Journal of Data Science and Modeling, Vol. 2, No. 2, 103-118, June 2024

Research Manuscript

A Comparative Study of RPEL and JPEL for Parameter Estimation

Mahdieh Bayati*[1],

[1]Department of Statistics, University of Qom, Qom, Iran

Recieved: 10/11/2024 Accepted: 12/03/2025

Abstract: This study generalizes the joint empirical likelihood (JEL), which is named the joint penalized empirical likelihood (JPEL), and presents a comparative analysis of two innovative empirical likelihood methods: the restricted penalized empirical likelihood (RPEL) and the joint penalized empirical likelihood. These methods extend traditional empirical likelihood approaches by integrating criteria based on the minimum variance and unbiasedness of the estimator equations. In RPEL, estimators are obtained under these two criteria, while JPEL facilitates the joint application of the estimator equations used in RPEL, allowing for broader applicability.

We evaluate the effectiveness of RPEL and JPEL in regression models through simulation studies and evaluate the performance of RPEL and JPEL, focusing on parameter accuracy, model selection (as measured by the Empirical Bayesian Information Criterion), predictive accuracy (Mean Squared Error), and robustness to outliers. Results indicate that RPEL consistently outperforms JPEL across all criteria, with RPEL yielding simpler models and more reliable estimates, particularly as sample sizes increase. These findings suggest that RPEL provides greater stability and interpretability for regression models, making it a superior choice over JPEL for the scenarios tested in this study.

Keywords: Empirical likelihood, restricted penalized empirical likelihood, restricted joint empirical likelihood, estimating equations, Gibbs sampling, MCMC.

Mathematics Subject Classification (2010): 62F10, 62J05, 62J12, 65C05.

*Corresponding Author: Bayati_2006@hotmail.com

1. Introduction

Empirical likelihood (EL) is a non-parametric statistical method that has evolved over several decades. Its development reflects a growing need for flexible statistical techniques that do not rely heavily on strict parametric assumptions. The concept of EL was first introduced by Owen (1988), who proposed a method for constructing confidence regions based on the empirical distribution function. This approach allows statisticians to perform inference without assuming a specific underlying distribution for the data. Initially, the method was developed to estimate confidence intervals and regions for parameters based on a single sample.

Today, many statisticians use this method to analyze various real-world data. Owen (1990) further advanced EL by pioneering the EL ratio statistics, applying them to various non-parametric problems and demonstrating that these statistics asymptotically follow a chi-squared distribution. He also developed confidence intervals and hypothesis tests for model parameters based on likelihood ratio statistics within a parametric framework. Subsequent work by DiCiccio, and Romano (1989) and Hall and La Scala (1990) addressed these statistics' asymptotic properties and essential corrections.

Qin and Lawless (1994) established that EL, in conjunction with appropriate estimating equations, offers a valid non-parametric fit for data. In the EL framework, parameter estimates are derived by maximizing the EL function subject to estimating equations and additional restrictions, notably the zero expectation value of these equations under the probability model $p_1, p_2, ..., p_n$.

Further contributions by Newey and Smith (2004) and Chen et al. (2009) demonstrated appealing statistical properties of the estimators obtained through this method. However, studies by Chen et al. (2009), Hjort et al. (2009), Tang and Leng (2010), and Leng and Tang (2012) indicated that conventional asymptotic results for EL estimators hold only when the dimension p of the parameters and the number r of estimating equations grow at a rate slower than the sample size n. In contrast, challenges arise in high-dimensional contexts where both p and r increase with n.

To address these challenges, Tang and Leng (2010), Leng and Tang (2012), and Chang et al. (2015) utilized sparsity assumptions and penalty functions to achieve parameter sparsity, showing that sparse parameter estimators with desirable properties are attainable. Nonetheless, issues persist in high-dimensional data analysis using EL.

Taso (2004) identified an under-coverage problem, where true parameter values are often not contained within EL-based confidence regions at the nominal level when n is fixed and p is relatively large. Tsao and Wu (2014) proposed an extended EL approach to mitigate this under-coverage by imposing additional constraints on the parameter space. Meanwhile, Bartolucci, F. (2007) introduced a penalized EL method that optimizes the product of probabilities penalized by a loss function related to model parameters. Lahiri and Mukhopadhyay (2012) conducted similar work with a different type of loss function, exploring the properties of the penalized EL ratio statistic within a high-dimensional framework. Chang et al. (2017) further investigated the penalized EL estimator in the context of high-dimensional sparse model parameters, addressing new challenges in this evolving field.

Bayati et al. (2021) introduced Restricted Empirical Likelihood (REL) and Restricted Penalized Empirical Likelihood (RPEL) estimators. These estimators are derived under the criteria of unbiasedness and minimum variance for estimating equations. As a result, they exhibit desirable properties, particularly demonstrating greater robustness against outliers compared to some existing estimators. Bayati et al. (2021) applied this method in autoregressive models.

Shantia and Ghoreishi (2024) presented the Restricted Joint Empirical Likelihood (RJEL) method for semi-parametric hierarchical models, focusing on the effect of distribution dispersion at the second level. Their simulation studies highlighted the effectiveness of shrinkage estimates from RJEL, particularly in cases with outlier data and heavy-tailed distributions. This article applied the penalty function on RJEL and named it JPEL. Then, it compares RPEL and JPEL.

