Thai **J**ournal of **M**athematics : (2019) 217–233
Special Issue: Structural Change Modeling
and Optimization in Econometrics 2018

http://thaijmath.in.cmu.ac.th
ISSN 1686-0209

Bayesian Empirical Likelihood Estimation of Smooth Kink Regression

Woraphon Yamaka[1] **and Paravee Maneejuk**

Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand
e-mail : `woraphon.econ@gmail.com (W. Yamaka)`
`mparavee@gmail.com (P. Maneejuk)`

Abstract : The smooth kink regression model is introduced in this study. The model provides more flexibility in investigating the nonlinear effect of independent variable on dependent variable. The logistic function is considered as a regime weighting function for separating our two-regime model. In the estimation point of view, we employ the Bayesian empirical likelihood (BEL) as it gives a flexible way of combining data with prior information from our knowledge and the empirical likelihood in order to avoid the misspecification of the likelihood function. The performance and accuracy of the estimation from our proposed model is examined by the simulation study and real data.

Keywords : smooth kink model; logistic function; Bayesian empirical likelihood estimation.
2010 Mathematics Subject Classification : 35K05; 91G20.

1 Introduction

In many applications, the causal effect studies are well described by a kink regression model [1],[2],[3]. This model is a piecewise linear regression segment, where coefficients suddenly switch from one stage to another stage at the kink point. In other words, there are two or more regimes of the casual effect. However, this sudden change is in fact difficult to exist in the real data analysis as the behavior of data is rather characterized by gradual change. Therefore, we modify a sudden-switch kink regression model with unknown

[1] Corresponding author.

threshold to a smooth-switch kink regression model. This model is called smooth kink regression throughout this paper.

This paper aims at estimating the parameters of this model and it can be achieved by various estimation techniques such as Ordinary least squares (OLS) and Maximum Likelihood (ML) estimators. However, these classical frequentist methods are often faced with the poor estimation such as biased and inconsistent results under the nonlinear model context. This is because the nonlinear model needs a large sample data in order to split the data into two or more regimes.

Nevertheless, the large sample size is not always obtainable in various fields of research. In particular, the studies involving primary data research with real people will have small data due to high cost of conducting in-person interviews. Not only the limitation in the primary data studies, the studies involving secondary data, which is a research data that has previously been gathered by government public services departments, libraries, internet searches and censuses, also face with the small sample size of the data as some data are difficult to collect, for example population of the country, household debts, and other macroeconomic variables. Collecting these data is a time-consuming process, thus, the limited data problem will lead to ill-posed problem which arises when the amount of sample information is insufficient for model estimation. In other words, the model contains more parameters than can be justified by the data [4],[5]. When the data is small, it sometime reduces the performance and efficiency of model estimation. In addition, Stein [6] showed that it is inadmissible when the number of coefficients in the model is large. It is widely understood, the larger sample size of data can bring the higher probability of finding a significant result [7]. Button et al. [8] suggested that when the sample is limited, it is often hard to get meaningful results. Therefore, the conventional estimations are difficult to reach the optimal solution.

To deal with this data limitation, the Bayesian estimation is considered as a common and reasonable estimator for small data. This estimation does not require a large sample, as a prior distribution can be incorporated in the model estimation. A smaller data size model can be estimated without losing any power while retaining precision [9]. Lee and Song [10] found that Bayesian estimation requires only 1:3 ratio of parameters to observations. Therefore, it is reasonable here to employ a Bayesian approach for constructing density distributions based on limited information.

Another concern is that ML and Bayesian estimations are not suitable for estimating the parameters when the data is not normally distributed. In the case of ML estimation, Li, Wong, Lamoureux, and Wong suggested that if the assumption of normal distribution does not hold, it will bring a problem to the confidence interval and standard error of the estimated parameter, and thereby leading to a wrong significant result. It sometimes produces a biased estimation or even misleading results when the true distribution of errors may not be normal and may exist with heavy-tailed or skewed nature [11]. In the Bayesian case, we know that the important step is the simulation from the posterior distribution using Markov Chain Monte Carlo (MCMC) sampler. However, it is very sensitive to the likelihood, and we need to evaluate the likelihood density, which is difficult, in order to use MCMC. Due to the complexity of the estimation and model, the likelihood is generally assumed to have normal likelihood or other parametric likelihood density. Therefore, there are required to develop a more robust likelihood to make statistical inference. To this end, throughout the study, the empirical likelihood is considered as it can relax the strong assumption of normality and get around the limited data problem. Basically, the empirical likelihood consists of the objective function and constraint function where the objective function has entropy discrete distribution whilst model equations and observed information are treated as the constraint function. In short, its a non-parametric likelihood, which is fundamental for the likelihood-based statistical methodology.

As we mentioned above, there are two issues in this study. First, the smooth kink regression is introduced as a new non-linear econometric model. Second, we concern about the limitation of the data and unknown likelihood distribution in the real data study, therefore, this study uses empirical likelihood to approximate the likelihood in the Bayesian computation. This estimation is called Bayesian empirical likelihood (BEL). Recently, BEL method has been discussed by Schennach [12], Fang and Mukerjee [13], Chang and Mukerjee [14]. The BEL inference has been applied to various time series models. For examples, Yang and He [15], Zhang and Tang [11] considered BEL estimation in Quantile model; Chib, Shin and Simoni [16] developed a BEL approach for linear regression model. Recently, Yamaka, Pastpipatkul, Sriboonchitta [17] extended this estimation to estimate nonlinear Kink regression model. However, to our knowledge, there is no work done on the BEL inference on smooth kink regression model yet. So, we develop a BEL inference on our proposed model. Moreover, a Markov Chain Monte Carlo (MCMC) method is also presented to make Bayesian inference on parameters using the Metropolis-Hastings algorithm.

