Statistica Sinica **33** (2023), 1219-1232
doi:https://doi.org/10.5705/ss.202019.0410

PENALIZED JACKKNIFE EMPIRICAL LIKELIHOOD IN HIGH DIMENSIONS

Zhouping Li, Jinfeng Xu, Na Zhao and Wang Zhou

Lanzhou University, City University of Hong Kong,
The University of Hong Kong and National University of Singapore

Abstract: The jackknife empirical likelihood (JEL) is an attractive approach for statistical inferences with nonlinear statistics, such as U-statistics. However, most contemporary problems involve high-dimensional model selection and, thus, the feasibility of this approach in theory and practice remains largely unexplored in situations in which the number of parameters diverges to infinity. In this paper, we propose a penalized JEL method that preserves the main advantages of the JEL and leads to reliable variable selection based on estimating equations with a U-statistic structure in high-dimensional settings. Under certain regularity conditions, we establish the asymptotic theory and oracle property for the JEL and its penalized version when the numbers of estimating equations and parameters increase with the sample size. Simulation studies and a real-data analysis are used to examine the performance of the proposed methods and illustrate its practical utility.

Key words and phrases: Estimating equations, high-dimensional data analysis, jackknife empirical likelihood, penalized likelihood, U-statistics, variable selection.

1. Introduction

Statistical inference based on estimating equations with a U-statistic structure (U-type estimating equations) is common in nonparametric and semiparametric situations, such as quantile and rank regressions (Jin et al. (2003)). Suppose that observations $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random vectors, and the unknown parameters $\theta = (\theta_1, \ldots, \theta_p)^T$ can be estimated by solving the following r ($r \geq p$) estimating equations:

$$U_n(\theta) = \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} h(X_{i_1}, \ldots, X_{i_k}; \theta) = 0, \tag{1.1}$$

where $h(\cdot) = (h_1(\cdot), \ldots, h_r(\cdot))^T$ are symmetric in $X = (X_1, \ldots, X_k)^T$ and satisfy $Eh(X_1, \ldots, X_k; \theta_0) = 0$ with θ_0, which denotes the true value of θ. Here, Σ

Corresponding author: Zhouping Li, School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China. E-mail: lizhp@lzu.edu.cn.

denotes the summation over subsets of k integers $\{i_1, \ldots, i_k\}$ from $\{1, \ldots, n\}$. For many estimation procedures in the literature, the estimator $\hat{\theta}$ is defined formally as the solution of the above U-type estimating equations (1.1), which is known as a U-type estimating problem; see Jin, Ying and Wei (2001); Song and Ma (2010), and Li, Xu and Zhou (2016), among others. In this study, we examine this problem under a high-dimensional setting; that is, p and r diverge with the sample size n. Hence we use p_n and r_n throughout the paper to emphasize the dependence of p and r on n.

High-dimensional data have become ubiquitous in many applications, including microarray data analysis, neuroimaging, and portfolio selection, where the number of parameters or variables, p_n, is very large, usually in thousands or more. Penalized methods are effective in analyzing such data, and various penalty functions have been proposed, including the lasso (Tibshirani (1996)), SCAD (Fan and Li (2001)), adaptive lasso (Zou (2006)), and least squares approximation (Wang and Leng (2007)). Although these methodologies significantly improve inference and variable selection procedures in high-dimensional settings, such as linear regression and generalized linear models, to the best of our knowledge, its feasibility in jackknife empirical likelihood (JEL) inferences with U-type estimating equations (1.1) remains largely unexplored.