Various studies have also explored recent advancements in penalized empirical likelihood methods. Arslan, and Ozdemir (2023) introduced a robust penalized empirical likelihood approach for linear regression, focusing on enhancing robustness against outliers and model misspecification. Wang et al. (2019) proposed a penalized empirical likelihood method for sparse Cox regression models, demonstrating improved predictor selection accuracy. Additionally, Ji, and Liu (2024) developed a penalized empirical likelihood technique for population size estimation, incorporating a half-normal prior to enhance robustness. These studies provide further insights into applying and developing penalized empirical likelihood techniques in regression analysis.

First, the basic concepts of likelihood were reviewed. This method was introduced by Owen (1988), which combines the reliability of non-parametric methods and the flexibility of parametric methods. Then, restricted penalized empirical likelihood (RPEL) was explained. It showed that RPEL has attractive properties, especially for handling outlier data. The joint penalized empirical likelihood (JPEL) is an adjusted version of RPEL, where the dependence among estimating equations is eliminated. The remainder of this paper is organized as follows. Section 2 provides a description of EL methods. In Section 3, Bayesian analysis for these methods is explained. The EBIC for RPEL and JPEL is introduced in

Section 4. In Section 5, RPEL and JPEL are applied and compared in a multiple regression model. Finally, a simulation study is presented. To validate our proposed methodology, we analyze the Boston Housing Data set using multiple regression techniques, including standard LR, RPEL, and JPEL.

2. Methods of EL

EL uses estimating equations to obtain the best estimates of parameters. The most notable feature of the EL method is that, under mild conditions, the null distribution of the EL ratio statistic follows the standard chi-square distribution, similar to parametric settings. Since its development, the EL method has found extensive application across various fields of statistics. However, the method is primarily designed for independent observations and encounters challenges when applied to dependent data, such as time series. Bayati et al. (2021) addressed this problem using REL and Bayesian techniques, which will be explained in more detail in this article.

Let $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n \in R^d$ be independent and identically distributed observations and $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)^T$ be a p-dimensional parameter in parameter space $\boldsymbol{\Theta}$. In addition, a r-dimensional vector of estimating equations is $\boldsymbol{g(\mathbf{X};\theta) = (g_1(\mathbf{X};\theta), ..., g_r(\mathbf{X};\theta))^T}$, where $g_l(\mathbf{X};\boldsymbol{\theta})$ for $l = 1, ..., r$ are unbiased estimating equations. It is assumed that $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is the true vector of parameters. Thus,

$$E(g_l(\mathbf{X};\boldsymbol{\theta}_0)) = \sum_{i=1}^{n} p_i g_l(\mathbf{x};\boldsymbol{\theta}_0) = 0, \qquad l = 1, ..., r,$$

where $p_1, ..., p_n$ are the corresponding probabilities. More details are provided by Hjort et al. (2009) and Chang et al. (2015). The EL approach estimates the probabilities by maximizing $\prod_{i=1}^{n} p_i$ under unbiased and probability constraints. That is

$$R_E(\boldsymbol{\theta}) = \sup_{p_1,\ldots,p_n} \Big\{ \prod_{i=1}^{n} p_i : p_i > 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i g_l(\mathbf{x}_i;\boldsymbol{\theta}) = 0 \Big\}.$$

Owen (1988) and Qin and Lawless (1994) by Lagrange multiplier $(\lambda_1, ..., \lambda_r)$ established that $p_i \propto \frac{1}{1+\boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}_i;\boldsymbol{\theta})}$ maximizes $R_E(\boldsymbol{\theta})$. So,

$$\ell(\boldsymbol{\theta}) \propto -\sum_{i=1}^{n} \log\left[1 + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}_i;\boldsymbol{\theta})\right],$$

is the logarithm of the EL function. Thus, the maximum EL estimator of $\boldsymbol{\theta}$ is obtained

$$(\hat{\boldsymbol{\theta}}^{EL}, \hat{\boldsymbol{\lambda}}) = \arg\min_{\boldsymbol{\theta}} \max_{\boldsymbol{\lambda}} \Big\{ \sum_{i=1}^{n} \log\left[1 + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}_i;\boldsymbol{\theta})\right] \Big\}. \tag{2.1}$$

Tang and Wu (2013) proposed a two-layer algorithm to compute $\hat{\boldsymbol{\theta}}^{EL}$. The inner layer maximizes (2.1) to obtain $\hat{\boldsymbol{\lambda}}$ for a given $\boldsymbol{\theta}$, while the outer layer determines the optimal value of $\boldsymbol{\theta}$ as a function of $\hat{\boldsymbol{\lambda}}$. Both layers utilize the coordinate descent algorithm, updating one component at a time. For further details, see Tang and Wu (2013).

2.1 REL

While the traditional EL method relies solely on the unbiasedness of estimating equations, the RPEL estimator satisfies both unbiasedness and minimum variance criteria. This approach offers a new estimation method with advantageous properties, improving robustness against outliers compared to the traditional EL method.