The paper is organized as follows. Section 2 presents the smooth kink regression while Section 3 details the Bayesian empirical likelihood estimation algorithm and the testing threshold effects in smooth kink regression model is explained. Simulation study and results are provided in Section 4. The application study is presented in Section 5 and we conclude with a summary in Section 6.

2 Smooth Kink Regression Model

In this study, two regime Kink regression model is considered, and it relies on.

$$\begin{aligned} Y_t = \beta_1{}^{-} x_{1t}(1-F(\gamma_1,s_1))_{-} + \beta_1{}^{+} x_{1t} F(\gamma_1,s_1)_{+} +, ..., \\ +\beta_K{}^{-} x_{Kt}(1-F(\gamma_K,s_K))_{-} + \beta_K{}^{+} x_{Kt} F(\gamma_K,s_K)_{+} + \alpha Z_t + \varepsilon_t \end{aligned} \tag{2.1}$$

where $t = 1,...,T$. Y_t is $[T \times 1]$ sequence of response variable at time t, x_{kt} is a matrix of $(T \times K)$ predictor variables at time t. Z_t is the regime independent exogenous variables. The relationship between Y_t and x_{kt} is non-linear while there is a linear relationship between Y_t and Z_t. Therefore, the relationship of x_{kt} with Y_t changes at the unknown location threshold or kink point γ_k, thus β has a matrix of $(K \times 2)$ unknown coefficient parameters. In other words, the kink regression function of this model is continuous in the variables x_{kt} and Z_t, but the slope with respect to x_{kt} is discontinuous at the kink point. The parameters $(\beta_1{}^{+},...,\beta_K{}^{+})$ and $(\beta_1{}^{-},...,\beta_K{}^{-})$ are the coefficients with respect to $x_{kt}F(\gamma_k,s_k)_{+}$ and with respect to $x_{kt}(1-F(\gamma_k,s_k))_{-}$ respectively. In other words, the predictor variables can be separated into two regimes according to unknown threshold parameter of threshold or kink point γ_k and smoothed parameter s_k. Note that $F(\gamma_k,s_k)$ is a continuous and smooth transition function, which is bounded between $[0,1]$. In this study, we define this transition function as a logistic function.

$$F(\gamma_k, s_k) = \frac{1}{1+e^{-s_k(\gamma_k - x_{kt})}}, \tag{2.2}$$

where the transition variable can be given as x_{kt}. As we consider the empirical likelihood, the assumption of the error term ε_t is relaxed from the normal or any other distributions, and this study only assumes $E(\varepsilon_t) = 0$.

3 Bayesian Empirical Likelihood

3.1 Constructing Empirical Likelihood

Since the distribution of errors ε_t in Eq.(2.1) is unspecified, the likelihood function is unavailable. Therefore, it is necessary to find an appropriate likelihood. In this study, adopted is the empirical likelihood (EL) of Owen [18] as an alternative parametric likelihood in our Bayesian estimation. This section of the study will briefly discuss the concept of empirical likelihood, and its relationship with estimating functions.

Let $p_1, ..., p_T$ be the set of probability weights allocated to the data and $\theta \in \left\{\beta_1{}^-, ..., \beta_K{}^-, \beta_1{}^+, ..., \beta_K{}^+, \alpha, \gamma_1, ..., \gamma_K, s_1, ..., s_K\right\}$. It carries a lot of information about the stochastic properties of the data. Then, let $\mathrm{x}_{kt}^- = x_{kt}(1 - F(\gamma_k, s_k))_-$ and $\mathrm{x}_{kt}^+ = x_{kt}F(\gamma_k, s_k)_+$, the empirical likelihood for estimated parameter in Eq.(2.1) , in the spirit of Owen [19], is

$$EL(\theta) = \max \prod_{t=1}^{T} p_t. \tag{3.1}$$

By taking logarithm Eq.(3.1), we have

$$EL(\theta) = \max \sum_{t=1}^{T} \log p_t, \tag{3.2}$$

where the maximization is subject to the constraints.

$$\sum_{t=1}^{T} p_t \frac{\partial m(X_{it}^-, X_{it}^+; \theta)}{\partial \theta} (y_t - m(X_{it}^-, X_{it}^+; \theta)) = 0, \tag{3.3}$$

$$\sum_{t=1}^{T} p_t = 1, \tag{3.4}$$

where, $m(X_{it}^-, X_{it}^+; \theta) = \beta_1{}^- x_{kt}^- + \beta_1{}^+ x_{kt}^+ +, ..., +\beta_K{}^- x_{Kt}^- + \beta_K{}^+ x_{Kt}^+ + \alpha Z_t$.

Sometime the high dimensionality of the parameter space $(\theta, p_1, ..., p_T)$ makes the above maximization problem difficult to solve and leads to expressions which are hard to maximize. Instead of maximizing $EL(\theta)$ with respect to the parameters $(\theta, p_1, ..., p_T)$ jointly, we use a profile likelihood.