The empirical likelihood (EL) method introduced by Owen (1988, 1990) has been studied extensively, and is widely used in the literature to construct confidence regions and test hypotheses. One nice feature of this method is that the confidence intervals and p-values of a test can be easily obtained without estimating the covariance matrix. More details can be found in Owen (2001) and Chen and Van Keilegom (2009). Note that the standard EL method works well when dealing with linear constraints. An effective way of formulating the EL ratio statistic is to use estimating equations, as in Qin and Lawless (1994), although for nonlinear constraints, the EL method is extremely complicated in terms of its computation. To overcome this computational difficulty, Jing, Yuan and Zhou (2009) proposed the JEL, focusing particularly on nonlinear statistics involving U-statistics. Subsequently, Li, Peng and Qi (2011) and Peng (2012) extended the JEL to include general estimating equations. Recently, high-dimensional data analyses incorporating the EL method have attracted increasing attention; see Hjort, McKeague and Van Keilegom (2009); Chen, Peng and Qin (2009); Tang and Leng (2010); Lahiri and Mukhopadhyay (2012); Leng and Tang (2012); Peng, Qi and Wang (2014); Chang, Chen and Chen (2015); Chen et al. (2015); Li, Liu and Liu (2017); Chang, Tang and Wu (2018); Wang, Wu and Zhao (2019); Tang, Yan and Zhao (2020); Chang et al. (2021), and the references therein. Motivated by these developments, when the dimension grows with the sample size n, the simultaneous estimation of the parameters and variable selection using U-type estimating equations (1.1) is of great interest and challenging, both in theory and in terms of computation. Our study yields two main theoretical results.

(1). We prove that the JEL method is efficient when dealing with high-dimensional U-type estimating equations, and provide the corresponding algorithms in this scenario. This extends the scope of JEL methods for U-type estimating equations from fixed dimensions (Li, Xu and Zhou (2016)) to the case of diverging dimensions.

(2). We propose a novel penalized JEL (PJEL) approach based on U-type estimating equations. By choosing a proper penalty function, the approach preserves the main advantages of the JEL and the penalized method, and the resulting estimator possesses good properties, such as the oracle property, correctly selecting the true sparse model with probability tending to one and with optimal efficiency. Furthermore, Wilks' theorem continues to hold when constructing confidence regions and when testing hypotheses.

The rest of the paper is organized as follows. Section 2 presents the JEL method in high-dimensional settings. In Section 3, we describe the PJEL methodology and derive its asymptotic theory. In Sections 4 and 5, we report our simulation results and present a real-data application to assess the finite-sample performance of our method and illustrate its practical utility. All proofs are relegated to the Supplementary Material.

2. JEL with a Diverging Number of Parameters

For U-type estimating problems, the estimator $\hat{\theta}$ of θ_0 is the solution to $U_n^T(\theta) = (U_{n,1}(\theta), \dots, U_{n,r_n}(\theta))^T = 0$, where

$$U_{n,l}(\theta) = \binom{n}{k}^{-1} \sum_{1 \le i_1 < \cdots < i_k \le n} h_l(X_{i_1}, \dots, X_{i_k}; \theta), \quad l = 1, \dots, r_n.$$

The straightforward application of the standard EL method involves many nonlinear constraints, resulting in a heavy computational burden, and thus is not favorable. To overcome this difficulty, Jing, Yuan and Zhou (2009) proposed first obtaining n jackknife pseudo values, and then applying the standard EL to the nonlinear constraints in the U-statistics. They demonstrated that this procedure is particularly efficient in this situation. However, their discussion is restricted to estimating the mean of one-sample and two-sample statistics, where the dimension is fixed. In this paper, we propose a general JEL procedure for inferences for this U-type estimation problem that avoids nonlinear constraints and permits high-dimensional estimating equations. The method is based on the fact that $U_n(\theta)$ is an unbiased and consistent estimator of $Eh(X_1, \ldots, X_k; \theta)$ and has mean zero at the true parameter value θ_0. The details are as follows.

Define $T_n = U_n(\theta)$ and $T_{n-1}^{(-i)} = T(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n; \theta)$, the statistic computed on the original data set, with the ith observation removed. The jackknife pseudo values

$$\widehat{V}_i(\theta) = nT_n - (n-1)T_{n-1}^{(-i)} \tag{2.1}$$

can be shown to be asymptotically independent under mild conditions. Because they also estimate $Eh(X_1, \ldots, X_k; \theta)$ in an unbiased and consistent manner, a standard EL can then be constructed on $\widehat{V}_i(\theta)$, for $i = 1, \ldots, n$, instead of on the original observations $X_1, \ldots, X_n$, as follows. Specifically, the JEL function is defined as

$$\mathcal{L}(\theta) = \max\left\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \widehat{V}_i(\theta) = 0\right\}, \tag{2.2}$$

with the corresponding likelihood ratio

$$\mathcal{R}(\theta) = \max\left\{\prod_{i=1}^{n} (np_i) : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \widehat{V}_i(\theta) = 0\right\}. \tag{2.3}$$

Throughout this paper, $\| \cdot \|$ refers to the l_2-norm $\| \cdot \|_2$ and q is any fixed integer. In the Supplementary Material, we prove the following theorems.