Bayati et al. (2021) established that, under the criteria of unbiasedness and minimum variance for the estimating equation, the p_i's must follow a specific formula

$$p_i \propto \frac{1}{1 + \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i; \boldsymbol{\theta})}. \tag{2.2}$$

Then

$$L_R(\boldsymbol{\theta}, \mathbf{w}) \propto \prod_{i=1}^{n} p_i = \prod_{i=1}^{n} \frac{1}{1 + \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i; \boldsymbol{\theta})},$$

where $\mathbf{g}^*(\mathbf{x}; \boldsymbol{\theta}) = (g_1^2(\mathbf{x}; \boldsymbol{\theta}), ..., g_r^2(\mathbf{x}; \boldsymbol{\theta}))^T$ and $\mathbf{w} = (w_1, ..., w_r)^T \in \mathbb{R}^{+r}$.

Practically, $L_R(\boldsymbol{\theta}, \mathbf{w})$ offers appealing properties compared to $L(\boldsymbol{\theta}) \propto \prod_{i=1}^{n} \frac{1}{1+\boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}_i;\boldsymbol{\theta})}$ in EL. First, $L_R(\boldsymbol{\theta}, \mathbf{w})$ allows us to assume that the elements of the vector $\mathbf{w}$ are independent of $\boldsymbol{\theta}$, whereas in $L(\boldsymbol{\theta})$, the dependence of λ on $\boldsymbol{\theta}$ is a fundamental assumption. This makes $L_R(\boldsymbol{\theta}, \mathbf{w})$ a semi-parametric version of the standard EL function. Second, by assuming sparsity in the vectors $\boldsymbol{\theta}$ and $\mathbf{w}$ (to reduce the number of parameters and the number of the estimating equation respectively) and leveraging the convexity of $-\log L_R(\boldsymbol{\theta}, \mathbf{w})$, the REL method becomes applicable to high-dimensional data.

To obtain the estimators of $(\boldsymbol{\theta}, \mathbf{w})$, namely $(\hat{\boldsymbol{\theta}}^{REL}, \hat{\mathbf{w}}^{REL})$, we use the logarithm of REL which is

$$\ell_R(\boldsymbol{\theta}, \mathbf{w}) = -\sum_{i=1}^{n} \log\left[1 + \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i; \boldsymbol{\theta})\right]. \tag{2.3}$$

Based on (2.3), the maximum estimators of $(\boldsymbol{\theta}, \mathbf{w})$ are given as

$$(\hat{\boldsymbol{\theta}}^{REL}, \hat{\mathbf{w}}^{REL}) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{\mathbf{w}\in\mathbf{W}_n} \left\{ \sum_{i=1}^{n} \log\left[1 + \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i; \boldsymbol{\theta})\right] \right\},$$

where $W_n(\boldsymbol{\theta}) = \{\mathbf{w} \in \mathbb{R}^{+r} : \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i; \boldsymbol{\theta}) < \gamma\}$ for the tuning parameter γ.

2.2 Joint Empirical Likelihood

According to (2.2), changes in w_l's or g_l's affect other w_l's or g_l's. Because of $1 + w_1 g_1^2(\mathbf{x}_i;\boldsymbol{\theta}) + ... + w_r g_r^2(\mathbf{x}_i;\boldsymbol{\theta})$, w_l's and g_l's are dependent. As a result, each estimating equation is influenced by others. JEL reduces the degree of dependence among the estimating equations. Shantia and Ghoreishi (2024) suggested using $(1 + w_1 g_1^2(\mathbf{x}_i;\boldsymbol{\theta}))...(w_r g_r^2(\mathbf{x}_i;\boldsymbol{\theta}))$ instead of $1 + w_1 g_1^2(\mathbf{x}_i;\boldsymbol{\theta}) + ... + w_r g_r^2(\mathbf{x}_i;\boldsymbol{\theta})$ so that the role of each estimating equations in parameter estimation is accounted separately. Therefore we have

$$L_J(\boldsymbol{\theta},\mathbf{w}) = \prod_{i=1}^{n}\prod_{l=1}^{r}\frac{1}{1 + w_l g_l^2(\mathbf{x}_i;\boldsymbol{\theta})}. \tag{2.4}$$

Since the likelihood function (2.4) consists of the product of two partially bounded empirical likelihood, Shantia and Ghoreishi (2024) refered to it as a JEL. Based on logarithm of $L_J(\boldsymbol{\theta},\mathbf{w})$, the maximum estimator of $(\boldsymbol{\theta},\mathbf{w}))$, namely $(\hat{\boldsymbol{\theta}}^{JEL}, \hat{\mathbf{w}}^{JEL})$, are given as

$$(\hat{\boldsymbol{\theta}}^{JEL}, \hat{\mathbf{w}}^{JEL}) = arg \min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \max_{\mathbf{w}\in\mathbf{W}_n} \left\{ \sum_{i=1}^{n}\sum_{l=1}^{r} \log\left[1 + w_l g_l^2(\mathbf{x}_i;\boldsymbol{\theta})\right] \right\}.$$

