Maximum product spacings method for the estimation of parameters of linear regression

Sukrit Thongkairat\textsuperscript{1,2}, Woraphon Yamaka\textsuperscript{1,3} and Songsak Sriboonchitta\textsuperscript{1}

\textsuperscript{1}Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Chiang Mai, 50200, Thailand
\textsuperscript{2}International College of Digital Innovation, Chiang Mai University, Chiang Mai 50200, Thailand
\textsuperscript{3}Email: woraphon.econ@cmu.ac.th

Abstract. Maximum product of spacing (MPS) estimator, which is a general method for estimating parameters from observations with continuous univariate distributions, is considered as an alternative approach in linear regression modelling. We describe the basic idea of the maximum spacings estimator and apply to the linear regression problem. Moreover, we conduct a simulation and experiment study to make the comparison between MPS method and maximum likelihood estimator under various distribution assumptions. Finally, a real data set has been implemented to illustrate the performance of this estimator.

1. Introduction

In the parametric estimation techniques, the assumption of the density has to be explicit in order to make the inference. However, in the real data application, the true distribution is not available and prior knowledge is needed to approximate the distance between the model and the true distribution. One of the most powerful and acceptable methods for making the statistical inference is the maximum likelihood (ML) as it often works properly especially for discrete distributions. Although, this method has been proven to be consistent, asymptotically efficient under very general conditions, it was found to be unbound and inefficient in the estimation of mixtures of continuous distributions, heavy-tailed distributions and J-shape distributions, (see [1], [2], [3]) respectively. [3] mentioned that the ML estimation often breaks down as one parameter tending to cause the likelihood to be infinite, rendering the other parameters inconsistent. To overcome the drawbacks in the ML estimator, [4] proposed a maximum product of spacings (MPS) estimator to deal with those problems as it will return valid results over a much wider range of distributions. Then, [2] justified the MPS estimation by showing that it has similar properties as ML estimators including asymptotic sufficiency, but with more robust properties for various classes of problems. In regular cases MPS estimators are consistent, asymptotically normal and efficient, as derived in the works of [5] and [6].

[4] and [2] explained that we need to maximize the geometric mean of spacings, which are the differences between the values of the cumulative distribution function at neighboring data points. To fix the idea, let \( \theta \) be the unknown parameter and \( X_i \) be independent observations from a cumulative distribution function \( F_\theta (\cdot) \) then applying the probability integral transform \( F_\theta (\cdot) \) to order \( X_i \) yielding \( 0 \leq D_1 < D_2 < \ldots < D_n \leq 1 \), where \( D_i = F(X_i) - F(X_{i-1}) \), \( i = 1, \ldots, n \). The MPS finds the \( \theta \) which
maximizes product of spacings $\prod_{i=1}^{n} D_i$. According to this explanation, if the ML and MPS estimators are well-defined, MPS and ML estimation are asymptotically equal and have the same asymptotic sufficiency, consistency and efficiency properties, when $i \rightarrow n$.

As mentioned above, MPS are at least as efficient as ML estimation, especially when the true densities are the heavy tailed, mixture distribution, or J-shape. Thus, MPS estimator is expected to be more efficient than the ML estimator when the density is skewed or heavy tail and also expected to be an alternative estimator in various models. In this study, MPS estimator has been introduced to the linear regression model. Finally, the proposed method will be compared to the conventional ML method which has received much attention in the literature working on the frequentist probability and which has been widely employed in the regression model. To the best of our knowledge, the MPS estimator in regression model has not been explored in any literatures. Therefore, the main contribution of this study is to propose a robust method for drawing inferences about linear regression problems from a likelihood-based perspective.

The remainder of this paper is divided into three sections. In section 2, we briefly present the maximum likelihood linear regression model. In Section 3, maximum product of spacings estimator for regression is presented. Section 3 presents Monte Carlo evidence on the numerical performance of MPS and ML. A real data example is analyzed in Section 5. Section 6 presents conclusion and suggestions for further works.

2. Maximum likelihood estimator

2.1. Regression model
The regression model can be written as

$$Y_i = \beta_k (x'_{k,t}) + \epsilon_i, \quad (2.1)$$

where $Y_i$ is $(T \times 1)$ continuous dependent variable at time $t$, $x'_{k,t}$ is a matrix of $(T \times K)$ continuous independent variables at time $t$, and $\beta_k$ is $(K \times 1)$ vector of unknown parameters with respect to the covariates. The error term of the model $\epsilon_i$ is $(T \times 1)$ dimensional vector which is assumed to have parametric distribution with mean zero and variance $\sigma^2$.

