ELR\left(\hat{b}_r\right)\right) + \left(\beta_r - \hat{b}_r\right) \left.\frac{d\log(ELR(\beta_r))}{d\beta_r}\right|_{\beta_r=\hat{b}_r} + \frac{1}{2}\left(\beta_r - \hat{b}_r\right)^2 \left.\frac{d^2\log(ELR(\beta_r))}{d\beta_r^{\ 2}}\right|_{\beta_r=\hat{b}_r} + R, \tag{A.0}$$

where $\hat{b}_r$ is the maximum EL estimator of β_r having the form

$$\hat{b}_r = \sum\nolimits_{i=1}^{n} X_{(i)} \left(\left(\frac{i}{n}\right)^{r+1} - \left(\frac{i-1}{n}\right)^{r+1}\right) / (r+1)$$

and

$$R = \frac{1}{6} \left.\frac{d^3\log(ELR(\beta_r))}{d\beta_r^{\ 3}}\right|_{\beta_r=a} \left(\beta_r - \hat{b}_r\right)^3$$

denotes the remainder term with $a = \beta_r + \varpi(\hat{b}_r - \beta_r)$ and $\varpi \in (0,1)$.

In the case where the function $\log(ELR(u))$ is considered at the argument $u = \hat{b}_r$, we obtain the maximum of this function, since, by virtue of the definition of the EL, $p_i = 1/n$, $i = 1,\dots n$, satisfy the constraints $\sum\nolimits_{i=1}^{n} p_i = 1$ and $\sum\nolimits_{i=1}^{n} X_{(i)} \left\{(S_i)^{r+1} - (S_{i-1})^{r+1}\right\} / (r+1) = \hat{b}_r$ as well as $p_i = 1/n$,

$i=1,\ldots n$, are arguments of $\max\left(\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1\right)$. Thus, $\log\left(ELR\left(\hat{b}_r\right)\right) = \log\left(EL\left(\hat{b}_r\right)/n^{-n}\right) = 0$ and $d\log(ELR(u))/du\big|_{u=\hat{b}_r} = 0$.

Taking into account the results by David and Nagaraja (2003) and Serfling (1980), we have that $\left(\beta_r - \hat{b}_r\right)^2$ is asymptotically χ^2-distributed. In order to show $-2\log(ELR(\beta_r)) \sim \chi_1^2$, applying (A.0), we will prove that the remainder term $R \to 0$ as $n \to \infty$.To achieve the proof scheme above, the next two lemmas are presented.

Lemma 2. We have

$$\frac{d\log(ELR(\beta_r))}{d\beta_r} = \lambda_2,$$

where λ_2 is the second Lagrange multiplier defined in Section 2.2 and $\lambda_2 = \lambda_2\left(\beta_r\right)$ is a function of β_r.

Proof of Lemma 2.

Note that

$$\log(ELR(\beta_r)) = \sum_{k=1}^{n} \log p_k - \log n^{-n} \ ,$$

where in accordance with the definition of *ELR* in Section 2.2 and Eq.(4), $ELR(\beta_r) = \Pi_{k=1}^{n} p_k / n^n$, $p_k = \left(n + \lambda_2 V_k\right)^{-1}$ with $V_k = X_{(k)} S_k^r + \sum_{i=k+1}^{n} X_{(i)}\left(S_i^r - S_{i-1}^r\right) - (r+1)\beta_r$, for $k = 1,\ldots n$.

Thus

$$\frac{d\log(ELR(\beta_r))}{d\beta_r} = \sum_{k=1}^{n} \frac{1}{p_k}\frac{dp_k}{d\beta_r} = -\sum_{k=1}^{n} p_k\left(V_k \frac{d\lambda_2}{d\beta_r} + \lambda_2 \frac{dV_k}{d\beta_r}\right) \qquad \text{(A.1)}$$

To prove Lemma 2, we will show that

$$\sum_{k=1}^{n}\left(p_k V_k \frac{d\lambda_2(\beta_r)}{d\beta_r}\right) = 0 \text{ and } -\sum_{k=1}^{n} p_k \frac{dV_k}{d\beta_r} = 1.$$