The empirical likelihood, Eq.(3.2), is a crucial constrained profile likelihood where the constraints are defined in Eqs.(3.3) and (3.4), respectively. Then, we can maximize the empirical likelihood at each candidate parameter value θ to obtain the optimal p_t. Suppose, we know θ, then the study can write the empirical likelihood as

$$\mathrm{EL}(\theta, p_1, ..., p_T) = EL(p_1, ..., p_T). \tag{3.5}$$

We then maximize this profile empirical likelihood to obtain $(p_1, ..., p_T)$. By conducting the Lagrange multipliers, the study can maximize the empirical likelihood in Eq.(3.2) subject to the constraints in Eqs.(3.3) and (3.4) as

$$\begin{aligned} L(p, \lambda_0, \lambda_1) = \textstyle\sum_{t=1}^{T} \log(p_t) + \lambda_0(\sum_{t=1}^{T} p_t - 1) + \lambda'{}_1 \\ \textstyle\sum_{t=1}^{T} p_t \frac{\partial m(X_{it}^-, X_{it}^+; \theta)}{\partial \theta}(Y_t - m(X_{it}^-, X_{it}^+; \theta)), \end{aligned} \tag{3.6}$$

where $\lambda \in \mathbb{R}$ is the Lagrange multipliers. It is a straightforward exercise to show that the first order conditions for Ł with respect to p_t, and setting the derivative to zero, the study can find that $\lambda_0 = -T$, and by defining $\lambda = -T\lambda_1$, the study obtains the optimal p_t as

$$p_t = \frac{1}{T}\left(1 + \lambda' \frac{\partial m(X_{it}^-, X_{it}^+; \theta)}{\partial \theta}(Y_t - m(X_{it}^-, X_{it}^+; \theta))\right)^{-1} \tag{3.7}$$

Then, substituting the optimal p_t into the empirical likelihood in Eq.(3.5) the study obtains

$$EL(\theta) = \max \prod_{t=1}^{T} \frac{1}{T}\left(1 + \lambda' \frac{\partial m(X_{it}^-, X_{it}^+; \theta)}{\partial \theta}(Y_t - m(X_{it}^-, X_{it}^+; \theta))\right)^{-1}. \tag{3.8}$$

By taking logarithm, the study gets

$$logEL(\theta) = \sum_{t=1}^{T} \log(1 + \lambda' \frac{\partial m(X_{it}^-, X_{it}^+; \theta)}{\partial \theta}(Y_t - m(X_{it}^-, X_{it}^+; \theta)) - T log(T). \tag{3.9}$$

Computing the profile empirical likelihood at θ involves two step estimations. Firstly, it is important to solve a nonlinear optimization to obtain p_t,λ, and $EL(\theta_i)$ which depends on θ_i. Second step, the profile empirical likelihood is then maximized with respect to candidate θ_i. Then, the study proposes another candidate θ_i to repeat the first step again. After $EL(\theta_i)$ is computed for all candidates θ_i, the maximum value of $EL(\tilde{\theta}_i)$ is preferred. The maximization problem can now be represented as the problem of minimizing $Q(\lambda)$

$$Q(\lambda) = -\sum_{t=1}^{T} \log(1 + \lambda' \frac{\partial m(X_{it}^-, X_{it}^+; \theta)}{\partial \theta}(Y_t - m(X_{it}^-, X_{it}^+; \theta)). \tag{3.10}$$

Subject to $0 \leq p_t \leq 1$, that is

$$1 + \lambda' \frac{\partial m(X_{it}^-, X_{it}^+; \theta)}{\partial \theta}(Y_t - m(X_{it}^-, X_{it}^+; \theta)) \geq 1/T \tag{3.11}$$

To compute θ, one uses a nested optimization algorithm where the outer maximization loop with respect to θ encloses the inner minimization loop with respect to λ. Some comments on the inner loop and the outer loop are in order. In the application study, the number of all possible θ_i can be so large that it becomes infeasible and insensible to evaluate them all. Thus, we can employ a standard least square estimator to get the estimated θ_{LS} and specify the sensible range of candidate $\theta_i = [-2\theta_{LS}, 2\theta_{LS}]$.

3.2 Bayesian Empirical Likelihood for Smooth Kink Regression

The posterior distribution consists of the estimation of empirical likelihood function and the prior distribution. It is derived using Bayes rule. Let (Ω, A, P) be a probability space. Let $A_n, n \geq 1$, be a countable, measurable partition of Ω, and $B \in A$ be an event with $P(B) > 0$.

Then, for any $n \geq 1$,

$$P(A_n | B) = \frac{P(B|A_n)P(A_n)}{\sum\limits_{j=1}^{\infty} P(B|A_j)P(A_j)} \tag{3.12}$$

Indeed, we have

$$P(A_n | B) = \frac{P(A_n \cap B)}{P(B)} = \frac{P(B|A_n)\ P(A_n))}{P(B)}. \tag{3.13}$$

And writing

$$B = B \cap \Omega = B \cap (\cup_{j=1}^{\infty} A_j) = U_{j=1}^{\infty}(B \cap A_j). \tag{3.14}$$

The study has

$$P(B) = \sum_{j=1}^{\infty} P(B \cap A_j) = \sum_{j=1}^{\infty} P(B|A_j)P(A_j) \tag{3.15}$$

Consider the discrete case, let X, Y be discrete random variables. Then

$$P(X = x|Y = y) = \frac{P(Y = y|X = x)P(X = x)}{\sum_{x'} P(Y = y|X = x')P(X = x')}, \tag{3.16}$$

is the conditional density of X given Y. Here, $P(X = x)$ denotes prior density of X, while $P(Y = y)$ denotes empirical likelihood density. Thus, $P(X = x|Y = y)$is the posterior density and the study can rewrite Eq.(3.16) as

$$P(\theta | Y, X) \propto EL(\theta) \cdot \pi(\theta), \tag{3.17}$$

where $\pi(\theta)$denotes a prior density of each estimated parameter.