2.2. Model estimation
In this study, various distributions consisting of normal, student-t, and skewed student-t types are considered. The density functions are written as follows:

Normal likelihood function.

$$f_{\epsilon}(\theta) = \prod_{t=1}^{T} \left( \frac{1}{\sqrt{2\pi(\sigma^2)}} \exp \left( -\frac{\epsilon_i^2}{2(\sigma^2)} \right) \right), \quad (2.2)$$

where $\epsilon_i^2 = (Y_i - \beta_k (x'_{t,k}))^2$ and $\sigma > 0$ is variance parameter.

Student’s t likelihood.

$$f_{\epsilon}(\theta) = \prod_{t=1}^{T} \left( \frac{1}{\sqrt{\Gamma(\nu/2)\Gamma(\nu/2)}} \left( 1 + \frac{\epsilon_i^2}{(\nu\sigma^2)} \right)^{-\nu/2} \right)^{-1}, \quad (2.3)$$

where $\nu > 2$ is degree of freedom.

Skewed student’s t likelihood
where 

\[ a = 4\gamma c \frac{v - 2}{v - 1}, \]

\[ b^2 = 1 + 3\gamma^2 - a^2, \]

\[ c = \frac{\Gamma\left(\frac{v + 1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi(v - 2)}}, \gamma > 0. \]

\[ \Gamma(\cdot) \] is gamma distribution. \( \gamma \) is skewed parameter. One usually uses the Maximum Likelihood (ML) estimator to obtain the parameter estimates in the linear regression model in equation (2.1). Then, let \( \theta \) be all unknown parameters, the likelihood function is given by

\[ L(\theta) = \prod_{i=1}^{T} f(\theta|Y_i, x_i), \]

where \( f(\theta|Y_i, x_i) \) is normal, student-t or skewed student-t likelihoods. Then, the log likelihood will be written as

\[ \log L(\theta) = \sum_{i=1}^{T} \log f(\theta|Y_i, x_i). \]

To estimate the unknown parameters, we use equation (2.7) to obtain the ML estimates of the parameters of the linear regression model in equation (2.1). The efficiency equation is

\[ \hat{\theta}_{ML} = \arg \max \sum_{i=1}^{T} \log f(\theta|Y_i, x_i). \]

### 3. Maximum product of spacings estimator

Suppose we have an ordered random sample \( X_1, ..., X_{T-1} \) drawn from a continuous distribution with probability distribution function \( F_\theta, \theta \in \mathbb{Y} \subset \mathbb{R}^d \) and let the order sample \( X_1 < X_2 < ... < X_{T-1} \) where \( X_0 = -\infty \) and \( X_T = \infty \). Thus, according to [4] and [2], we can construct the 1-step spacing as

\[ D_i(\theta) = F_\theta(X_i) - F_\theta(X_{i-1}), \quad i = 1, ..., T \]  

where \( F_\theta(\cdot) \) is probability distribution function and \( \sum_{i=1}^{T} D_i = 1 \). To estimate unknown parameter \( \theta \), we have to define product spacings and maximize the geometric mean of the spacings

\[ G = \left\{ \prod_{i=1}^{T} D_i \right\}^{\frac{1}{T}} = \left\{ \prod_{i=1}^{T} F_\theta(X_i) - F_\theta(X_{i-1}) \right\}^{\frac{1}{T}}. \]

or, equivalently, its logarithm

\[ H = \log(G) = \frac{1}{T} \sum_{i=1}^{T} \log \left\{ F_\theta(X_i) - F_\theta(X_{i-1}) \right\}, \]

where \( F_\theta(X_0) \equiv 0 \) and \( F_\theta(X_{T-1}) \equiv 1 \). [7] omit the \( 1/T \) factor in front of the summation \( \sum_{i=1}^{T} (\cdot) \) as these are constants with respect to \( \theta \), the modifications do not alter the location of the maximum of the function \( H \). Therefore, in this study, we investigate the estimator of \( \theta \), defined as any parameter values in \( \Theta \) that maximizes the quantity in equation (3.3).
\[ \hat{\theta}_{\text{MPS}} = \arg \max \sum_{i=1}^{T} \log \left\{ F_{\theta}(X_i) - F_{\theta}(X_{i-1}) \right\}. \]  

[6] and [8] mentioned that MPS and ML estimation are asymptotically equal and have the same asymptotic sufficiency, consistency and efficiency properties as differentiating \( H \) with respect to \( \hat{\theta}_{\text{MPS}} \) is essentially the same as \( \log L(\theta) \) since \( F_{\theta}(X_i) - F_{\theta}(X_{i-1}) \to 0 \) when \( t \) increases.