We have

$$\sum_{k=1}^{n}\left(p_k V_k \frac{d\lambda_2(\beta_r)}{d\beta_r}\right) = \left[\sum_{k=1}^{n} p_k X_{(k)} S_k^r + \sum_{k=1}^{n} p_k \left(\sum_{i=k+1}^{n} X_{(i)}\left(S_i^r - S_{i-1}^r\right)\right) - (r+1)\beta_r\right]\frac{d\lambda_2}{d\beta_r}.$$

By virtue of Eq. (3), one can easily show that

$$\sum_{k=1}^{n} p_k X_{(k)} S_k^r + \sum_{k=1}^{n} p_k \left(\sum_{i=k+1}^{n} X_{(i)}\left(S_i^r - S_{i-1}^r\right)\right) = (r+1)\beta_r .$$

The two equations above conclude that

$$\sum_{k=1}^{n}\left(p_k V_k \frac{d\lambda_2(\beta_r)}{d\beta_r}\right) = 0 .$$

The definition of V_k leads to

$$\frac{dV_k}{d\beta_r} = rX_{(k)} S_k^{r-1}\frac{dS_k}{d\beta_r} + r\sum_{i=k+1}^{n} X_{(i)}\left(S_i^{r-1}\frac{dS_i}{d\beta_r} - S_{i-1}^{r-1}\frac{dS_{i-1}}{d\beta_r}\right) - (r+1),$$

and then

$$\sum_{k=1}^{n} p_k \frac{dV_k}{d\beta_r} = r\sum_{k=1}^{n} p_k X_{(k)} S_k^{r-1}\frac{dS_k}{d\beta_r} + r\sum_{k=1}^{n} p_k \sum_{i=k+1}^{n} X_{(i)}\left(S_i^{r-1}\frac{dS_i}{d\beta_r} - S_{i-1}^{r-1}\frac{dS_{i-1}}{d\beta_r}\right) - (r+1) \ . \quad \text{(A.2)}$$

Thus we need to show that

$$\sum_{k=1}^{n} p_k X_{(k)} S_k^{r-1}\frac{dS_k}{d\beta_r} + \sum_{k=1}^{n} p_k \sum_{i=k+1}^{n} X_{(i)}\left(S_i^{r-1}\frac{dS_i}{d\beta_r} - S_{i-1}^{r-1}\frac{dS_{i-1}}{d\beta_r}\right) = 1,$$

in order to conclude that $-\sum_{k=1}^{n} p_k \left(dV_k / d\beta_r\right) = 1$.

To this end, we apply Lemma 1 to rewrite the second term on right side of (A.2) in the form

$$\sum_{k=1}^{n} p_k \sum_{i=k+1}^{n} X_{(i)} \left(S_i^{r-1} \frac{dS_i}{d\beta_r} - S_{i-1}^{r-1} \frac{dS_{i-1}}{d\beta_r} \right) = \sum_{i=2}^{n} X_{(i)} \left(S_i^{r-1} \frac{dS_i}{d\beta_r} - S_{i-1}^{r-1} \frac{dS_{i-1}}{d\beta_r} \right) \sum_{j=1}^{i-1} p_j$$
$$= \sum_{i=2}^{n} X_{(i)} \left(S_i^{r-1} \frac{dS_i}{d\beta_r} - S_{i-1}^{r-1} \frac{dS_{i-1}}{d\beta_r} \right) S_{i-1}$$
$$= \sum_{i=2}^{n} \left(X_{(i)} S_i^{r-1} \frac{dS_i}{d\beta_r} \left(S_i - p_i \right) - X_{(i)} S_{i-1}^{r-1} \frac{dS_{i-1}}{d\beta_r} \right)$$
$$= \sum_{i=2}^{n} X_{(i)} S_i^{r} \frac{dS_i}{d\beta_r} - \sum_{i=2}^{n} X_{(i)} S_i^{r-1} \frac{dS_i}{d\beta_r} p_i - \sum_{i=2}^{n} X_{(i)} S_i^{r} \frac{dS_{i-1}}{d\beta_r}.$$