Theorem 1. *Let $\hat{\theta}_E$ be the minimizer of* (2.3). *Under Conditions* C1–C5 *in the Supplementary Material, as $n \to \infty$ and with probability tending to one, we have (i) $\hat{\theta}_E \xrightarrow{p} \theta_0$ and (ii) $\|\hat{\theta}_E - \theta_0\| = O_p(a_n)$, where a_n is given in the Supplementary Material.*

Theorem 2. *Under Conditions* C1–C5 *in the Supplementary Material, we have*

$$A_n \Omega_{22.1}^{1/2} \sqrt{n}(\hat{\theta}_E - \theta_0) \xrightarrow{d} N(0, \Delta),$$

where $A_n \in R^{q \times p_n}$, such that $A_n A_n^T \to G$ and G is a $q \times q$ nonnegative matrix; $\Omega_{22.1}$ is given in the Supplementary Material.

3. PJEL

In this section, we establish the asymptotic theory for the PJEL estimator. Suppose that θ can be partitioned as $\theta = (\theta_1^T, \theta_2^T)^T$, where $\theta_1 \in R^d$ and $\theta_2 \in R^{p_n - d}$ are the nonzero and zero components, respectively. Then, the true parameter value $\theta_0 = (\theta_{10}^T, \mathbf{0})^T$. We estimate the unknown parameter vector θ_0 by minimizing

$$l_p(\theta) = \sum_{i=1}^{n} \log\{1 + \lambda^T \widehat{V}_i(\theta)\} + n \sum_{j=1}^{p_n} p_\tau(|\theta_j|), \tag{3.1}$$

where $p_\tau(|\theta_j|)$ is some penalty function, and τ is a tuning parameter that controls the trade-off between the bias and the model complexity (see Fan and Li (2001)). Then, the PJEL has the following oracle property.

Theorem 3. *Let* $\hat{\theta} = (\hat{\theta}_1^T, \hat{\theta}_2^T)^T$ *be the minimizer of* (3.1). *Under Conditions* C1–C7 *in the Supplementary Material, as* $n \to \infty$, *we have*

(i) $P(\hat{\theta}_2 = 0) \to 1$;

(ii) $n^{1/2} B_n \Delta_{11.2}^{-1/2} (\hat{\theta}_1 - \theta_{10}) \xrightarrow{d} N(0, G^*)$, *where* B_n *is a* $q \times d$ *matrix, such that* $B_n B_n^T \to G^*$, G^* *is a* $q \times q$ *nonnegative symmetric matrix, and* $\Delta_{11.2}$ *is given in the Supplementary Material.*

Because the number of parameters may diverge to infinity, we next study the problem of testing the following linear hypothesis:

$$H_0 : L_n \theta_{10} = 0, \quad v.s. \quad H_1 : L_n \theta_{10} \neq 0,$$

where L_n is a $q \times d$ matrix, such that $L_n L_n^T = I_q$ for fixed q. Note that this type of hypothesis includes simultaneously testing whether a few variables are statistically significant. In the penalized likelihood context, Fan and Peng (2004) investigated this type of hypothesis testing in a parametric likelihood framework, and Leng and Tang (2012) considered the problem in a standard EL setting for general estimating equations. Our results further generalize their results. Specifically, a PJEL ratio test statistic is defined as

$$\tilde{l}(L_n) = -2\{l_p(\hat{\theta}) - \min_{\theta_1, L_n \theta_1 = 0} l_p(\theta)\}. \tag{3.2}$$

The following theorem derives the asymptotic null distribution of the above test statistic.