Note that, three distributions consisting of normal, student-t, and skew-student-t are considered. In the MPS and ML estimations, the probability distribution functions \( F_{\theta}(X_i) \) are plotted in figure 1. In particular, this figure shows similarities between two functions and also confirms that the maximum spacing estimates are identified.

Figure 1. Likelihood function and spacings function for three distributions. Parameters have been set as \( u = 0, \sigma = 1, \gamma = 1 \) and \( v = 4 \).

4. Simulation study

In order to compare the finite sample properties of the MPS method and the ML method in parameter estimates, a set of simulations was performed based on three parametric distributions namely normal, Student-t and skewed Student-t. The simulation procedure is implemented using the following linear model

\[ Y_i = \beta_0 + \beta_1 (x_i') + \varepsilon_i. \]  

Our interest is to estimate the fixed effects parameters \( \beta_0, \beta_1 \) and nuisance parameter \( \sigma \), degree of freedom parameter \( v \) and skewed parameter \( \gamma \). The simulated dataset was generated as follows. We considered a one covariate model thus \( x_i' \) is generated from standard normal \( N(0,1) \). The parameters are chosen as \( \beta_0 = 0.1, \beta_1 = 3, \sigma = 1, v = 4 \) and \( \gamma = 10 \). The error term \( \varepsilon_i \) has been generated independently from normal \( N(0,1) \), student-t \( T(0,1,v) \), skewed student-t \( sT(0,1,v,\gamma) \) and mixture uniform-normal \( \text{Unif}(0,1)-N(0,1) \) distributions. The different sample sizes, say, \( n = 10, 15, 20, 30, \) and 50 are considered for each case. For each case, we generated \( M = 1000 \) datasets. To examine the performance of the ML and MPS estimations, we employ the bias (Bias) for each parameter over the 100 replicates.

\[ \text{Bias}(\hat{\theta}) = \left| \frac{1}{M} \sum_{j=1}^{M} (\hat{\theta}_j - \theta) \right|. \]
where $\tilde{\theta}_j = \{\tilde{\beta}_0, \tilde{\beta}_1\}$ and $\bar{\theta} = \{0.1, 2\}$ and $M = 1,000$ is number of repetitions.

Figure 2. Bias plots of each parameter.

The result of the simulation study is presented in figure 2. The results show the bias of parameters under various distributions and sample sizes. It can be observed that the Bias for the regression parameters $\beta_0$ and $\beta_1$ tends to close to zero when sample size is increased indicating that the MPS estimation has consistent asymptotic properties and the reliable results are obtained (see, Bias of MPS estimator (dotted rad line) and ML estimator (solid black line)). We can confirm that MPS estimator is
a consistent estimator in that it converges in probability to the true value of the parameter as the sample size increases. In addition, when we compare the performance between MPS estimator and ML estimator, we can observe that MPS estimator does not completely outperform the ML estimator in terms of Bias values. However, we find some interesting evidences form the results that the Bias of $\beta_1$ obtained from MPS is low when the number of observation is small ($n<20$), except for $\beta_0$ skewed student-t error. In the last experiment study, we simulate the data based on the mixture uniform-normal distribution in order to assess the performance of these two estimators when the data is mixture distribution. [2] claimed that maximum likelihood will fail while MPS estimation succeeds, when there exists mixture distribution error. Thus, it is interesting to examine this aspect. In this case, we consider only the ML and MPS based on the normal distribution. According to the result, the MPS estimator is likely to be superior to ML estimator when the sample size is small ($n < 15$). This result is in line with the suggestion of [2] that in cases where the underlying distribution is mixture, maximum likelihood will fail while MSP estimation succeeds. We can conclude that MPS estimation makes it possible to use smaller data sets compared to ML estimation when the true distribution is mixture.

5. Application

In this section, we show an application based on a dataset from [9] on characteristics of Australian athletes available from the Australian Institute of Sport (AIS). We consider body mass index (bmi), lean body mass (lbm) and body fat (fat) associated with $n = 20$ Australian athletes. We consider the following linear regression model

$$bmi_i = \beta_0 + \beta_1 l bm_i + \beta_2 f at_i + \varepsilon_i$$

(5.1)

where $\beta_0$ is the constant term, $\beta_1$ is the coefficient of the lean body mass and $\beta_2$ is the coefficient of the gender, $\varepsilon_i$ belongs to the distribution for $i=1,...,20$.