Applying this result to (A.2) we obtain

$$\sum_{k=1}^{n} p_k X_{(k)} S_k^{r-1} \frac{dS_k}{d\beta_r} + \sum_{k=1}^{n} p_k \sum_{i=k+1}^{n} X_{(i)} \left(S_i^{r-1} \frac{dS_i}{d\beta_r} - S_{i-1}^{r-1} \frac{dS_{i-1}}{d\beta_r} \right)$$
$$= \sum_{k=1}^{n} p_k X_{(k)} S_k^{r-1} \frac{dS_k}{d\beta_r} + \sum_{i=2}^{n} X_{(i)} S_i^{r} \frac{dS_i}{d\beta_r} - \sum_{i=2}^{n} X_{(i)} S_i^{r-1} \frac{dS_i}{d\beta_r} p_i - \sum_{i=2}^{n} X_{(i)} S_i^{r} \frac{dS_{i-1}}{d\beta_r}$$

$$= \sum_{i=1}^{n} X_{(i)} \left(S_i^{r} \frac{dS_i}{d\beta_r} - S_i^{r} \frac{dS_{i-1}}{d\beta_r} \right). \quad \text{(A.3)}$$

The constraint used in the definition of the EL

$$\beta_r = \sum_{i=1}^{n} X_{(i)} \left(\frac{S_i^{r+1}}{r+1} - \frac{S_{i-1}^{r+1}}{r+1} \right)$$

implies the equation $d\beta_r / d\beta_r = d\left(\sum_{i=1}^{n} X_{(i)} \left(S_i^{r+1} - S_{i-1}^{r+1} \right) / (r+1) \right) / d\beta_r$, i.e.,

$$\sum_{i=1}^{n} X_{(i)} \left(S_i^{r} \frac{dS_i}{d\beta_r} - S_i^{r} \frac{dS_{i-1}}{d\beta_r} \right) = 1.$$

This result and Eq. (A.3) complete the proof of $-\sum_{k=1}^{n} p_k \left(dV_k / d\beta_r \right) = 1$. The proof of Lemma 2 is complete.

The next lemma is presented to evaluate $d^2 \log(ELR(\beta_r)) / d\beta_r^2 = d\lambda_2(\beta_r) / d\beta_r$.

Lemma 3. The function $\lambda_2 = \lambda_2(\beta_r)$ satisfies

$$\frac{d\lambda_2}{d\beta_r} = \frac{\sum_{k=1}^{n} p_k V_k' - \lambda_2 \sum_{k=1}^{n} p_k^2 V_k V_k'}{\sum_{k=1}^{n} p_k^2 V_k^2},$$

where V_k' denotes the derivative $dV_k / d\beta_r$.

Proof of Lemma 3.

By virtue of definition of λ_2 in Section 2.3 and Eq. (4) with $V_k = J_k - (r+1)\beta_r$ for $k = 1,..,n$, we have

$$\sum_{k=1}^{n} \frac{V_k}{n + \lambda_2 V_k} = 0 .$$

Then

$$0 = d\left(\sum_{k=1}^{n} \frac{V_k}{n + \lambda_2 V_k}\right) / d\beta_r = \sum_{k=1}^{n} \frac{V_k'(n + \lambda_2 V_k) - V_k(\lambda_2' V_k + \lambda_2 V_k')}{(n + \lambda_2 V_k)^2}$$

$$= \sum_{k=1}^{n} \frac{V_k'(n + \lambda_2 V_k)}{(n + \lambda_2 V_k)^2} - \sum_{k=1}^{n} \frac{V_k(\lambda_2' V_k + \lambda_2 V_k')}{(n + \lambda_2 V_k)^2} = \sum_{k=1}^{n} p_k V_k' - \lambda_2' \sum_{k=1}^{n} p_k^2 V_k^2 - \lambda_2 \sum_{k=1}^{n} p_k^2 V_k V_k' .$$

The above equation leads to the formula of λ_2' given in Lemma 3. This completes the proof of Lemma 3.