To estimate the posterior distribution in Kink regression models, Yang and He [15] suggested that the value of the empirical likelihood is relatively easy to compute given θ which makes the MetropolisHastings algorithm of Hastings [20] feasible for sampling from the posterior. However, it remains for this study to derive the conditional posterior distribution for unknown parameter θ. If we select a proper prior, the posterior in Eq.(3.17) is also proper. In this case, it is necessary to first find a working likelihood, and then a BEL approach is developed to make inference on Kink regression.

For any proposed θ, its profile empirical likelihood ratio is given by

$$R(\theta) = \arg\max \left\{ \prod_{t=1}^{T} (T p_t) \middle| \sum_{t=1}^{T} p_t (Y_t - m(X_{it}^{-}, X_{it}^{+}; \theta)) = 0, p_t \geq 0, \sum_{t=1}^{T} p_t = 1 \right\}. \tag{3.18}$$

By a standard Lagrange multiplier, the study obtains optimal

$$p_t = \frac{1}{T} \left(1 + \lambda_n(\theta)'(Y_t - m(X_{it}^{-}, X_{it}^{+}; \theta))\right)^{-1}. \tag{3.19}$$

Again, substituting the optimal p_t into the empirical likelihood ratio in Eq.(3.18) we obtain

$$R(\theta) = \max \prod_{t=1}^{T} \frac{1}{T} \left(1 + \lambda_n(\theta)'(Y_t - m(X_{it}^{-}, X_{it}^{+}; \theta))\right)^{-1}, \tag{3.20}$$

where $\lambda_n(\theta)$ satisfies the following equation:

$$\sum_{t=1}^{T} \frac{(Y_t - m(X_{it}^{-}, X_{it}^{+}; \theta))}{1 + \lambda(\theta)'(Y_t - m(X_{it}^{-}, X_{it}^{+}; \theta))} = 0. \tag{3.21}$$

Thus, the empirical likelihood function of $\theta$is given by $EL(\theta) = \log R(\theta)/T^T$. The study can consider this $EL(\theta)$ as the likelihood in the posterior density in Eq.(3.17). Note that to compute θ, one uses a mixed algorithm where the outer MCMC loop with respect to θ encloses the other inner optimization loop with respect to λ.

This BEL computation is inspired by Yang and He [15]. They suggested that the value of the empirical likelihood ratio $R(\theta)$ is easy to estimate at given θ and so is the Metropolis-Hastings algorithm for sampler the parameter in the posterior. In this study, we choose the priors as follows.

We take $\theta = \{\beta^{-}, \beta^{+}\}$ to be normally distributed with mean θ_0 and variance 0.01, γ_k and s_k are assumed to have uniform distribution. Hence, the conditional posteriors of Θ, γ_k, and s_k can be computed as in the following:

1)The conditional posterior distribution for Θ is

$$\theta^* = \left(\frac{X'X}{(\varepsilon'\varepsilon/n)^2} + 0.01 \right)^{-1} \times \left(\frac{X'X}{(\varepsilon'\varepsilon/n)^2} \tilde{\theta} + 0.01(\theta_0) \right), \tag{3.22}$$

where $\tilde{\theta} = (X'X)^{-1} X'Y$. And $X = \left\{X_{it}^{-}, X_{it}^{+}\right\}$.

2) The conditional posterior distribution for γ_k can be written as

$$P(\gamma_k \,|\Theta, Y, X) = \sum_{1} EL(\theta \,|Y, X) \cdot \pi(\theta), \tag{3.23}$$

3) The conditional posterior distribution for s_k can be written as

$$P(s_k\,|\theta,Y,X\,) = \sum_{1} EL(\theta\,|Y,X\,)\cdot\pi(\theta), \tag{3.24}$$

To sample all of these parameters based on conditional posterior distribution, the study employs the Markov chain Monte Carlo, Metropolis-Hastings algorithm for obtaining the sequence of parameter samples from fully conditional distributions. The study runs the Metropolis-Hastings algorithm for 20,000 iterations where the first 5,000 iterations serve as a burn-in period. For Metropolis-Hastings algorithm, the study applies it to find Kink value γ_k where the acceptance ratio is

$$r = \frac{EL(\theta^*\,|Y,X)\pi(\theta_{i-1}\,|\theta^*\,)}{EL(\theta_{i-1}\,|Y,X)\;\pi(\theta^*\,|\theta_{i-1})} \tag{3.25}$$

Then, we set

$$\theta_j = \begin{cases} \theta_{j-1} & if \quad U < r \\ \theta_j^* & if \;\; U > r \end{cases}, \tag{3.26}$$

where θ_{j-1} is the estimated vector of parameter at $(j-1)^{th}$ draw and θ_j^* is proposal vector of parameters which are generated from normal distribution $N(\theta_{j-1}, 0.01)$. U is Uniform(0,1). This means that if the proposal θ_j^* looks good, keep it; otherwise, keep the current value θ_{j-1}. By using a MetropolisHastings algorithm, the study estimates the parameters using the average of the Markov chain on θ as an estimate of θ, when the posterior density is likely to be close to normal and the trace of θ_j looks stationary.