Theorem 4. *Under the null hypothesis and Conditions* C1–C7 *in the Supplementary Material, we have*

$$\tilde{l}(L_n) \xrightarrow{d} \chi_q^2, \qquad \text{as } n \to \infty.$$

From Theorem 4, a $(1-\alpha)$-level confidence region for $L_n\theta_1$ can be constructed as

$$C_\alpha = \{v : -2\{l_p(\hat{\theta}) - \min_{\theta_1, L_n\theta_1 = v} l_p(\theta)\} \le \chi^2_{q,1-\alpha}\},$$

where $\chi^2_{q,1-\alpha}$ is the $(1-\alpha)$th quantile of the χ_q^2 distribution.

Finally, the algorithms for implementing the proposed PJEL method and constructing the related confidence region are summarized as follows.

Algorithm 1 Algorithm of PJEL

.

Input $\{X_i, Y_i\}$. Suppose γ and ϵ_0 are two predefined small numbers, for example, $\gamma = 10^{-4}$.
For fixed θ, define $\lambda(\theta)$ that minimizes $l_p(\theta; \lambda)$;
Let $k = 0$, and initialize $\theta^{(0)}$ using a general PIM estimator (e.g., probabilistic index model in the simulations, etc.).
repeat
  Let $\Theta_k = \{j : \theta_j^{(k)} \neq 0\}$, and find θ such that

$$\theta = \operatorname*{argmin}_{\theta, \theta_{\Theta_k^c} = 0} l_p(\theta; \lambda(\theta)),$$

  where Θ_k^c is the complimentary set of Θ_k and $\theta_{\mathcal{A}} = \{\theta_j, j \in \mathcal{A}\}$.
  for $j \in \Theta_k$ **do**
    if $|\theta_j| < \gamma$ **then**
      Take $\theta_j^{(k+1)} = 0$,
    else
      Take $\theta_j^{(k+1)} = \theta_j$;
    end if
  end for
  Calculate $L_{k+1} = l_p(\theta^{(k+1)}; \lambda(\theta^{(k+1)}))$;
  $k = k + 1$;
until $\max_{j \in \Theta_{k-1}} |\theta_j^{(k-1)} - \theta_j^{(k)}| < \epsilon_0$.

---

4. Simulation Studies

In this section, we use the probabilistic index model (PIM) (Thas et al. (2012)) as an example to illustrate the proposed method for U-structured problems. The effect of the covariates X on the response Y is evaluated using the

Algorithm 2 Construct confidence region by PJEL

Input $\{X_i, Y_i\}$, and the matrix $L_n \in R^{q\times d}$ for the hypothesis testing.
Let $\hat{\theta}$ be the optimal estimator of θ_0 by Algorithm 1, and calculate the corresponding $l_p(\hat{\theta};\lambda(\hat{\theta}))$. $\mathcal{C}$ is a set of grid points $\{c_j, j=1,\ldots,K\}$ with some predefined constant $K>0$, where c_j are evenly spaced in the interval $C_0=[L_n\hat{\theta}-c_0, L_n\hat{\theta}+c_0]$. Here c_0 is some predefined positive constant that ensures C_0 can cover the $(1-\alpha)$th-level confidence interval.
for j from 1 to K **do**
  Fix θ_{10} such that $L_n\theta_{10}=c_j$, obtain the optimal estimator $\tilde{\theta}^T=(\theta_{10},\theta_{20})^T$ that minimizes $l_p(\tilde{\theta};\lambda(\tilde{\theta}))$ by Algorithm 1, and obtain $l_p(\tilde{\theta};\lambda(\tilde{\theta}))$;
  if $-2(l_p(\hat{\theta};\lambda(\hat{\theta}))-l_p(\tilde{\theta});\lambda(\tilde{\theta}))\le\chi^2_{q,1-\alpha}$; **then**
    Add c_j into Θ_0;
  end if
end for
Construct the $(1-\alpha)$th-level confidence region for $L_n\theta_{10}$: $\mathcal{C}_\alpha=[c_l,c_u]$, where c_l and c_u are the minimum and maximum of $\mathcal{C}$, respectively.