By virtue of Lemmas 2 and 3, we can rewrite (A.0) as

$$\log(ELR(\beta_r)) = \frac{1}{2}(\beta_r - \hat{b}_r)^2 \left.\frac{d\lambda_2(\beta_r)}{d\beta_r}\right|_{\beta_r = \hat{b}_r} + R, \qquad \text{(A.4)}$$

where

$$\left(\left. \frac{d\lambda_2(\beta_r)}{d\beta_r} \right|_{\beta_r = \hat{b}_r} \right)^{-1}$$

$$= \frac{-1}{n} \sum_{k=1}^{n} \left(X_{(k)} \left(\frac{k}{n} \right)^r + \sum_{i=k+1}^{n} X_{(i)} \left(\left(\frac{i}{n} \right)^r - \left(\frac{i-1}{n} \right)^r \right) - \sum_{i=1}^{n} X_{(i)} \left(\left(\frac{i}{n} \right)^{r+1} - \left(\frac{i-1}{n} \right)^{r+1} \right) \right)^2 .$$

In order to evaluate the remainder term *R*, we present the following lemma.

Lemma 4. $\lambda_2(\beta_r) = O_p(n^{2/3})$.

In Section 2.3 and Eq. (4), λ_2 is defined to be a root of

$$\sum_{k=1}^{n} \frac{J_k - (r+1)\beta_r}{n + \lambda_2 (J_k - (r+1)\beta_r)} = 0, \tag{A.5}$$

where $J_k = X_{(k)} S_k^r + \sum_{i=k+1}^{n} X_{(i)} (S_i^r - S_{i-1}^r)$ and denote $L(\lambda_2) = \sum_{k=1}^{n} \frac{J_k - (r+1)\beta_r}{n + \lambda_2 (J_k - (r+1)\beta_r)}$.

Then we have

$$\sqrt{n} L(\lambda_2) = \frac{\sum_{k=1}^{n} (J_k - (r+1)\beta_r)}{\sqrt{n}} - \frac{\lambda_2}{\sqrt{n}} \frac{1}{n} \sum_{k=1}^{n} \frac{(J_k - (r+1)\beta_r)^2}{1 + \lambda_2 n^{-1} (J_k - (r+1)\beta_r)}. \tag{A.6}$$

Employing the following facts,

$X_k = O_p(n^{1/3})$, $J_k = O_p(n^{1/3})$, $k = 1,..,n$, and, $\sum_{i=k}^{n} X_i p_i \le \sum_{i=1}^{n} |X_i| p_i < \infty$ considering the assumption that $E|X|^3 < \infty$ and $p_i, S_i \le 1$, $i = 1,..,n$.

Thus we obtain that $\sqrt{n} L(\lambda_2) \to -\infty$, as $n \to \infty$ and $\sqrt{n} L(\lambda_2) \to \infty$, as $n \to \infty$ when $\lambda_2 = O_p(n^{2/3})$. Thus the proof of Lemma 4 is complete.

To evaluate $\lambda_2' = d\lambda_2(\beta_r)/d\beta_r$, we will use the following scheme. Since $p_k = (n + \lambda_2 V_k)^{-1}$ and $V_k = X_{(k)} S_k^r + \sum_{i=k+1}^{n} X_{(i)} (S_i^r - S_{i-1}^r) - (r+1)\beta_r$, *k*=1,…,*n*, applying $\lambda_2(\beta_r) = O_p(n^{2/3})$ and

$X_{(k)} = O_p\left(n^{1/3}\right)$ to evaluate p_k, we can conclude that $p_k = O_p(n^{-1})$. Consider $V_n' = d\left(\left(X_{(n)} - (r+1)\beta_r\right)\right)/d\beta_r = -(r+1)$. Then $p_n' = dp_n(\beta_r)/d\beta_r = \lambda_2' O_p(n^{-5/3})$. Since

$$V_k - V_{k+1} = \left(X_{(k)} - X_{(k+1)}\right)S_k^r = \left(X_{(k)} - X_{(k+1)}\right)\left(1 - \sum_{i=k+1}^{n} p_i\right),$$

the results above implies

$$\begin{aligned} V_{n-1}' &= \frac{d\left(V_n + (X_{(n-1)} - X_{(n)})S_{n-1}^r\right)}{d\beta_r} = V_n' + r(X_{(n-1)} - X_{(n)})(1-p_n)^{r-1}\left(\sum_{i=n}^{n} p_i'\right) \\ &= O_p(1) + \lambda_2' O_p\left(n^{-4/3}\right), \end{aligned}$$

