Int. J. Nonlinear Anal. Appl. 12 (2021) No. 2, 743-750
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2021.5125

Comparison between Renyi GME and Tsallis GME for estimation Kink regression

Saad Obaid Jameela[a,*], Basim Shlaibah Msallamb[b]

[a]Baghdad University, College of Administration and Economics, Baghdad, Iraq
[b]Al-Iraqia University, College of Administration and Economics, Baghdad, Iraq

(Communicated by Madjid Eshaghi Gordji)

Abstract

In this paper, an estimation of the parameters of the Kink regression model was made by relying on higher order from the general maximum entropy method with two measures Tsallis and Renyi ($\alpha = 3$, $\alpha = 4$, $\alpha = 5$, $\alpha = 6$), A practical application has been made to data for a real phenomenon in the Iraqi economy, which is inflation with the Debt/GDP ratio and a statistical analysis of it. After making a comparison with other estimation methods, as the results showed that the explanatory variable for the kink point is equal to (5), and the results show a decrease in the Debt/GDP ratio after observations Kink.

Keywords: GME , Kink regression, Renyi, Tsallis, Entropy, Debt/GDP

1. Introduction

The Kink Regression method is one of the important techniques in statistical analysis. This statistical method works with real phenomena in which the marginal slope is not continuous when one or more observations of the explanatory variable are observed, although the regression function is continuous with the explanatory variables that are related to the fixed slope and the constant term parameter. And that this observation, whose marginal slopes are not continuous, represents the cut off point (cut off or kink). Regular analysis represented by the normal regression of the phenomenon data that has a kink point because this leads to inaccurate results and this requires us to employ kink to do statistical analysis of any phenomenon that has a kink point in order to obtain accurate and efficient results for the regression model. And that the Kink Regression model is a

*Corresponding author

Email addresses: saad.iraq100@gmail.com (Saad Obaid Jameela), drbasimstatistics@gmail.com (Basim Shlaibah Msallamb)

Received: December 2020 *Accepted:* May 2021

broader and general method for the discontinuous regression model, and with the great importance of it in the statistical analysis of most economic and social phenomena ... etc. However, most of the research that was dealt with was using traditional methods, as the most famous of which was the Least squares method (LS) [6]. As for the two researchers (Tarkhamtham and Yamaka) [8, 9], they used the general great entropy method, depending on the Renyi and Tsallis measures. As for the researchers (Böckerman, Kanninen, Suoniemi, 2018) [1] they talked about the semi parametric Kink regression model and the estimation of its parameters in the Kernal method. By using the general maximum entropy method with the Renyi and Tsallis measure and with different order, the parameters of the kink regression model will be estimated in this research by applying it to data for a real phenomenon in the Iraqi economy that represents the phenomenon of inflation and the Debt/GDP ratio, and then a comparison between these estimates will be made to find the best method. The second topic of the research included the type of regression model (kink) and some of its characteristics, as well as the general maximum entropy method by adopting the Renyi measure and the general maximum entropy method by adopting the Tsallis measure, while the third section included the results of the analysis of the real phenomenon data on the Iraqi economy represented by the phenomenon of inflation and what is its impact on Debt/GDP ratio.

The model

The general model for kink regression, which contains kink points for a number of explanatory variables at once, and the number of (K) points from the kink is as follows [9]:

$$Y_i = \beta_1^-(x_{1i} - \lambda_1) + \beta_1^+(x_{1i} - \lambda_1) + \cdots + \beta_k^-(x_{ki} - \lambda_k) + \beta_k^+(x_{ki} - \lambda_k) + \beta_3 Z_0 + \epsilon_i \tag{1.1}$$

$$i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, k, \quad \acute{\beta} = \begin{bmatrix} \beta_1^- & \beta_1^+ & \beta_a \end{bmatrix}$$

As Y_i is the view value (i) from the observations of the response variable, X_{ji}: is the view value (i) from the views of the explanatory variable j. Z_0 : is the value of the observation (i) from the observations of the explanatory variable associated with the constant term and its value is always the same,$(\lambda_1, \ldots, \lambda_k)$: represent the parameters of the kink point [5].