probabilistic index, which is defined as the probability $\mathrm{P}(Y_i \preccurlyeq Y_j) := \mathrm{P}(Y_i<Y_j)+0.5\mathrm{P}(Y_i=Y_j)$, where Y_i and Y_j are independent response variables with an identical distribution function F. The data consist of i.i.d. observations (Y_i, X_i), for $i=1,\ldots,n$. The PIM is defined as

$$\mathrm{P}(Y_i \preccurlyeq Y_j|X_i,X_j)=m(X_i,X_j;\beta)=g^{-1}\left(Z_{ij}^T\beta\right),$$

where $g(\cdot)$ is a link function and Z_{ij} depends on X_i and X_j. Following (Thas et al. (2012)), let $Z_{i,j}=X_j-X_i$. For the probit and logit link functions, the models are referred to as the linear PIM and the Cox PIM, respectively. De Neve and Thas (2015) proposed the following U-type estimating equations:

$$\sum_{(i,j)} U_{ij}(\beta)=\mathbf{0}, \quad U_{ij}(\beta):=Z_{ij}\left[\mathrm{I}(Y_i \preccurlyeq Y_j)-g^{-1}\left(Z_{ij}^T\beta\right)\right].$$

As in Thas et al. (2012), we consider two scenarios.

(a) *Normal linear model*: $Y_i|X_i$ are i.i.d. $N(\alpha_1 X_{1i}+\alpha_2 X_{2i},1)$. In this setting, the corresponding PIM is given by $\Phi^{-1}\{P(Y_i \preccurlyeq Y_j|X_i,X_j)\}=\beta_1(X_{1j}-X_{1i})+\beta_2(X_{2j}-X_{2i})$, where $\beta_i=\alpha_i/\sqrt{2}$ and Φ is the distribution function of the standard normal distribution.

(b) *Exponential model*: $Y_i|X_i$ are i.i.d. Exponential $\{\exp(\alpha_1 X_{1i}+\alpha_2 X_{2i})\}$. In this setting, the corresponding PIM is given by logit $\{P(Y_i \preccurlyeq Y_j|X_i,X_j)\}=\beta_1(X_{1j}-X_{1i})+\beta_2(X_{2j}-X_{2i})$, where $\beta_i=-\alpha_i$.

In the above settings, the covariate X_1 is a Bernoulli random variable with success probability 0.5, X_2 follows $U[0,10]$, $\alpha_1 = 1$, and $\alpha_2 = 1$. We conducted simulations for various combinations of p and n, with 1,000 repetitions.

The tuning parameters τ are taken from a fine grid and chosen using the BIC-type criterion (Wang, Li and Leng (2009))

$$BIC(\tau) = 2\ell_p(\beta_\tau) + C_n \log n df_\tau,$$

where β_τ is the PJEL estimator with tuning parameter τ, and df_τ is the number of nonzero coefficients in β_τ. For fixed p, $C_n = 1$ and for diverging p, $C_n = \max\{\log\log p, 1\}$. In our numerical study, we find that the selected tuning parameter for BIC decreases as the sample size increases, whereas the remaining parameters, such as the dimension of the covariates and the value of β stay the same. Furthermore, the ratio of λ to the square root of p/n increases with the sample size (not shown in the tables). These results are consistent with the regularized condition in our assumptions, which validates the BIC-type criterion used here.

We also compare the proposed method with the penalized empirical likelihood (PEL) introduced by Tang and Leng (2010), and the penalized maximum smoother rank correlation (PMSRC) introduced by Lin and Peng (2013), using the same data sets and tuning criteria. The method of Tang and Leng (2010) considers only linear constraints. Therefore, we first use the sequential linearization method of Wood, Do and Broom (1996) to linearize the nonlinear constraints, and then apply the PEL to it. The latter method achieves the estimation using a penalized smoothing objective function for the maximum rank correlation, which is a typical U-statistic. The simulation results are summarized in Table 1 and Table 2, which report the performance of the proposed method in terms of the median of the L_2 distance (MD) of the estimation, the average of the correctly selected zero coefficients (C), and the average of the incorrectly estimated zero coefficients (IC) for variable selection. In both scenarios, the PJEL method performs well in terms of the estimation and variable selection. Moreover, it identifies most of the true zero coefficients as zero, whereas for the PEL, the performance is not good enough for the variable selection. For the PMSRC, although the estimation is quite good for the linear PIM model, for the Cox model, the performance is not as good as that of the PJEL. Moreover, the variable selection capability is not comparable with that of the PJEL. In addition, in order to illustrate the computational improvement of including the jackknife method in the EL model, we report the average computational time of the PJEL and PEL in the above setting