and then $p_{n-1}' = \lambda_2' \, O_p(n^{-5/3})$. Sequentially

$$\begin{aligned} V_{n-2}' &= \frac{d\left(V_{n-1} + (X_{(n-2)} - X_{(n-1)})S_{n-2}^r\right)}{d\beta_r} = V_{n-1}' + r(X_{(n-2)} - X_{(n-1)})\left(1 - \sum_{i=n-1}^{n} p_i\right)^{r-1}\left(\sum_{i=n-1}^{n} p_i'\right) \\ &= O_p(1) + \lambda_2' O_p\left(n^{-4/3}\right), \end{aligned}$$

and then $p_{n-2}' = \lambda_2' \, O_p(n^{-5/3})$. In this induction type manner one can conclude that

$V_k' = O_p(1) + \lambda_2' O_p\left(n^{-4/3}\right)$, for all k=1,…,n.

Now, using Lemma 3, we represent $\lambda_2' = d\lambda_2(\beta_r)/d\beta_r$ in the form

$$\frac{d\lambda_2}{d\beta_r} = \frac{\sum_{k=1}^{n} p_k V_k' - \lambda_2 \sum_{k=1}^{n} p_k^2 V_k V_k'}{\sum_{k=1}^{n} p_k^2 V_k^2}.$$

Thus applying the above results, $\lambda_2(\beta_r) = O_p\left(n^{2/3}\right)$, $p_k = O_p(n^{-1})$ and $X_{(k)} = O_p\left(n^{1/3}\right)$ to the formula of $\lambda_2' = d\lambda_2(\beta_r)/d\beta_r$, we obtain that $\lambda_2' = O_p(n)$. In a similar manner, one can show that $\lambda_2'' = d\lambda_2'(\beta_r)/d\beta_r = O_p(n)$.

We use the obtained results to analyze the remainder term R in (A.0). In this case we have

$$R = \lambda_2''(a)\left(\beta_r - \hat{b}_r\right)^3/6 = o_n(1)\text{, as } n \to \infty .$$

Taking into account (A.4), we will show that

$$-\left(\beta_r - \hat{b}_r\right)^2 \frac{d\lambda_2(\beta_r)}{d\beta_r}\bigg|_{\beta_r = \hat{b}_r} \xrightarrow{d} \chi_1^2 .$$

To this end, we note that

$$\hat{b}_r = \sum_{i=1}^n X_{(i)}\left(\left(\frac{i}{n}\right)^{r+1} - \left(\frac{i-1}{n}\right)^{r+1}\right)/(r+1).$$

where $\left((i/n)^{r+1} - ((i-1)/n)^{r+1}\right)/(r+1) \cong (i/n)^r/n$.

Thus the proposed estimator $\hat{b}_r$ is asymptotically equal to the estimator $\bar{b}_r = n^{-1}\sum_{i=1}^n X_{(i)}(i/n)^r$ presented in David and Nagaraja (2003, pp.332-333). Following the book material in David and Nagaraja (2003, pp. 332-333) and Serfling (1980, p. 276), we have that the estimator $\bar{b}_r$ is asymptotically normally distributed with mean β_r and a variance that can be estimated in the form

$$\hat{\sigma}^2(\bar{b}_r)$$
$$= \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\left(\frac{i-1}{n}\right)^r\left(\frac{j-1}{n}\right)^r\left(\frac{\min(i-1, j-1)}{n} - \left(\frac{i-1}{n}\right)\left(\frac{j-1}{n}\right)\right)(X_{(i)} - X_{(i-1)})(X_{(j)} - X_{(j-1)}).$$

To shorten notations, we denote $\lambda'|_{\beta_r=\hat{b}_r} = d\lambda_2(\beta_r)/d\beta_r|_{\beta_r=\hat{b}_r}$.

Lemma 5. Let the conditions of Proposition 1 be held, then

$$n\left(-1/\lambda_2'|_{\beta_r=\hat{b}_r} - \hat{\sigma}(\bar{b}_r)\right) \xrightarrow{p} 0\text{, as } n \to \infty .$$

Proof of Lemma 5.