First of all, we compare the performance of our estimator to the MLE under various distribution assumptions. To compare them, we consider the mean square error (MSE) and the lowest MSE is preferred. Table 1. presents the mean squared errors from different distributions for comparison between the maximum likelihood method (ML) and the method of maximum product of spacings (MPS). It can be seen from the results that the MPS estimation has a mean square error less than the ML estimation in all cases, excepted for a Student-t distribution. However, the overall result of MPS under student-t shows the lowest MSE. According to these measures, it can be concluded that the best estimation is the MPS under student-t.

| Model comparison by mean squared error for the linear regression model using the MPS and ML estimates. |
|----------------------------------|----|----|----|
|                                | norm | std | std |
| ML         | 62.7932 | 62.9545 | 62.6023 |
| MPS         | 62.9239 | 62.2521 | 62.4587 |

| Table 2. Parameter estimation under Student-t distribution. |
|------------------------------------------|--------------------|----------|--------------------|----------|----------|
| Estimate | Std. Error | P-value | Estimate | Std. Error | P-value |
| $\beta_0$ | 8.2855 | 3.0696 | 0.0069 | 8.2725 | 1.4651 | 0.0000 |
| $\beta_1$ | 0.1661 | 0.0535 | 0.0019 | 0.1681 | 0.0114 | 0.0000 |
| $\beta_2$ | 0.2015 | 0.0814 | 0.0132 | 0.1972 | 0.0186 | 0.0000 |
| $\sigma$ | 1.1520 | 0.1921 | 0.0000 | 1.3355 | 1.6070 | 0.2107 |
| $v$ | 30.0000 | 101.9525 | 0.7685 | 29.9999 | 370.8756 | 0.9352 |
Table 2 presents the parameter estimation for the maximum product of spacings method under the Student-t distribution. It is important to note that the significance of the coefficient is verified at $\alpha = 0.01$ level. We can observe that the standard errors of $\beta_0$, $\beta_1$ and $\beta_2$ obtained from MPS are considerably smaller than that of the ML estimation. This result confirms that MPS show a good performance in the real data application and it can be adopted as an alternative estimation in the analysis of regression models.

Finally, the estimated parameters obtained from the MPS method have the following meanings: A 1% change in lean body mass (lbm) will correspond to the body mass index (bmi) change by 0.1681% in the same direction. In addition, 1% change in body fat will correspond to the body mass index around 0.1972%.

6. Conclusions
In this paper, we propose the maximum product of spacings method for the estimation of the linear regression model based on different distributions, namely normal, Student-t, and skewed Student-t, and compare the results with those from the maximum likelihood method (ML). We employ an experiment study to examine the performance of MPS over the wide range of error distribution assumption. From the result, it is observed that all the estimates appear to be consistent and outperform the ML method in some cases. The MPS and ML estimators are equally efficient when the error has parametric distribution. For mixture distribution, however, the MPS estimator is likely to be superior to ML estimator when the sample size is small. Finally, the proposed method can be extended to other time econometric models, such as discrete model, non-linear model and switching model.

References
[1] Pitman E J G 1979 Some basic theory for statistical inference Chapman & Hall
[2] Ranneby B 1984 The maximum spacing method. An estimation method related to the maximum likelihood method Scandinavian Journal of Statistics pp 93-112
[3] Cheng R C H and Amin N A K 1983 Estimating parameters in continuous univariate distributions with a shifted origin Journal of the Royal Statistical Society pp 394-403
[4] Cheng R C H and Amin N A K 1979 Maximum product of spacings estimation with application to the lognormal distributions Math Report 79
[5] Shao Y and Hahn M G 1999 Strong consistency of the maximum product of spacings estimates with applications in nonparametrics and in estimation of unimodal densities Annals of the Institute of Statistical Mathematics 51(1) pp 31-49
[6] Ghosh K and Jammalamadaka S R 2001 A general estimation method using spacings Journal of Statistical Planning and Inference 93(1-2) pp 71-82
[7] Cheng R C H and Stephens M A 1989 A goodness-of-fit test using Moran's statistic with estimated parameters Biometrika 76(2) pp 386–392
[8] Kumar Singh R, Kumar Singh S and Singh U 2016 Maximum product spacings method for the estimation of parameters of generalized inverted exponential distribution under Progressive Type II Censoring Journal of Statistics and Management Systems 19(2) pp 219-245
[9] Cook R D and Weisberg S 1994 Transforming a response variable for linearity Biometrika 81(4) pp 731-737