Table 1. Simulation results for the linear PIM model with the PJEL, PEL, and PMSRC methods.

		PJEL			PEL			PMSRC		
n	p	MD	C	IC	MD	C	IC	MD	C	IC
200	5	0.0548	2.936	0.000	0.0527	2.346	0.000	0.0354	2.442	0.000
300	10	0.0418	7.934	0.000	0.0392	5.878	0.000	0.0295	6.763	0.000
400	15	0.0378	12.918	0.000	0.0366	9.395	0.000	0.0207	11.275	0.000

Table 2. Simulation results for the Cox PIM model with the PJEL, PEL, and PMSRC methods.

		PJEL			PEL			PMSRC		
n	p	MD	C	IC	MD	C	IC	MD	C	IC
200	5	0.0647	2.930	0.053	0.0850	2.278	0.000	0.1252	2.112	0.000
300	10	0.0551	7.905	0.000	0.0548	6.598	0.000	0.1058	6.245	0.000
400	15	0.0446	12.904	0.000	0.0460	9.732	0.000	0.0919	7.246	0.000

Table 3. The average time (in seconds) for the PJEL and PEL with a deterministic sample size n and covariates p.

n	p	Linear PIM		Cox PIM	
		PJEL	PEL	PJEL	PEL
200	5	4.140	16.394	5.236	34.895
300	10	18.093	115.586	40.399	344.050
400	15	58.542	451.907	191.416	1,692.695

for the simulation shown in Table 3. Here, we find that in both the linear and the Cox PIM models, the PEL takes much longer than the PJEL, and when the sample size or the dimension of the covariates grows, the ratio of the time expenditure between these two methods increases. Thus, when we have a large sample size data set and U-statistic structure estimating equations with high dimension covariates, the PEL is not applicable, owing to the computational burden. This reveals the advantage of the PJEL method over the PEL method in terms of the computational efficiency.

Next, we illustrate the performance of the PJEL when constructing the confidence region. We set L_n in (3.2) to $(1, 0, \ldots, 0)$, and test the null hypothesis $H_0 : \beta_{10} = a$, with $a = \beta_1 - 0.2, \beta_1 - 0.1, \beta_1, \beta_1 + 0.1$, and $\beta_1 + 0.2$ separately, where β_{10} denotes the first component of β. Under the nominal level $\alpha_0 = 0.05$, we summarize the empirical size for the deterministic value of a in Table 4. The table shows that the size of our test is close to the nominal level, and increases when the null value differs from the true value β_1. These results validate the likelihood ratio test under the PJEL, and indicates the feasibility of constructing

Table 4. The empirical percentages that a given value does not fall in the 95% confidence interval.

PIM model	n	p	$\beta_{10}-0.2$	$\beta_{10}-0.1$	β_{10}	$\beta_{10}+0.1$	$\beta_{10}+0.2$
Normal linear	200	5	24.40	11.20	4.60	10.00	26.10
	300	10	40.40	15.80	6.40	12.00	33.80
	400	15	54.70	22.50	4.70	12.20	40.70
Exponential	200	5	26.90	14.10	6.50	9.10	16.80
	300	10	31.70	15.60	6.30	10.10	21.40
	400	15	41.60	20.00	6.90	10.60	26.70

Table 5. The 15 variables in the air pollution study.