3.3 Bayes Factor for a Kink Effect

Since the nonlinear structure model has been employed in this study, the Bayes factor, which is a reliable testing procedure in model comparison and Kink effect test in the Bayesian approach, is developed here. The purpose of this Bayes factor is to check whether a Kink parameter does significantly exist or not. It can be used to assess the models of interest namely linear and Kink regression, so that the best fit model will be identified given a data set and a possible model set. In other words, Bayes factor is a useful tool for selecting a possible model [21]. Bayes factor is used for the ratio of the posterior under one model to another model. In this study, we consider the linear model to be a null model denoted by M_1 and the smooth kink model to be an alternative model denoted by M_2. More specifically, Bayes factor BF is given by

$$BF = \frac{\mathrm{P}(Y,\mathrm{X}\,|M_1)}{\mathrm{P}(Y,X_{it}^-,X_{it}^+\,|M_2)} = \frac{\int EL(Y,\mathrm{X}\,|\theta_1)\pi(\theta_1\,|M_1)d\theta_1}{\int EL(Y,X_{it}^-,X_{it}^+\,|\theta_2)\pi(\theta_2\,|M_2)d\theta_2} \tag{3.27}$$

where $\mathrm{P}(Y,\mathrm{X}\,|M_1)$ and $\mathrm{P}(Y,X_{it}^-,X_{it}^+\,|M_2)$ are the posterior density of the null model and alternative model, respectively. θ_1 and θ_2 are the vector of parameters of M_1 and M_2, respectively. For choosing the appropriate model, the study follows the idea of Kass and Raftery [22]. If $log(BF) < 2, 2 < log(BF) < 6, 6 < log(BF) < 10$, and $log(BF) > 10$, there is a chance that the log(BF) favors the M_2 with weak support evidence M_1, strong evidence M_1 and decisive evidence for M_1, respectively.

4 Simulation Study: A Comparison of Some Estimation Techniques

Next, we turn our attention to investigating the performance of BEL estimator and testing whether BEL is robust. We conduct experiment with errors drawn from different distributions. Errors are generated from either normal or non-normal distributions. In this experiment, we also compare the BEL estimator to the traditional MLE, Bayesian with normal likelihood (BAY) and Maximum Empirical likelihood (MEL). In the simulation, the following equation is used to generate the dataset Y_t.

$$Y_t = \beta_0 + \beta_1{}^- x_{1t}(1 - F(\gamma_1, s_1))_- + \beta_1{}^+ x_{1t}(F(\gamma_1, s_1)_+ + \varepsilon_t, \tag{4.1}$$

where the true values for parameters α, $\beta_1{}^-$, and $\beta_1{}^+$ are $\beta_0 = 0.5$, $\beta_1{}^- = 1$, and $\beta_1{}^+ = -1$, respectively. The threshold value is $\gamma_1 = 3$ and smooth parameter value is $s_1 = 20$. The covariate x_{1t} is independently generated from the standard normal distribution $N(\gamma_1, 1)$ to guarantee that γ_1 is located in x_{1t}. To make a fair comparison, the study considers the following random errors ε_t: (i) $N(0,1)$, (ii) $t(0,1,4)$ and (iii) $Unif(-2,2)$. In this Monte simulation study $N = 50$ and $N = 100$. Then, the performance of these estimators are evaluated through the Bias which is given as

$$Bias = \left| M^{-1} \sum_{m=1}^{M} (\tilde{\theta}_m - \theta_m) \right|, \tag{4.2}$$

where $M = 100$ is the number of replications; and $\tilde{\theta}_m$ and θ_m are , respectively, the estimated parameters and their true parameter values. The simulation results are shown in Tables 1, 2 and 3, respectively.

In Tables 1, 2 and 3, we report bias of parameters under 4 different estimations with sample sizes N=50 and N=100. Each Table is arranged in 12 rows and 5 columns corresponding to the Bias of parameters and estimations, respectively. The traditional MLE and BAY outperform MEL and BEL estimators for both sample sizes when the error term is assumed to have normal distribution. As we know that the MLE and BAY are based on the normal likelihood, thus, it is not surprising that these two estimators perform better than the estimation based on empirical distribution.

Next, we investigate the performance of the BEL estimator when the errors are generated from some non-normal distributions, namely$t(0,1,4)$ and $U(-2,2)$. The estimation results for the $t(0,1,4)$case are reported in Table 2. The overall result is different from the normal error case. MLE and BAY estimators do not have a lower Bias than BEL estimator. We observe the higher Bias of MLE and BAY estimators for both sample sizes. Similar to the student-t distribution case, we also observe the higher performance of MEL and BEL estimators when the errors are generated from uniform distribution. According to these comparisons, we find that BEL and MEL yield a similar bias result. However, when we examine the effect of sample size on these two estimators, the performance of BEL estimator seems to be better than MEL, particularly when N=50.

Table 1: Smooth transition kink regression with $N(0,1)$ errors

	MEL	**BEL**	**MLE**	**BAY**
$N=50$	Bias	Bias	Bias	Bias
α	0.1444	0.1112	0.041	0.0436
$\beta_1{}^-$	0.0484	0.0403	0.0307	0.0379
$\beta_1{}^+$	0.0916	0.0873	0.0012	0.0329
γ_1	0.7727	0.6208	0.1261	0.0252
s_1	0.1698	0.1341	0.0515	0.0352
$N=100$	**Bias**	**Bias**	**Bias**	**Bias**
α	0.0232	0.0225	0.0116	0.0182
$\beta_1{}^-$	0.0234	0.0129	0.0015	0.0068
$\beta_1{}^+$	0.0421	0.0446	0.0116	0.003
γ_1	0.0516	0.0742	0.0259	0.0232
s_1	0.0648	0.0648	0.0491	0.0078