2.3 A Penalty Framework for REL and JEL

While the REL or JEL estimator $\hat{\boldsymbol{\theta}}$ generally performs well in practical applications, it becomes inefficient in high-dimensional settings with numerous parameters. Moreover, many models exhibit sparsity in their parameters, particularly in high-dimensional regression, where several independent variables may not influence the response. To shrink these irrelevant parameters to zero, an effective approach is to apply a suitable penalty function that reduces the dimensionality of $\boldsymbol{\theta}$ and $\mathbf{w}$. Bayati et al. (2021) suggested two forms of penalty functions for $\boldsymbol{\theta}$ and $\mathbf{w}$, for $\boldsymbol{\theta}$ applied $P_1(\boldsymbol{\theta}) = \sum_{k=1}^{p} \theta_k^2$, and for $\mathbf{w}$ and some fixed $a > 0$ used

$$P_2(\mathbf{w}) = \sum_{l=1}^{r}\left\{\frac{(w_l - a)^2}{2w_l} + \frac{3}{2}\log w_l - \log a\right\}, \tag{2.5}$$

where A and B are some positive constants called hyper-parameters.. They demonstrated that the penalty function (2.5) is sensitive to both small and large values of $\mathbf{w}$, a crucial property that enhances the efficacy of estimation by selectively down-weighting less significant equations.

3. Baysian analysis

Computing RPEL or JPEL estimates presents significant challenges. In this paper, we conduct a Bayesian analysis of the parameters, adopting an approach to RPEL and JPEL that differs from Lazar (2003) methodology for EL by incorporating distinct priors for $\boldsymbol{\theta}$ and $\mathbf{w}$. This framework provides closed-form conditional posteriors for various estimating equations, facilitating practical implementation. Bayati et al. (2021) applied penalty functions to construct posterior density for $\boldsymbol{\theta}$ and $\mathbf{w}$. In this paper, we extend this approach, with the posterior density for RPEL given by

$$\pi_R(\boldsymbol{\theta}, \mathbf{w}) \propto \prod_{i=1}^{n} \frac{1}{1 + \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i; \boldsymbol{\theta})} \times e^{-u_1 P_1(\boldsymbol{\theta})} \times e^{-u_2 P_2(\mathbf{w})}, \tag{3.6}$$

and the posterior density for JPEL is given by

$$\pi_J(\boldsymbol{\theta}, \mathbf{w}) \propto \prod_{i=1}^{n} \prod_{l=1}^{r} \frac{1}{1 + w_l g_l^2(\mathbf{x}_i; \boldsymbol{\theta})} \times e^{-u_1 P_1(\boldsymbol{\theta})} \times e^{-u_2 P_2(\mathbf{w})}. \tag{3.7}$$

Since $(1 + \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i; \boldsymbol{\theta})) \geq 0$ in Equation (3.6), we can exploit the properties of the exponential distribution. Let s follow an exponential distribution with parameter $(1 + \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i; \boldsymbol{\theta}))$, so that

$$\frac{1}{1 + \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i; \boldsymbol{\theta})} = \int_0^\infty e^{-s(1 + \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i; \boldsymbol{\theta}))} ds.$$

Considering $s_i \sim \mathit{Exponential}(1 + \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i; \boldsymbol{\theta}))$ for $i = 1, \ldots, n$. We can rewrite $\pi_R(\boldsymbol{\theta}, \mathbf{w})$ as

$$\pi_R(\boldsymbol{\theta}, \mathbf{w}) \propto \int_0^\infty \cdots \int_0^\infty e^{-\sum_{i=1}^{n} s_i [1 + \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i; \boldsymbol{\theta})] - u_1 P_1(\boldsymbol{\theta}) - u_2 P_2(\mathbf{w})} ds_1 ds_2 \ldots ds_n.$$

Similarly, for JPEL, let $s_{il} \sim \mathit{Exponential}(1 + w_l g_l^2(\mathbf{x}_i; \boldsymbol{\theta}))$ for $i = 1, \ldots, n$ and $l = 1, \ldots, r$, noting that $1 + w_l g_l^2(\mathbf{x}_i; \boldsymbol{\theta}) \geq 0$ in Equation (3.7). We can then rewrite $\pi_J(\boldsymbol{\theta}, \mathbf{w})$ as

$$\pi_J(\boldsymbol{\theta}, \mathbf{w}) \propto \int_0^\infty \cdots \int_0^\infty e^{-\sum_{i=1}^{n} \sum_{l=1}^{r} s_{ij} [1 + w_l g_l^2(\mathbf{x}_i; \boldsymbol{\theta})] - u_1 P_1(\boldsymbol{\theta}) - u_2 P_2(\mathbf{w})} ds_{11} ds_{12} \ldots ds_{nr}.$$

For statistical inference, random samples can be drawn from the marginal densities using Markov Chain Monte Carlo (MCMC) methods. These samples allow for approximating the posterior distributions of the parameters, enabling efficient estimation and uncertainty quantification, even in complex models. We utilize

Gibbs sampling, a specific MCMC technique, for the parameters $\mathbf{w}, \boldsymbol{\theta}, s_i$ and s_{ij}. This method simplifies the sampling process by providing closed-form conditional distributions. For this purpose we define

$$\pi_R(\boldsymbol{\theta}, \mathbf{w}, s_1, \ldots, s_n) \propto e^{-\sum_{i=1}^{n} s_i[1+\mathbf{w}^T\mathbf{g}^*(\mathbf{x}_i;\boldsymbol{\theta})]-u_1P_1(\boldsymbol{\theta})-u_2P_2(\mathbf{w})},$$

and

$$\pi_J(\boldsymbol{\theta}, \mathbf{w}, s_{11}, \ldots, s_{nr}) \propto e^{-\sum_{i=1}^{n}\sum_{l=1}^{r} s_{il}[1+w_l g_l^2(\mathbf{x}_i;\boldsymbol{\theta})]-u_1P_1(\boldsymbol{\theta})-u_2P_2(\mathbf{w})},$$

Gibbs sampling enables us to sample each variable conditionally, holding the others constant, which is computationally simpler than sampling from the joint distribution. By employing penalty functions as prior densities for the parameters, our method, termed RPEL and JPEL, allows for flexible control over the number of parameters and estimating equations.