From equation (1.1) we notice that one of its most important details is that it contains two relationships. The first is a linear relationship between the response variable and the explanatory variable associated with the fixed slope Z_0, which is neglected for ease of dealing with the estimation process and in this relationship the regression function is continuous, while the second type of relationship is nonlinear between the response variable and the explanatory variables that have $(\lambda_1, \ldots, \lambda_k)$ that represent Kink Points and in this relationship the response variable is continuous, while the marginal slope for it are discontinuous at the Kink point, which represents one of the observations of the explanatory variable on it, for the marginal slopes associated with the explanatory variables that have A point (Kink) has two values before and after Kink (Hansen: 2017) denoted by a negative sign$(-)$ for the first value of the boundary slope before the Kink, and by a positive sign $(+)$ for the value of the marginal slope at the values of the observations of the explanatory variable after Kink [4, 7].

Since the real data that will be analyzed has only one explanatory variable, equation (1.1) will be written for one explanatory variable having only one Kink as follows:

$$Y_i = \beta_1^-(x_1 - \lambda)^- + \beta_1^+(x_1 - \lambda)^+ + \beta_a Z_0 + \epsilon_i \tag{1.2}$$

It is noted from the equation (1.2) that it is difficult to write it in the traditional regression form when writing for (n) of the equations. The reason for this is the difficulty of determining which of the observations of the explanatory variable that Kink possesses. Therefore, the method of the researcher (Hansen: 2017) was relied on in writing the equation (1.2) In terms of matrices as follows [6]:

$$Y_i = \acute{\beta} X_{i(\lambda)} + \epsilon_i, \quad i = 1, \dots, n \tag{1.3}$$

As $\acute{\beta} = \begin{bmatrix} \beta_1^- & \beta_1^+ & \beta_a \end{bmatrix}, \quad \acute{X}_{i_{(\lambda)}} = \begin{bmatrix} (X_{1i} - \lambda)^- & (X_{1i} - \lambda)^+ & Z_0 \end{bmatrix}$

The test of the continuity or discontinuity hypothesis of the Kink regression model is one of the main steps that must be performed before estimating the parameters of equation (1.3), according to what the researcher said (Hansen: 2017), as he showed that the continuity test step is difficult to implement in Kink's regression, so it will be assumed that the marginal slope function The present in the Kink regression equation does not have continuity according to the explanatory variables associated with it and without going to conduct the test, and this method is what will also be adopted in this study when conducting an estimation of the parameters of equation (1.3) as well as determining the value of the Kink point.

2. Generalized Maximum Entropy Method

The process of estimating the regression parameters, whether linear or nonlinear regression, using the general maximum entropy method, by directly reparametrized as well as for errors, as well as directly reparametrized Kink . This is done in the form of predictions for random variables or in a convex combination. In terms of support points, which are denoted by the symbol (S) and that their value is specified within the following period $(2 \leq S \leq 7)$ and that their number is determined according to the researcher's own viewpoint, and in some times the type of the data or what imposes them [6, 2] In this paper, the case of $(S = 5)$will be dealt with in the theoretical and practical side as follows [4]:

$$\begin{aligned} Z_i &= \left[-c, -\frac{c}{2}, 0, \frac{c}{2}, c\right] \\ q &= \left[-c, -\frac{c}{2}, 0, \frac{c}{2}, c\right] \\ v_i &= \left[-c, -\frac{c}{2}, 0, \frac{c}{2}, c\right] \end{aligned} \tag{2.1}$$

where (c) is a positive default numerical value between $(0, \infty)$, and z_j: is a vector of rank (1xs) which represents support points for the model parameters β in addition to the constant term parameter, q represents a vector of order $(1 \times s)$ to points Support for the Kink parameter, while v_i: is a vector of order $(1 \times s)$ and represents the support points for the random error vocabulary.