Based on the formula,

$$\left(\frac{d\lambda_2(\beta_r)}{d\beta_r}\bigg|_{\beta_r=\hat{b}_r} \right)^{-1}$$

$$= \frac{-1}{n}\sum_{k=1}^{n}\left(X_{(k)}\left(\frac{k}{n}\right)^{r} + \sum_{i=k+1}^{n} X_{(i)}\left(\left(\frac{i}{n}\right)^{r} - \left(\frac{i-1}{n}\right)^{r}\right) - \sum_{i=1}^{n} X_{(i)}\left(\left(\frac{i}{n}\right)^{r+1} - \left(\frac{i-1}{n}\right)^{r+1}\right)\right)^{2},$$

and,

$$\hat{\sigma}^2\left(\overline{b}_r\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{i-1}{n}\right)^{r}\left(\frac{j-1}{n}\right)^{r}\left(\frac{\min(i-1, j-1)}{n} - \left(\frac{i-1}{n}\right)\left(\frac{j-1}{n}\right)\right)(X_{(i)} - X_{(i-1)})(X_{(j)} - X_{(j-1)}),$$

heavy algebra can be used to show that

$$-1/\lambda_2'\big|_{\beta_r=\hat{b}_r} = \sum_{k=1}^{n} C_{1k} X_{(k)}^{2} + \sum_{k=1}^{n-1}\sum_{i=k+1}^{n} C_{1ki} X_{(k)} X_{(i)} \text{, and}$$

$$\hat{\sigma}^2\left(\overline{b}_r\right) = \sum_{k=1}^{n} C_{2k} X_{(k)}^{2} + \sum_{k=1}^{n-1}\sum_{i=k+1}^{n} C_{2ki} X_{(k)} X_{(i)} ,$$

where

$$C_{1k} = 2(k-n)\frac{1}{n}\left(\frac{k-1}{n}\right)^{1+r}\left(\frac{k}{n}\right)^{r} + k(n-k)\frac{1}{n^2}\left(\frac{k}{n}\right)^{2r} + (n+1-k)\frac{1}{n}\left(\frac{k-1}{n}\right)^{1+2r},$$

$$C_{1ki} = \frac{1}{n^2}\left(k(i-1-n)\left(\frac{i-1}{n}\right)^{r}\left(\frac{k}{n}\right)^{r} + k(n-i)\left(\frac{i}{n}\right)^{r}\left(\frac{k}{n}\right)^{r} + n(i-n)\left(\frac{i}{n}\right)^{r}\left(\frac{k-1}{n}\right)^{1+r} \right.$$
$$\left. + n(n+1-i)\left(\frac{i-1}{n}\right)^{r}\left(\frac{k-1}{n}\right)^{1+r} \right),$$

$$C_{2k} = 2(k-n)\frac{1}{n}\left(\frac{k-1}{n}\right)^{1+r}\left(\frac{k}{n}\right)^{r} + k(n-k)\frac{1}{n^2}\left(\frac{k}{n}\right)^{2r} + (n+1-k)\frac{1}{n}\left(\frac{k-1}{n}\right)^{1+2r},$$

$$C_{2ki} = \frac{1}{n^2}\left(k(i-1-n)\left(\frac{i-1}{n}\right)^r \left(\frac{k}{n}\right)^r + k(n+1-i)\left(\frac{i}{n}\right)^r \left(\frac{k}{n}\right)^r + n(i-n)\left(\frac{i}{n}\right)^r \left(\frac{k-1}{n}\right)^{1+r} \right.$$
$$\left. + n(n+1-i)\left(\frac{i-1}{n}\right)^r \left(\frac{k-1}{n}\right)^{1+r} \right).$$

Based on the above results, it can be easily seen that the two variance formulas $\hat{\sigma}^2(\overline{b}_r)$ and $-1/\lambda_2'\big|_{\beta_r=\hat{b}_r}$ are asymptotically equivalent. This completes the proof of Lemma 5.

Thus, we conclude that

$$-\left(\beta_r - \hat{b}_r\right)^2 \frac{d\lambda_2(\beta_r)}{d\beta_r}\bigg|_{\beta_r=\hat{b}_r} \xrightarrow{d} \chi_1^2 \text{ as } n \to \infty.$$

Remark: Our limited Monte Carlo study showed that the values of $-1/\lambda_2'\big|_{\beta_r=\hat{b}_r}$ and $\hat{\sigma}^2(\overline{b}_r)$ are very close even when the sample size *n* is relatively small. We do not show this numerical study in this manuscript.