4. Empirical Bayesian Information Criterion (EBIC) for RPEL and JPEL

The EBIC is primarily used for model selection in high-dimensional data settings, such as when the number of predictors is close to or exceeds the number of observations. Additionally, it can be used to compare various models. According to Bayati et al. (2021), EBIC is applied for RPEL and JPEL and has the following formula:

$$EBIC(RPEL) := \frac{p(M_0)+r(M_0)}{2}\log(n) + \sum_{i=1}^{n}\frac{1}{1+\|\mathbf{w}\|_1}\log\left[1+\hat{\mathbf{w}}^T\mathbf{g}^*(\mathbf{x}_i;\hat{\boldsymbol{\theta}})\right],$$

$$EBIC(JPEL) := \frac{p(M_0)+r(M_0)}{2}\log(n) + \sum_{i=1}^{n}\sum_{l=1}^{r}\frac{1}{1+w_l}\log\left[1+\hat{w}_l g_l^2(\mathbf{x}_i;\hat{\boldsymbol{\theta}})\right],$$

where $p(M_0)$ and $r(M_0)$ are the number of parameters and estimating equation in model M_0 respectively and $\|.\|$ stands for the l_1-norm.

5. RPEL and JPEL in a multiple regression model

There are many widely used regression models for analyzing high-dimensional data, including linear regression, multiple linear regression, Poisson regression, logistic regression, bridge regression, etc. In this section, we focus on the multiple linear regression model. In many statistical problems, the response variable is influenced by more than one independent variable. Let p represent the number of independent variables $(x_{i1}, x_{i2}, \dots, x_{ip})$ affecting the response variable y_i. We define the regression model as follows:

$$y_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \cdots + \theta_p x_{ip} + \varepsilon_i \qquad i = 1, 2, \dots, n,$$

where $x_{i1} = 1$, with θ_1 as the intercept and θ_j (for $j = 2, \dots, p$) as the slope of $\mathbf{x}_i$. The errors ε_i are independent random variables.
Using ordinary least squares, the estimating equations take the form:

$$g_l(\mathbf{x}_i; \boldsymbol{\theta}) = x_{il}\big(y_i - \theta_1 x_{i1} - \cdots - \theta_p x_{ip}\big), \qquad l = 1, 2, \dots, p. \tag{5.8}$$

Additionally, other estimating equations may be available. Specifically, we take $r = p$, meaning the number of estimating equations matches the number of parameters. Bayati et al. (2021) demonstrate that when $r > p$, setting a penalty function on w can highlight important estimating equations. We will discuss how RPEL and JPEL are used for the regression model.
With the penalty functions $P_1(\boldsymbol{\theta}) = \sum\limits_{l=1}^{p} \theta_l^2$ and $P_2(\mathbf{w})$ as defined in (2.5), the prior distributions are:

$$\begin{aligned}
\pi_1(\boldsymbol{\theta}) &= e^{-u_1 P_1(\boldsymbol{\theta})} = \frac{1}{B^{\frac{p}{2}}} e^{-\frac{1}{2B}\sum\limits_{j=1}^{p}\theta_j^2}, \\
\pi_2(\mathbf{w}) &= e^{-u_2 P_2(\mathbf{w})} = \prod_{l=1}^{r} \frac{A}{w_l^{3/2}} e^{-\frac{(w_l - A)^2}{2w_l}}.
\end{aligned} \tag{5.9}$$

The REL and JEL functions for the regression model are defined as:

$$\begin{aligned}
L_R(\boldsymbol{\theta}, \mathbf{w}) &= \int_0^\infty \cdots \int_0^\infty e^{-\sum\limits_{i=1}^{n} s_i \left(1 + w_1 g_1^2(\mathbf{x}_i;\boldsymbol{\theta}) + \cdots + w_r g_r^2(\mathbf{x}_i;\boldsymbol{\theta})\right)} ds_1 \dots ds_n, \\
L_J(\boldsymbol{\theta}, \mathbf{w}) &= \int_0^\infty \cdots \int_0^\infty e^{-\sum\limits_{i=1}^{n}\sum\limits_{l=1}^{r} s_{il}\left(1 + w_l g_l^2(\mathbf{x}_i;\boldsymbol{\theta})\right)} ds_{11} \dots ds_{nr}.
\end{aligned}$$