In this paper, the parameters of equation (1.3) will be estimated in two ways, the first being the general maximum entropy method (GME) with the Tsallis measure and the general maximum entropy method on the Renyi measure [9] and then applying that to real phenomenon data and comparing the estimation results as follows:

3. Tsallis Measure

The entropy function is defined according to the Tsallis measure as follows [2, 8]:

$$H_\alpha^T = \frac{\sum_k P_k^\alpha - 1}{1 - \alpha} \tag{3.1}$$

as α represents the order of the function, and g is a fixed positive quantity and most research tends to make it a value equal to 1 to facilitate the process of dealing with the equation (3.1), and

(Tarkhamtham and Yamaka) stated in 2018 that it is best to deal with the equation of a regression model Kink without containing the variable Z_0 and the purpose of that and as mentioned previously is to facilitate the procedure of estimating the parameters using the general maximum entropy method (GME), and accordingly, the reparametrized will be for the parameters $(\epsilon_i, \lambda, \beta_1^+, \beta_1^-)$ only, and this is done by By writing it in the form of a convex combination for a number of support points, which were previously identified with $(S = 5)$ [4, 6] as follows:

$$\beta_1^- = \sum_m p_{1m}^- z_{1m}^-, \qquad x_{1,t} \le \lambda \tag{3.2}$$

$$\beta_1^+ = \sum_m p_{1m}^+ z_{1m}^+, \qquad x_{1,t} > \lambda \tag{3.3}$$

As: p_{1m}^- It is the probability distribution of the S dimension of the observation before the Kink point, and $\left(\sum_m^5 p_{1m}^- = 1\right), p_{1m}^+$: Is the probability distribution of the S dimension of the observation after the Kink point , and $\left(\sum_m p_{1m}^+ z_{1m}^+ = 1\right)$, where $(m = 1, \dots, (S = 5))$.

As for the Kink point, it can be calculated according to the following formula:

$$\lambda = \sum_m h_m q_m \tag{3.4}$$

As: $\sum_m^5 h_m = 1$.
The same applies to the random error ϵ_i we rewrite it as a convex combination [10, 7]:

$$\epsilon_i = \sum_m \omega_{im} v_{im}, \qquad i = 1, 2, \dots, n \tag{3.5}$$

As: $\sum_m^5 \omega_{im} = 1$.
Using equations (3.1) (3.2) (3.3) (3.4) (3.5), the general maximum entropy model of the Tsallis measure can be constructed as follows [9, 10]:

$$\begin{aligned} &H^T = argmax\left\{H^T(p) + H^T(h) + H^T(\omega)\right\} \equiv \\ &\frac{1}{1-\alpha}\Big(\sum_k\sum_m p_k^{\alpha,-} - 1\Big) + \frac{1}{1-\alpha}\Big(\sum_k\sum_m p_k^{\alpha,+} - 1\Big) + \frac{1}{1-\alpha}\Big(\sum_k\sum_m h_{km}^{\alpha} - 1\Big) + \\ &\frac{1}{1-\alpha}\Big(\sum_k\sum_m \omega_{km}^{\alpha} - 1\Big) \end{aligned} \tag{3.6}$$

Equation (3.6) is subject to limitations (consistency and standard limitations) [3]:

$$Y_i = \sum_m p_{1m}^- z_{1m}^- \Big(x_{1i} - \sum_m h_{1m} q_{1m}\Big) + \sum_m p_{1m}^+ z_{1m}^+ \Big(x_{1i} - \sum_m h_{1m} q_{1m}\Big) + \sum_m \omega_{im} v_{im} \tag{3.7}$$

where $\sum_m p_{im}^- = 1, \quad \sum_m p_{im}^+ = 1, \quad \sum_m h_m = 1, \quad \sum_m \omega_{im} = 1$
Now we rewrite the lagrange function as follows:

$$L = H^T(p, h, \omega) + \gamma_1(\theta) + \gamma_2\Big(1 - \sum_m p_{1m}^-\Big) + \gamma_3\Big(1 - \sum_m p_{1m}^+\Big) + \gamma_4\Big(1 - \sum_m h_{1m}\Big) + \gamma_5\Big(1 - \sum_m \omega_{tm}\Big) \tag{3.8}$$

by put th equation (3.6) and equation (3.7) into equation (3.8), we get:

$$\begin{aligned} L &= \frac{1}{1-\alpha}\Big(\sum_k\sum_m p_k^{\alpha,-}-1\Big)+\frac{1}{1-\alpha}\Big(\sum_k\sum_m p_k^{\alpha,+}-1\Big)+\frac{1}{1-\alpha}\Big(\sum_k\sum_m h_{km}^{\alpha}-1\Big)+ \\ &\quad \frac{1}{1-\alpha}\Big(\sum_k\sum_m \omega_{km}^{\alpha}-1\Big)+\gamma_1\Bigg(Y_i-\sum_m p_{1m}^{-}z_{1m}^{-}\Big(x_{1i}-\sum_m h_{1m}q_{1m}\Big)+ \\ &\quad \sum_m p_{1m}^{+}z_{1m}^{+}\Big(x_{1i}-\sum_m h_{1m}q_{1m}\Big)+\sum_m \omega_{im}v_{im}\Bigg)+\gamma_2-\gamma_2\sum_m p_{1m}^{-}+\gamma_3- \\ &\quad \gamma_3\sum_m p_{1m}^{+}+\gamma_4-\gamma_4\sum_m h_{1m}+\gamma_5-\gamma_5\sum_m \omega_{im} \end{aligned} \tag{3.9}$$

By partially deriving equation (3.9) and finding the solution, we can find $\hat{p}_{1m}^{-}, \hat{p}_{1m}^{+}, \hat{\omega}_{tm}$ *and* $\hat{h}_{tm}$ as follows:

$$\hat{p}_{1m}^{-} = \left(\frac{1-\alpha}{\alpha}\right)\left[\sum_m \hat{\gamma}_{1m}z_{1m}^{-}\left(\acute{x}_{1i}-\sum_m h_{1m}q_{1m}\right)+\gamma_{2k}\right] \tag{3.10}$$

$$\hat{p}_{1m}^{+} = \left(\frac{1-\alpha}{\alpha}\right)\left[\sum_m \hat{\gamma}_{1m}z_{1m}^{+}\left(\acute{x}_{1i}-\sum_m h_{1m}q_{1m}\right)+\gamma_{3k}\right] \tag{3.11}$$

By performing the same previous calculation $\hat{\omega}_{tm}$ and $\hat{h}_{1m}$:

$$\hat{\omega}_{tm} = \left(\frac{1-\alpha}{\alpha}\right)\left[\sum_m \hat{\gamma}_{1m}v_{1m}+\gamma_{5k}\right] \tag{3.12}$$

$$\hat{h}_{1m} = \left(\frac{1-\alpha}{\alpha}\right)\left[\sum_m \gamma_{1m}p_{1m}^{-}z_{1m}^{-}\Big(\acute{x}_{1i}-\sum_m q_{1m}\Big)-\sum_m \lambda_{1m}p_{1m}^{-}z_{1m}^{-}\Big(\hat{x}_{1i}-\sum_m q_{1m}\Big)-\gamma_{4k}\right] \tag{3.13}$$

4. Renyi Measure

The entropy function according to the Renyi measure is defined as follows [2, 8]:

$$H_{\alpha}^{R}(x)=\frac{1}{1-\alpha}\log\sum p_k^{\alpha} \tag{4.1}$$

As α represents the order of the function, the Renyi function can be written in another form and as with the following equation:

$$H_{\alpha}^{R}=\log\left[\sum_{i=1}^{n}p_i^{\alpha}\right]^{\frac{1}{1-\alpha}}=\log\left[\sum_{i=1}^{n}p_i^{\alpha}\right]^{-\frac{1}{\alpha-1}}=-\log\left[\sum_{i=1}^{n}p_i^{\alpha}\right]^{\frac{1}{\alpha-1}} \tag{4.2}$$

Researcher Renyi has shown that entropy also represents information that has been disclosed (or demystified) after conducting a statistical analysis.
Like the previous steps that were implemented in Tsallis Measure, the parameters $(\epsilon_i, \lambda, \beta_1^{+}, \beta_1^{-})$ will be performed by rewriting them in the form of a convex combination of a number of support points