Table 2: Smooth transition kink regression with $t(0,1,4)$ errors

	MEL	**BEL**	**MLE**	**BAY**
$N=50$	Bias	Bias	Bias	Bias
α	0.0359	0.0199	0.1045	0.3402
$\beta_1{}^-$	0.0178	0.0101	0.0133	0.0278
$\beta_1{}^+$	0.0026	0.0033	0.0039	0.0236
γ_1	0.024	0.0262	0.2067	0.1911
s_1	0.0103	0.0398	0.2051	0.1842
$N=100$	**Bias**	**Bias**	**Bias**	**Bias**
α	0.0275	0.1536	0.0507	0.0266
$\beta_1{}^-$	0.0193	0.0643	0.0046	0.0182
$\beta_1{}^+$	0.0187	0.0013	0.0019	0.0107
γ_1	0.0287	0.0171	0.0604	0.0326
s_1	0.0301	0.0161	0.0139	0.0469

This simulation results allow us to conclude that the performance of BEL and MEL are better than the conventional MLE and BAY when the data are generated from normal distribution. However, the performance of the BEL estimator in the model performs well over a wide range of error distributions as it produces an acceptable Bias values compared with MLE, BAY, and MEL when the errors are generated from both normal and non-normal distributions. In addition, BEL is found to outperform the competing estimators with small sample size under various error distributions.

Table 3: Smooth transition kink regression with $Unif(-2,2)$ errors

	MEL	**BEL**	**MLE**	**BAY**
$N = 50$	Bias	Bias	Bias	Bias
α	0.0178	0.0117	0.0318	0.0402
${\beta_1}^-$	0.0167	0.0119	0.068	0.0233
${\beta_1}^+$	0.0138	0.0258	0.0371	0.0165
γ_1	0.0282	0.015	0.0396	0.0466
s_1	0.1692	0.0488	0.0807	0.0953
$N = 100$	**Bias**	**Bias**	**Bias**	**Bias**
α	0.0714	0.0119	0.0136	0.041
${\beta_1}^-$	0.0193	0.0166	0.0044	0.0234
${\beta_1}^+$	0.0331	0.0017	0.0089	0.0133
γ_1	0.0024	0.0012	0.0075	0.05
s_1	0.0490.	0.0084	0.0033	0.0127

5 Example Data

This section illustrates the performance of the proposed smooth kink regression and BEL estimation when applied to a real data set. The considered data lasts 39 years of consumer price index (CPI) and unemployment rate (UNE). We collect the data from 1979 to 2017. We first plot the raw data in Figure 1. These plots motivate our empirical study, as there exhibits the structural change along the sample period. The movement of CPI seems to have descending trend before 2000 before changing to be the ascending trend. Likewise, the movement of UNE series is low along 1980-1990, and then it had substantially increased and reached the highest value in 2000. We then suspect that there may involve a regime change in this dataset. The data description is provided in Table 4.

5.1 Data Analysis

In the application study, the number of all possible θ_i can be so large that it becomes infeasible and insensible to evaluate them all. Thus, here we can employ a standard least squares estimator to get the estimated θ_{LS} and specify the sensible range of candidate $\theta_i = [-2\theta_{LS}, 2\theta_{LS}]$. This study investigates the basic question whether the CPI had an effect on the UNE. Motivated by this, we fit our model for answering this basic question, which is given as follows:

$$UNE_t = \beta_0 + \beta_1{}^{-}CPI_t(1 - F(\gamma_1, s_1))_{-} + \beta_1{}^{+}CPI_{1t}(F(\gamma_1, s_1)_{+} + \varepsilon_t. \tag{5.1}$$

Table 4: Data Description

	UNE	CPI
Mean	3.544445	1.07085
Median	3.341667	0.595587
Maximum	5.541667	7.778582
Minimum	2.016667	-1.35284
Std. Dev.	1.158324	1.804973
Skewness	0.270713	1.590211
Kurtosis	1.67552	6.248693
Observations	39	39

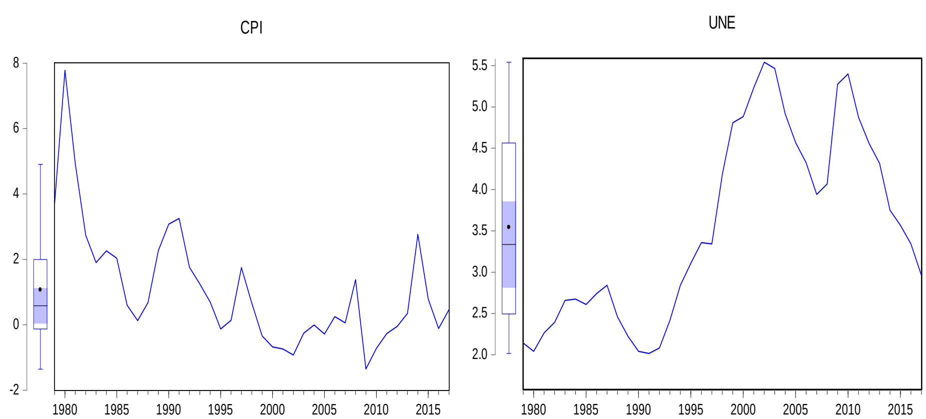

Figure 1: Annual CPI and UNE of Japan, 1979-2017

Prior to estimating the smooth kink regression model, we need to confirm the nonlinear behavior in our model. In this study, therefore, nonlinear structure test is conducted for our data. Bayes factor test is used to determine whether our model appears to have a nonlinear behavior. Using the Bayes factor formula, the result shown in Table 5 provides the values of Bayes factor of these two models in which we can observe that the value of $\log(BF)$ is equal to -0.3287. This result means the smooth kink model M_2 is more anecdotally supported by the data under consideration than the modelM_1, and hence, the data is more likely to have the nonlinear structure. This result suggests rejecting the null hypothesis of linear regression M_1 and accepting the alternative hypothesis of smooth kink regression.