Thus, the corresponding RPEL and JPEL functions for the regression model (5.8) are:

$$\pi_R(\boldsymbol{\theta}, \mathbf{w}, s_1, \dots, s_n) \propto e^{-\sum\limits_{i=1}^{n} s_i[1 + \mathbf{w}^T \mathbf{g}^*(\mathbf{x}_i;\boldsymbol{\theta})] - \frac{1}{2B}\sum_{j=1}^{p}\theta_j^2 - \frac{3}{2}\sum_{l=1}^{r}\ln w_l - \sum_{l=1}^{r}\frac{(w_l - A)^2}{2w_l}},$$

$$\pi_J(\boldsymbol{\theta}, \mathbf{w}, s_{11}, \dots, s_{nr}) \propto e^{-\sum\limits_{i=1}^{n}\sum\limits_{l=1}^{r} s_{il}[1 + w_l g_l^2(\mathbf{x}_i;\boldsymbol{\theta})] - \frac{1}{2B}\sum_{j=1}^{p}\theta_j^2 - \frac{3}{2}\sum_{l=1}^{r}\ln w_l - \sum_{l=1}^{r}\frac{(w_l - A)^2}{2w_l}}.$$

Next, we employ a Bayesian method to find the posterior density for each parameter and apply Gibbs sampling. The posterior density distribution of each parameter is obtained as follows:

$$\begin{cases} s_i \sim \mathbf{Exp}\Big(1+\sum\limits_{l=1}^{r} w_l\, g_l^*(\mathbf{x}_i,\boldsymbol{\theta})\Big), & i=1,\dots,n \\ w_l \sim \mathbf{GIG}\Big(2\sum\limits_{i=1}^{n} s_i g_l^*(\mathbf{x}_i,\boldsymbol{\theta})+1, A^2, -\frac{1}{2}\Big), & l=1,\dots,r \\ \theta_j \sim \mathbf{N}\Big(\mu_j,\sigma_j^2\Big), \quad j=1,\dots,p \end{cases} \tag{5.10}$$

The value of μ_j and σ_j for RPEL and JPEL are:

$$\mu_j^{RPEL} = \frac{\sum\limits_{i=1}^{n}\sum\limits_{l=1}^{r} s_i w_l x_{il}^2 x_{ij}\Big(y_i - \sum\limits_{k\neq j} x_{ik}\theta_k\Big)}{\sum\limits_{i=1}^{n}\sum\limits_{l=1}^{r} s_i w_l x_{il}^2 x_{ij}^2 + \frac{1}{2B_0}}, \sigma_j^{RPEL} = \Big[2\sum_{i=1}^{n}\sum_{l=1}^{r} s_i w_l x_{il}^2 x_{ij}^2 + \frac{1}{B_0}\Big]^{-0.5},$$

$$\mu_j^{JPEL} = \frac{\sum\limits_{i=1}^{n}\sum\limits_{l=1}^{r} s_{il} w_l x_{il}^2 x_{ij}\Big(y_i - \sum\limits_{k\neq j} x_{ik}\theta_k\Big)}{\sum\limits_{i=1}^{n}\sum\limits_{l=1}^{r} s_{il} w_l x_{il}^2 x_{ij}^2 + \frac{1}{2B_0}}, \sigma_j^{JPEL} = \Big[2\sum_{i=1}^{n}\sum_{l=1}^{r} s_{il} w_l x_{il}^2 x_{ij}^2 + \frac{1}{B_0}\Big]^{-0.5},$$

where GIG stands for the generalized inverse Gaussian distribution, with the probability density function $f(x;a,b,p) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})} x^{p-1} e^{-\frac{ax+b/x}{2}}$, where $a>0$, $b>0$, $p\in\mathbb{R}$, and K_p is the modified Bessel function of the second kind.
Using the conditional densities from equation (5.10), we can easily apply the Gibbs sampling method to generate random samples from the parameters' marginal distribution for statistical inferences.

6. Simulation study and application

In this section, we compare the performance of the RPEL and JPEL estimators for regression models using a simulation study. The simulation results indicate that RPEL performs better than JPEL. Therefore, we apply RPEL and JPEL to a real data analysis.

6.1 Simulation study

To evaluate the estimators, we use EBIC and MSE. We execute the Gibbs sampling scheme from equation (5.10) to generate $M = 5000$ samples from the marginal distributions of the parameters, with a burn-in period of 20%.
Let the regression model be

$$y_i = \theta_1 + \theta_2 x_{i2} + \theta_3 x_{i3} + \varepsilon_i, \qquad i=1,2,\dots,n,$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$ is the parameter vector of our interest, x_{i2}'s and x_{i3}'s are independent samples from the exponential $Exp(1)$ and $N(0, 0.25)$ respectively. Also, the errors are independent samples from the mixture distribution

$$\varepsilon_i \sim 0.5N(0, 0.5) + 0.5Unif(-1, 1).$$

We assume the true parameter values $\boldsymbol{\theta} = (1, 2, 0)^T$, and generate for various sample sizes $n = 20, 50, 500$. The results are in Table 1. For a small sample size ($n = 20$), RPEL provides estimates very close to the true values, especially for $\hat{\theta}_2$. However, $\hat{\theta}_3$ shows a significant deviation from the true value of 0, indicating challenges in estimating parameters that are close to zero. JPEL's estimates, while also relatively close, show slightly greater deviations, especially in the $\hat{\theta}_2$ estimate, which is more than 0.08 away from the true value.