Supplementary Tables related to the MC evaluation of the EL inference of the Gini index

Table 1 and 2 show the MC Type I error rates of the tests for $H_0 : G = \theta$, given that θ is the value of the Gini index of the corresponding underlying data distribution. In this study we used that θ= 0.05, 0.28, 0.5, 0.64 are the values of the Gini index corresponding to *Pareto*(1), *Lognormal*(0.5), *Exp*(1) and χ_1^2 underlying distributions, respectively. We repeated 3000 times sampling data with sizes *n*=20, 25 and 100 for each underlying distribution to evaluate the proposed method.

Table 1. The Type I error control of the EL procedure based on the method by Qin et al. (2010). The expected Type I error rate is 0.05 and G denotes the value of the Gini index for corresponding data distributions.

Baseline distribution	n= 20	50	100
Pareto(1) G=0.05	0.999	0.617	0.259
Lognormal(0.5) G=0.28	0.131	0.078	0.061
Exp(1) G=0.50	0.100	0.074	0.054
χ_1^2 G= 0.64	0.092	0.079	0.058

Table 2. The Type I error control of the proposed method. The expected Type I error rate is 0.05 and G denotes the value of the Gini index for corresponding data distributions.

Baseline distribution	n= 20	50	100
Pareto(1) G=0.05	0.191	0.225	0.197
Lognormal(0.5) G=0.28	0.045	0.070	0.057
Exp(1) G=0.50	0.086	0.059	0.051
χ_1^2 G= 0.64	0.092	0.070	0.053