PREC	Average annual precipitation in inches
JANT	Average January temperature in degrees F
JULT	Average July temperature in degrees F
OVR65	Percent of 1960 SMSA population aged 65 or older
POPN	Average household size
EDUC	Median school years completed by those over 22
HOUS	Percent of housing units which are sound & with all facilities
DENS	Population per sq. mile in urbanized areas, 1960
NONW	Percent of non-white population in urbanized areas, 1960
WWDRK	Percent of employed in white collar occupations
POOR	Percent of families with income less than $3,000
HC	Relative hydrocarbon pollution potential
NOX	Relative nitric oxides pollution potential
SO_2	Relative sulphur dioxide pollution potential
HUMID	Annual average percent of relative humidity at 1pm

the confidence region by taking the hypothesis at finer grid points.

5. An Application to an Air Pollution Study

The adverse effects of air pollution on human health have been explored in many scientific studies, particularly in terms of the effect on human mortality. Here, we apply the proposed method to an air pollution study[1] (McDonald and Schwing (1973)) to identify the factors associated with air that most affect mortality. The data set consists of 60 observations and 15 features. The response variable is “Mortality”, representing the total age-adjusted mortality rate per 100,000. Detailed descriptions of the 15 variables in the air pollution study are given in Table 5. McDonald and Schwing (1973) proposed using a multiple linear regression to assess the covariate effects on the mortality rate. Here, we apply the proposed method to fit a normal linear PIM to identify those factors that

[1] Data source: `http://lib.stat.cmu.edu/datasets/pollution`

Table 6. The 95% confidence intervals of the estimated nonzero coefficients for the air pollution study.

Covariate	Estimated coefficient	Confidence interval by PJEL
PREC	0.731	[0.361, 1.203]
JANT	-1.113	[-1.472, -0.822]
EDUC	-0.390	[-0.712, -0.161]
NONW	1.743	[1.36, 2.23]
HC	-3.542	[-5.612, -2.65]
NOX	3.956	[2.096, 5.89]

are relevant to mortality. The BIC-type criterion is used to choose the parameter λ. Then, using the PJEL method, six covariates are identified as relevant, and are reported in Table 6, along with the constructed 95% confidence intervals, using the PJEL procedures for the inference. The identified covariates are as follows: *average annual precipitation (PREC), average January temperature (JANT), median school years completed by those over 22 (EDUC), non-white percentage (NONW), hydrocarbon pollution potential (HC), and nitric oxide pollution potential (NOX).* The variables selected here are similar to those selected by McDonald and Schwing (1973). They used two different criteria to choose the factors. When using the ridge regression, we identify four features in common with their six (annual precipitation, January temperature, education, and non-white percentage). In addition, we find that *hydrocarbon* has adverse effect on the human mortality rate.

Supplementary Material

The online Supplementary Material includes detailed proofs of the main theorems.

Acknowledgments

The authors sincerely thank the editor, associate editor, and two anonymous reviewers for their very helpful and constructive comments. Zhouping Li's research was supported by the National Natural Science Foundation of China (11571154). Jinfeng Xu and Na Zhao's research was supported by the National Natural Science Foundation of China (72033002), University of Hong Kong-Zhejiang Institute of Research and Innovation Seed Fund, and General Research Fund (17306619, 17308018 and 17308820) of Hong Kong. Wang Zhou's research was partially supported by grant R-155-000-192-114 from the National University of Singapore.

References

Chang, J., Chen, S. and Chen, X. (2015). High dimensional generalized empirical likelihood for moment restrictions with dependent data. *Journal of Econometrics* **185**, 283–304.

Chang, J., Chen, S., Tang, C. and Wu, T. (2021). High-dimensional empirical likelihood inference. *Biometrika* **108**, 127–147.

Chang, J., Tang, C. and Wu, T. (2018). A new scope of penalized empirical likelihood with high-dimensional estimating equations. *The Annals of Statistics* **46**, 3185–3216.

Chen, B., Pan, G., Yang, Q. and Zhou, W. (2015). Large dimensional empirical likelihood. *Statisitica Sinica* **25**, 1659–1677.

Chen, S. X., Peng, L. and Qin, Y. L. (2009). Effects of data dimension on empirical likelihood. *Biometrika* **96**, 711–722.

Chen, S. X. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regression (with Discussion). *Test* **18**, 415–447.