Table 1: Estimation of the parameters

Sample size	Method	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	EBIC	MSE
$n = 20$	RPEL	0.9446	2.0066	0.2748	10.5478	0.2262
	JPEL	0.9822	2.0859	0.2094	10.6113	0.3078
$n = 50$	RPEL	0.9726	2.0140	0.1065	13.7892	0.0940
	JPEL	1.0966	1.9271	-0.0960	14.4516	0.1243
$n = 500$	RPEL	0.9965	1.9962	0.0106	24.7085	0.0920
	JPEL	1.0290	1.9674	-0.0744	34.3325	0.10310

For a moderate sample size ($n = 50$), **with** a larger sample size, RPEL continues to produce increasingly accurate estimates. The estimate for $\hat{\theta}_1$ is very close to 1, and $\hat{\theta}_2$ remains very close to 2. JPEL shows a more notable deviation in $\hat{\theta}_3$, which is estimated at -0.0960, highlighting the sensitivity of JPEL to parameter estimation when the true value is near zero.

The estimates reveal that as **the** sample size increases, RPEL's performance remains robust, while JPEL begins to exhibit instability. For a large sample size ($n = 500$), RPEL estimates are remarkably close to the true parameter values, particularly $\hat{\theta}_1$, and $\hat{\theta}_2$, demonstrating high accuracy.

JPEL's estimates, while improving somewhat with the larger sample size, still deviate significantly for $\hat{\theta}_3$, which reflects potential issues with model specification or penalization in estimating near-zero coefficients.

The results indicate that RPEL consistently provides better estimates as **the** sample size increases, highlighting its efficiency in parameter estimation for regression models.

For each sample size, RPEL exhibits lower EBIC values compared to JPEL. The lower EBIC values for RPEL indicate a preference for simpler models, which can lead to more interpretable and reliable parameter estimates. In the context of model selection, RPEL's performance suggests it balances model fit and complexity more effectively than JPEL, especially at larger sample sizes where complexity increases.

The MSE values provide insights into the predictive accuracy of the estimators. **RPEL** consistently outperforms JPEL in terms of MSE across all sample sizes, indicating that RPEL provides more accurate predictions of the response variable. Notably, the MSE values decrease as the sample size increases, which is expected; however, JPEL does not demonstrate the same level of accuracy improvement as RPEL.

Table 2: Estimation of the parameters with outliers

Sample size	Method	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	EBIC	MSE
$n = 20$	RPEL	0.9752	2.0792	0.0649	14.5594	4.4316
	JPEL	1.0513	2.0504	0.0339	17.0852	5.0941
$n = 50$	RPEL	0.9616	2.0278	0.1156	17.0723	3.3126
	JPEL	1.0847	1.9478	-0.0042	28.46	3.81
$n = 500$	RPEL	1.0467	1.9916	-0.0166	36.9725	3.1386
	JPEL	1.1080	1.9895	-0.0765	110.6261	3.5448

To assess the impact of outliers on parameter estimates, we replace 5% of the values in the response variable y with random values drawn from a $N(10, 1)$ distribution. The results of this outlier analysis are presented in Table 2.

Across all sample sizes, JPEL consistently shows higher EBIC values than RPEL, especially for larger sample sizes, where JPEL's EBIC reaches over 110 compared to RPEL's 36.97. This suggests that RPEL might be better suited for achieving simpler models (lower EBIC) when outliers are present, indicating a potential advantage in model selection with RPEL in the presence of outliers. RPEL appears more robust to outliers across sample sizes, offering estimates closer to the true parameter values, lower EBIC (indicating simpler models), and lower MSE values (indicating better predictive accuracy).

JPEL tends to show more variability in parameter estimates, higher EBIC values (suggesting increased model complexity), and higher MSE values, especially in larger sample sizes, indicating a greater sensitivity to the presence of outliers.

6.2 Real Data Analysis

To illustrate our methodology, we utilize the Boston Housing Dataset, which is publicly available through the MIT OpenCourseWare website. This dataset contains information on housing in Boston, Massachusetts, including various features such as crime rate, average number of rooms per dwelling, and distance to employment centers. The target variable in our analysis is the median house price.

The dataset is well-suited for multiple regression analysis, as it allows us to examine the relationships between multiple predictor variables and the housing prices. The analysis helps demonstrate the effectiveness of our proposed approach in real-world scenarios.

The variables in the dataset that we consider for regression analysis include:

- X_1: Per capita crime rate by town.
- X_2: Proportion of non-retail business acres per town.
- X_3: Nitrogen oxide concentration (pollution level).
- X_4: Average number of rooms per dwelling.
- X_5: Proportion of owner-occupied units built before 1940.
- X_6: Weighted distances to five Boston employment centers.
- X_7: Property tax rate per 10,000.
- X_8: Pupil-teacher ratio by town.
- Y: Median value of owner-occupied homes (in $1000s$) *(Target variable)*.

We define our multiple regression model as follows:

$$Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \cdots + \theta_8 X_8 + \epsilon,$$

Table 3:

Method	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$	$\hat{\theta}_6$	$\hat{\theta}_7$	$\hat{\theta}_7$	MSE
LR	0.350	-0.084	-2.629	7.488	-0.064	-0.844	-0.011	-0.519	6.662
RPEL	0.301	-0.0857	-2.707	7.483	-0.065	-0.851	-0.011	-0.593	6.127
JPEL	0.069	-0.0753	-2.887	7.470	-0.064	-0.857	-0.011	-0.599	6.150

The performance of our methodology is evaluated using three different approaches, LR, RPEL, and JPEL. Table 3 presents the estimated coefficients ($\hat{\theta}_i$) for each method, along with the MSE as a measure of model performance.