Table 5: Bayes factor of Kink effect

	$P(Y,X\|M_1)$	$P(Y,X\|M_2)$	**BF**	$\log(BF)$	**Interpret**
Regime 1 vs. Regime 2	121.5668	168.8735	0.7198	-0.3287	Support M2

5.2 Result

Table 6: Coefficients (standard errors) from Kink regression

Parameter	**Japan**
α	4.2428 (-0.0845)
$\beta_1{}^{-}$(regime 1)	-0.1845 (-0.3581)
$\beta_1{}^{+}$(regime 2)	-0.3064 (-0.0473)
γ_1	1.3156 (-0.7345)
s_1	0.723 (-0.1317)
Acceptance rate	0.7031

The smooth kink regression model is then estimated by BEL estimator, and the estimated results are shown in Table 6. In this study, we can interpret regime 1 and 2 as low and high CPI regime, respectively. Although, CPI is found to have a negative relationship with UNE, the effects of CPI on UNE are not the same. We observe that CPI shows negative coefficient (-0.1845) in regime 1 and (-0.3064) in regime 2. Consider the

Kink or threshold point (γ_1), the value of $\gamma_1 = 1.3156$. Figure 2 depicts the histograms based on the MCMC Metropolis hasting draws, which gives some basic insight into the geometry of the posteriors obtained in this application analysis. The results show a good convergence behavior and it seems to converge to the normal distribution; thus, we can get accurate posterior inference for parameters that appear to have good mixing. In addition, the estimation of all 15,000 parameters in posterior draws do not contain zero, indicating that there is strong evidence of the significant parameter estimates.

Figure 3 plots the smooth kink line of our model. The result illustrates a steep negative curve for CPI with a kink point around 1.3156, changing to a low negative slope above that point.