R-code

```
#############################################################
####Here is the R-code to implement the proposed EL test, the classical EL test#####
####and the empirical test for beta r=1 with underlying standard normal distribution##
########################################################################
```

```
#The sample size, r value and library#
n<-50                                             #sample size
r<-1                                              # r value
library(emplik)
#Calculate the true value beta r=1 as M1 assuming underlying normal distribution#
FF<-function(u) u*(pnorm(u,0,1))^r*dnorm(u,0,1)        # x*F^r*density
M1<-integrate(Vectorize(FF),-Inf,Inf)[[1]]              #calculate true beta value
MC<-3000                                  # number of simulations
L<-array()                                #store results of the proposed method #
ELtest<- array()                          #store results of classical EL method #
Btest<-0                                  #store results of the br- based method#
for(mc in 1:MC){
      x<-rnorm(n,0,1)                     # generate random sample
      xs<-sort(x)                         #sort them in order statistics
#####################################################################
###############these two lines are for test in el.test comparison###############
#####################################################################
      xx<- x*(pnorm(x,0,1))^r
      ELtest[mc]<- el.test(xx,mu=c(M1))$'-2LLR'      # ELtest#
#####################################################################
############ calculate the value of test statistic of the EL ratio method ##########
#####################################################################
ssx<-sort(x)
```

```
LLL<-array()
	for( ii in 1:n)  LLL[ii]<-ssx[ii]*(ii/n)^(r)
	P0<-el.test(LLL,mu=M1)$wts/n
	for(kkk in 1:10){
	Sw<-array()                #To store Sw that is Sn ,n=1,...,n, the cumulative probabilities#
	Sw[1]<-P0[1]
	for( ii in 2:n) {
		Sw[ii]<-sum(P0[1:ii])
		LLL[ii]<-ssx[ii]*(Sw[ii]^r-Sw[(ii-1)]^r)
			}
	J0<-array()       #store value for Jk's....k=1,...,n
	J0[n]<-ssx[n]
		for( ii in 1:(n-1)) J0[ii]<-ssx[ii]*Sw[ii]^r+sum(LLL[(ii+1):n])
	M2=M1*(1+r)
	P1<-el.test(J0,mu=M2)$wts/n
	Swn<-array()
	Swn[1]<-P1[1]
		for( ii in 2:n) Swn[ii]<-sum(P1[1:ii])
	zz<-ssx*(Swn)^r
	P2<-el.test(zz,mu=M1)$wts/n
	P0<-P2
	}
Sw<-array()
```

```
Sw[1]<-P1[1]

LLL[1]<-ssx[1]*P1[1]^(r+1)/(r+1)

for( ii in 2:n) {

    Sw[ii]<-sum(P1[1:ii])

    LLL[ii]<-ssx[ii]*(Sw[ii]^(r+1)-Sw[(ii-1)]^(r+1))/(r+1)

    }

print(c(M1,sum(LLL),sum(P1)))      #here M1 should be close to sum (LLL), and sum(P1) #

UU<-array()

UU[1]<-0

for( ii in 2:n) UU[ii]<-ssx[ii]*(Sw[ii]^r-Sw[(ii-1)]^r)

LambdaI<-(1/P1[1]-n)/(ssx[1]*P1[1]^r+sum(UU[2:n])-(r+1)*M1)

PP<-function(u){

    p<-array()              #p vector of pi's

    S<-array()              #S vector of Si's

    S[n]<-1

    p[n]<-1/(n+u*(xs[n]-M1*(r+1)))       #calculate Pn

    SS<-0

    for( i in (n-1):1) {

        S[i]<-S[i+1]-p[i+1]

        SS<-SS+xs[i+1]*(S[i+1]^r-S[i]^r)

        p[i]<-1/(n+u*(xs[i]*S[i]^r+SS-M1*(r+1)))

    }

    ELM<-xs[1]*(S[1]^(r+1))/(r+1)
```

```
        for( i in 2:n) {
        ELM<-ELM+xs[i]*(S[i]^(r+1)-S[i-1]^(r+1))/(r+1)
                }
        return(ELM)          # ELM is the second constraint
        }
        PP<-Vectorize(PP)
        PP1<-function(u) (PP(u)-M1)^2
l0<-optimize(Vectorize(PP1),c(LambdaI-2*abs(LambdaI),LambdaI+2*abs(LambdaI)))$minimum
###########################################################################
############## calculate the value of test statistic of the proposed method########
###########################################################################
        lamb<-l0
        p<-array()
        S<-array()
        S[n]<-1
        p[n]<-1/(n+lamb *(xs[n]-M1*(r+1)))
        SS<-0
        for( i in (n-1):1) {
            S[i]<-S[i+1]-p[i+1]
            SS<-SS+xs[i+1]*(S[i+1]^r-S[i]^r)
            p[i]<-1/(n+lamb*(xs[i]*S[i]^r+SS-M1*(r+1)))
            }
        L[mc]<-(-2*sum(log(p*n)))
```

```
##################################################################################
#######calculate the br-based normality test #######################
#######################################################################
        xs<- sort(x)  #sort x from small#
        ys<-xs
        bhat<- n^(-1)*sum(xs*(rank(xs)/n)^r)
        nvar<- 0
        for (i in 2:n ){
                for (j in 2:n ) {
                        deltax<- xs[i]-xs[i-1]
                        deltay<- ys[j]-ys[j-1]
                nvar<- nvar+((i-1)/n)^r*((j-1)/n)^r*( min(i-1,j-1)/n-(i-1)*(j-1)/n^2 )*deltax*deltay
                                }
                        }
        c31<- bhat-1.96*sqrt(nvar/n)
        c32<- bhat+1.96*sqrt(nvar/n)
        if ((c31<=M1) & (M1<=c32)) Btest<-Btest+1
######################################################################
##########print the results for the three methods in comparison##############
#######################################################################
        print(c(mc,L[mc],mean(1*(abs(L[L!='NaN'])>3.84))))
        print(c(mc,ELtest[mc],mean(1*(ELtest[ELtest!='NaN']>3.84))))
        print(c(mc,1-Btest/mc))
```

}

REFERENCES

David, H. A., and Nagaraja, H. N., 2003. Order statistics. 3rd ed. Wiley Interscience (John Wiley and Sons), Hoboken, NJ.

Serfling, R. J., 1980. Approximation Theorems of Mathematical Statistics, JohnWiley and Sons, New York.

Vexler, A., Tao, G., and Hutson, A. D., 2014. Posterior expectation based on empirical likelihoods. Biometrika, 101, 711-7118.