From the results in Table 3, we observe that the estimates for the regression coefficients vary slightly across the methods. The LR approach provides a baseline for comparison. The RPEL and JPEL methods incorporate penalization techniques, which can improve generalization and reduce overfitting by imposing constraints on the regression coefficients.

In particular, the RPEL and JPEL methods yield lower MSE compared to **the** LR, suggesting that these approaches enhance predictive performance. The RPEL achieves the lowest MSE (6.127), demonstrating its effectiveness in improving model accuracy. The JPEL follows closely with the MSE of 6.150, indicating that joint penalization also contributes to improved estimation.

Overall, the results confirm that employing penalized regression techniques such as RPEL and JPEL can lead to improved predictive accuracy and robustness in the estimation of housing prices.

Discussion and Results

Overall, the results demonstrate that RPEL outperforms JPEL in terms of parameter estimation accuracy, model selection (EBIC), predictive accuracy (MSE), and robustness to outliers. RPEL provides more accurate and reliable estimates across all sample sizes and is better suited for model selection and prediction in both normal and outlier-affected data. JPEL, while useful in some contexts, exhibits greater instability and sensitivity to data challenges, especially when the true parameter values are close to zero or when outliers are present. Thus, for regression models where estimation stability and interpretability are critical, RPEL is the preferred choice over JPEL.

References

Arslan, O., Özdemir, Ş. (2023). Robust penalized empirical likelihood estimation method for linear regression. *Statistics*, **57(2)**, 423–443.

Bartolucci, F. (2007). Penalized empirical likelihood estimation in semiparametric models. *Journal of Statistical Planning and Inference*, **137(10)**, 3362-3375.

Bayati, M., Ghoreishi, Sk, and Wu, J.(2021) Bayesian analysis of restricted penalized empirical likelihood. *Journal of computational statistics.* **20(1)**, 11-20

Bayati, M., Ghoreishi, Sk, and Wu, J.(2021) Restricted Empirical Likelihood Estimation for Time Series Autoregressive Models. *Journal of Statistical Theory and Applications.*, **36**, 1321–1339

Chang, J., Chen, S. X., and Tang, C. Y. (2015). High dimensional generalized empirical likelihood for moment restrictions with dependent data. *Journal of Econometrics*, **185(1)**, 283-304.

Chang, J., Tang, C. Y., and Wu, Y. (2017). A new scope of penalized empirical likelihood with high-dimensional estimating equations. *Biometrika*, **104(2)**, 329-343.

Chen, J., and Cui, H. J. (2006). On Bartlett correction of empirical likelihood. *Biometrika*, **93(4)**, 627-643.

Chen, J., and Cui, H. J. (2007). Properties of empirical likelihood with moments restrictions. *Journal of Multivariate Analysis*, **98(4)**, 742-755.

Chen, J., Variyath, A. M., and Abraham, B. (2009). Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, **18(3)**, 570-589.

DiCiccio, T. J., and Romano, J. P. (1989). On adjustments based on the signed root of the empirical likelihood ratio statistic. *Biometrika*, **76(3)**, 447-456.

Ji, Y., and Liu, Y. (2024). A Penalized Empirical Likelihood Approach for Estimating Population Sizes under the Negative Binomial Regression Model. *Mathematics*, **12(17)**, 2674.

Hall, P., and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Journal of the American Statistical Association*, **85(412)**, 1122-1129.

Hjort, N. L., McKeague, I. W., and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *The Annals of Statistics*, **37(3)**, 1079-1111.

Lahiri, S. N., and Mukhopadhyay, S. (2012). Empirical likelihood for high-dimensional data. *The Annals of Statistics*, **40(5)**, 2511-2540.

Lazar, N. A. (2003). Bayesian empirical likelihood. *Biometrika*,**90(2)**, 319–326.

Leng, C., and Tang, C. Y. (2012). Penalized empirical likelihood and growing dimensional general estimating equations. *Biometrika*, **99(3)**, 703-716.

Newey, W. K., and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, **72(1)**, 219-255.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75(2)**, 237-249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, **18(1)**, 90-120.

Owen, A. B. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, **19(4)**, 1725-1747.

Qin, J., and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, **22(1)**, 300-325.

Shantia, E., and Ghoreishi, S. (2024). Restricted joint empirical likelihood methods for hierarchical models with application in high-dimensional analysis. *Journal of Statistical Research*, **58(1)**, 14-35.

Tang, C. Y., and Leng, C. (2010). Penalized high-dimensional empirical likelihood. *Biometrika*, **97(4)**, 905-920.

Tang, C. Y., and Wu, T. T. (2013). Nested coordinate descent algorithms for empirical likelihood. *Journal of Statistical Computation and Simulation*, **84(9)**, 1917–1930.

Tsao, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *The Annals of Statistics*, **32(3)**, 1215-1221.

Tsao, M., and Wu, F. C. (2013). Empirical likelihood on the full parameter space. *The Annals of Statistics*, **41(4)**, 2176-2196.

Tsao, M., and Wu, F. C. (2014). Extended empirical likelihood for estimating equations. *Biometrika*, **101(3)**, 703-710.

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