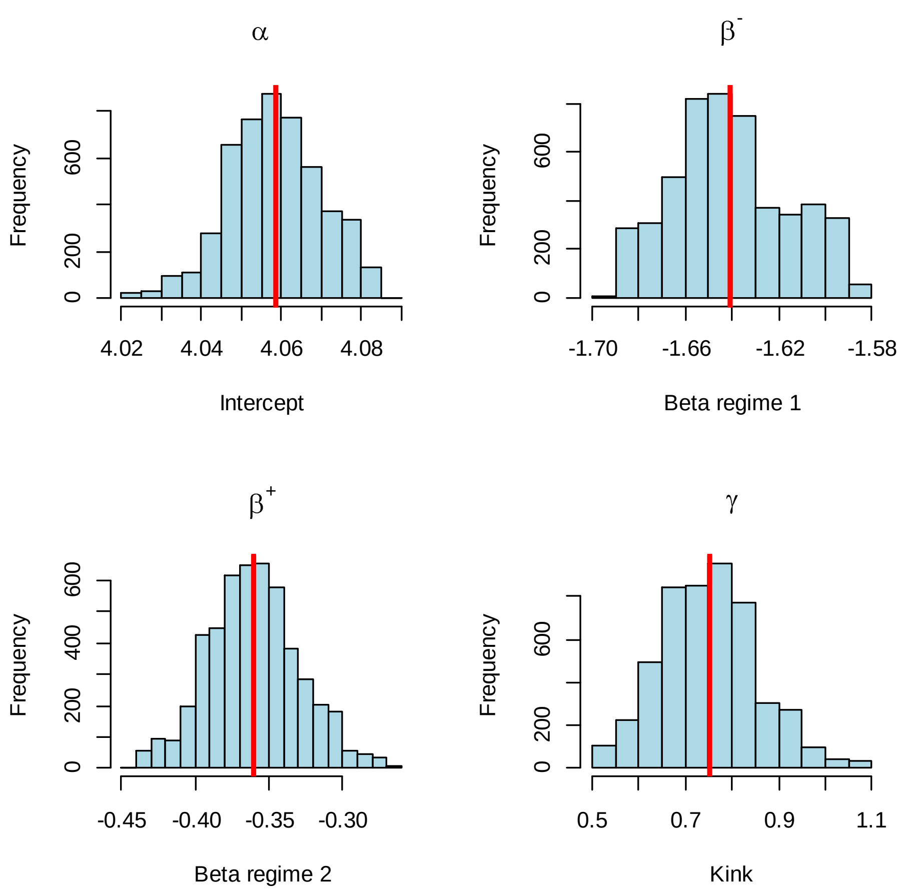

Figure 2: Histograms based on the MCMC draws

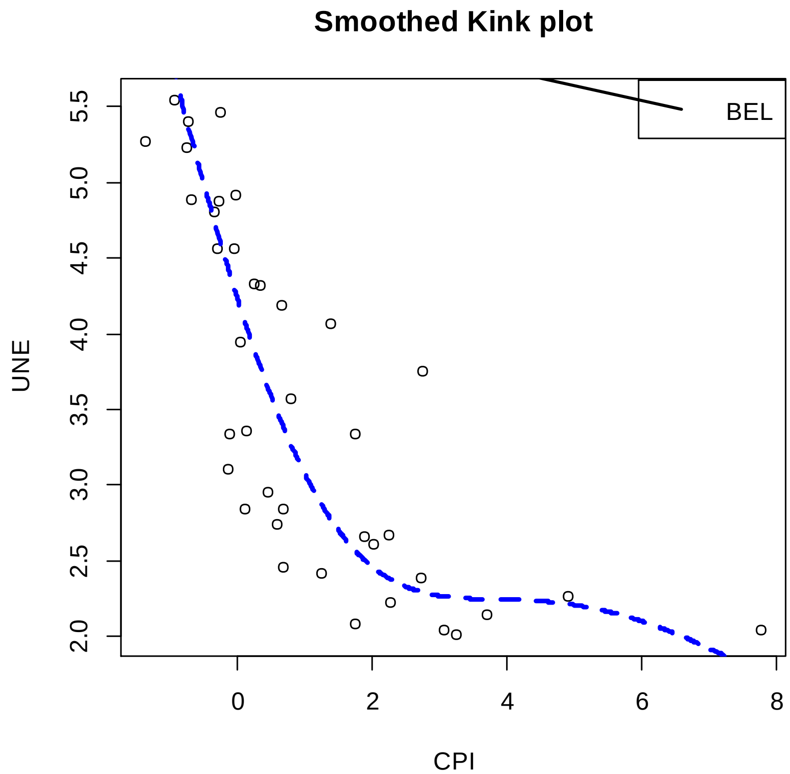

Figure 3: Kink plot of UNE and CPI

6 Conclusion

In this study, we proposed a new nonlinear model called smooth kink regression. We consider the Bayesian empirical likelihood as the estimator of our model. This approach can relax a strong assumption of normality and the limited data. Although the idea of BEL approach is not new, this study provides an important addition to the literature by employing this approach for smooth kink regression model.

The study then conducts a simulation study to show the performance and accuracy of BEL estimation for our model. The results of the simulation study confirm that BEL estimator can give an accurate result for all unknown parameters. The overall result reveals that our BEL estimator applied to our model performs well over a wide range of error distributions. The BEL estimator produces acceptable Bias values compared with MLE, Bayesian and MEL when the errors are generated from both normal and non-normal distributions. In addition, BEL can outperform the competing methods with small data under various error distributions and small sample size.

Finally, the empirical results demonstrate that the consumer price index provides a negative effect on unemployment rate for high and low CPI regimes, respectively. Although, the effects of consumer price index on unemployment rate are the same for both regimes, but the size of the effect of regime 1 is larger than regime 2, confirming the usefulness of our proposed model under the structural change of the relationship between predictor and dependent variable.

Acknowledgements : We are grateful for financial support from Centre of Excellence in Econometrics, Faculty of Economics, Chiang Mai University. We thank Nguyen, H. T. for valuable comment to improve this paper.

References

[1] Card, D., Lee, D. S., Pei, Z., and Weber, A. (2015). Inference on causal effects in a generalized regression kink design. Econometrica, 83(6), 2453-2483.

[2] Hansen, B. E. (2017). Regression kink with an unknown threshold. Journal of Business and Economic Statistics, 35(2), 228-240.

[3] Sriboochitta, S., Yamaka, W., Maneejuk, P., and Pastpipatkul, P. (2017). A generalized information theoretical approach to non-linear time series model. In Robustness in Econometrics (pp. 333-348). Springer, Cham.

[4] Golan, A. (2002). Information and entropy econometricseditor's view. Journal of Econometrics, 107(1), 1-15.

[5] Heckelei, T., Mittelhammer, R. and Jansson, T. (2008). A Bayesian alternative to generalized cross entropy solutions for underdetermined econometric models. Food and Resource Economics Discussion Paper, 2.

[6] Stein, C. (1955), Inadmissibility of the Usual Estimator for the Mean of Multivariate Normal Distribution, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1, 197-206.

[7] Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., ... and Smith, P. G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. British journal of cancer, 35(1), 1.

[8] Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., and Munaf, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience, 14(5), 365-376.

[9] Van de Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., and Van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. European Journal of Psychotraumatology, 6.

[10] Lee, S. Y., and Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. Multivariate Behavioral Research, 39(4), 653-686.

[11] Zhang, Y. Q., and Tang, N. S. (2017). Bayesian empirical likelihood estimation of quantile structural equation models. Journal of Systems Science and Complexity, 30(1), 122-138.

[12] Schennach, S. M. (2005). Bayesian exponentially tilted empirical likelihood. Biometrika, 92(1), 31-46.

[13] Fang K T and Mukerjee R. (2006). Empirical-type likelihoods allowing posterior credible sets with frequentist validity: Higher-order asymptotics, Biometrika, 93, 723733.

[14] Chang I H and Mukerjee R (2008). Bayesian and frequentist confidence intervals arising from empirical type likelihoods, Biometrika, 95. 139147.

[15] Yang, Y., and He, X. (2012). Bayesian empirical likelihood for quantile regression. The Annals of Statistics, 40(2), 1102-1131.

[16] Chib, S., Shin, M., and Simoni, A. (2016). Bayesian empirical likelihood estimation and comparison of moment condition models (No. 2016-21).

[17] Yamaka, W., Pastpipatkul, P., and Sriboonchitta, S. (2018). Bayesian Empirical Likelihood Estimation for Kink Regression with Unknown Threshold. In International Conference of the Thailand Econometrics Society (pp. 752-766). Springer, Cham.

[18] Owen, A. B. (1998). Empirical Likelihood. John Wiley and Sons, Inc.

[19] Owen, A. (1990). Empirical likelihood ratio confidence regions. The Annals of Statistics, 90-120.

[20] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97109.

[21] Geweke, J., and Terui, N. (1993). Bayesian threshold autoregressive models for nonlinear time series. Journal of Time Series Analysis, 14(5), 441-454.

[22] Kass, R. E., and Raftery, A. E. (1995). Bayes factors. Journal of the american statistical association, 90(430), 773-795.

(Received 30 November 2018)
(Accepted 7 February 2019)