De Neve, J. and Thas, O. (2015). A regression framework for rank tests based on the probabilistic index model. *Journal of the American Statistical Association* **110**, 1276–1283.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.

Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics* **32**, 928–961.

Hjort, N. L., McKeague, I. and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *The Annals of Statistics* **37**, 1079–1111.

Jin, Z., Ying, Z. and Wei, L. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381–390.

Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time models. *Biometrika* **90**, 341–353.

Jing, B.-Y., Yuan, J. and Zhou, W. (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association* **104**, 1224–1232.

Lahiri, S. N. and Mukhopadhyay, S. (2012). A penalized empirical likelihood method in high dimensions. *The Annals of Statistics* **40**, 2511–2540.

Leng, C. and Tang, C. (2012). Penalized empirical likelihood and growing dimensional general estimating equations. *Biometrika* **99**, 703–716.

Li, M., Peng, L. and Qi, Y. (2011). Reduce computation in profile empirical likelihood method. *The Canadian Journal of Statistics* **39**, 370–384.

Li, Z., Liu, Y. and Liu, Z. (2017). Empirical likelihood and general relative error criterion with divergent dimension. *Statistics* **51**, 1006–1022.

Li, Z., Xu, J. and Zhou, W. (2016). On nonsmooth estimating functions via jackknife empirical likelihood. *Scandinavian Journal of Statistics* **43**, 49–69.

Lin, H. and Peng, H. (2013). Smoothed rank correlation of the linear transformation regression model. *Computational Statistics and Data Analysis* **57**, 615–630.

McDonald, G. C. and Schwing, R. C. (1973). Instabilities of regression estimates relating air pollution to mortality. *Technometrics* **15**, 463–481.

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

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics* **18**, 90–120.

Owen, A. B. (2001). *Empirical Likelihood.* Chapman and Hall, London.

Peng, L. (2012). Approximate jackknife empirical likelihood method for estimating equations. *The Canadian Journal of Statistics* **40**, 110–123.

Peng, L., Qi, Y. and Wang, R. (2014). Empirical likelihood test for high dimensional linear models. *Statistics & Probability Letters* **86**, 85–90.

Qin, J. and Lawless, J. F. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics* **22**, 300–325.

Song, X. and Ma, S. (2010). Penalized variable selection with U-estimates. *Journal of Nonparametric Statistics* **22**, 499–515.

Tang, C. and Leng, C. (2010). Penalized high-dimensional empirical likelihood. *Biometrika* **97**, 905–919.

Tang, N., Yan, X. and Zhao, X. (2020). Penalized generalized empirical likelihood with a diverging number of general estimating equations for censored data. *The Annals of Statistics* **48**, 607–627.

Thas, O., Neve, J. D., Clement, L. and Ottoy, J. P. (2012). Probabilistic index models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**, 623–671.

Tibshirani, R. J. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **58**, 267–288.

Wang, H. and Leng, C. (2007). Unified Lasso estimation via least squares approximation. *Journal of the American Statistical Association* **102**, 1039–1048.

Wang, H., Li, B. and Leng, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**, 671–683.

Wang, D., Wu, T. and Zhao, Y. (2019). Penalized empirical likelihood for the sparse Cox regression model. *Journal of Statistical Planning and Inference* **201**, 71–85.

Wood, A. T. A., Do, K. A. and Broom, N. M. (1996). Sequential linearization of empirical likelihood constraints with application to U-statistics. *Journal of Computational and Graphical Statistics* **5**, 365–385.

Zou, H. (2006). The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association* **101**, 1418–1429.

Zhouping Li

School of Mathematics and Statistics, Center for Data Science, Lanzhou University, Lanzhou 730000, China.

E-mail: lizhp@lzu.edu.cn

Jinfeng Xu

Department of Biostatistics, City University of Hong Kong, Hong Kong, China.

E-mail: jinfenxu@cityu.edu.hk

Na Zhao

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China.

E-mail: zhaona@connect.hku.hk

Wang Zhou
Department of Statistics and Data Science, National University of Singapore, Singapore.
E-mail: stazw@nus.edu.sg

(Received October 2019; accepted August 2021)