S which were considered Is equal to (5) and as written by equations (3.2), (3.3), (3.4) and (3.5). As a result of the above, the entropy function of the Renyi Measure will be as follows [10, 4] [10] [4]:

$$H^R(P,h,\omega) = H^R_\alpha(p,h,\omega) = argma\left\{H^R_\alpha(p) + H^R_\alpha(h) + H^R_\alpha(\omega)\right\} x \equiv$$
$$\frac{1}{1-\alpha}\log\sum_k\sum_m p^{\alpha,-}_{km} + \frac{1}{1-\alpha}\log\sum_k\sum_m p^{\alpha,+}_{km} + \frac{1}{1-\alpha}\log\sum_k\sum_m h^{\alpha}_{km} + \frac{1}{1-\alpha}\log\sum_k\sum_m \omega^{\alpha}_{km} \tag{4.3}$$

In the same manner that was used in the Tsallis measure, we find find $\hat{p}^-_{1m}, \hat{p}^+_{1m}, \hat{\omega}_{tm}$ *and* $\hat{h}_{tm}$ as follows:

$$\hat{p}^-_{1m} = \left(\frac{1-\alpha}{1}\right)\left[\frac{\exp\left[-z^-_{1m}\sum_i \hat{\lambda}_i\left(\acute{x}_{1i} - \sum_m h_{1m}q_{1m}\right)\right]}{\sum\exp\left[-z^-_{1m}\sum_i \hat{\lambda}_{1i}\left(\acute{x}_{1i} - \sum_m h_{1m}q_{1m}\right)\right]}\right] \tag{4.4}$$

$$\hat{p}^+_{1m} = \left(\frac{1-\alpha}{1}\right)\left[\frac{\exp\left[-z^+_{1m}\sum_i \hat{\lambda}_i\left(\acute{x}_{1i} - \sum_m h_{1m}q_{1m}\right)\right]}{\sum\exp\left[-z^+_{1m}\sum_i \hat{\lambda}_{1i}\left(\acute{x}_{1i} - \sum_m h_{1m}q_{1m}\right)\right]}\right] \tag{4.5}$$

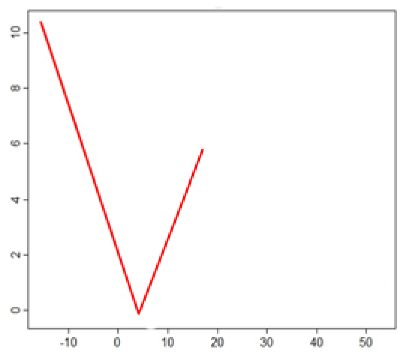
$$\hat{\omega}_{im} = \left(\frac{1-\alpha}{\alpha}\right)\left[\frac{\exp(-\hat{\gamma}_{1i}v_{1m})}{\sum\exp(-\hat{\gamma}_{1i}v_{1m})}\right] \tag{4.6}$$

$$\hat{h}_{1m} = \left(\frac{1-\alpha}{\alpha}\right)\left[\frac{\exp\left[-\sum_i \hat{\gamma}_{1i}p^-_{1m}z^-_{1m}\left(\acute{x}_{1i} - \sum_m q_{1m}\right) + \sum_i \hat{\gamma}_{1i}p^+_{1m}z^+_{1m}\left(\acute{x}_{1i} - \sum_m q_{1m}\right)\right]}{\sum\exp\left[-\sum_i \hat{\gamma}_{1i}p^-_{1m}z^-_{1m}\left(\acute{x}_{1i} - \sum_m q_{1m}\right) + \sum_i \hat{\gamma}_{1i}p^+_{1m}z^+_{1m}\left(\acute{x}_{1i} - \sum_m q_{1m}\right)\right]}\right] \tag{4.7}$$

And γ represents the Lagrange coefficient.

Case Study

In this aspect of the research, data of a real phenomenon in the Iraqi economy will be analyzed that represents inflation as an explanatory variable and the Debt/GDP ratio as a response variable and that the data collected from Ministry of Planning - Iraqi Central Bureau of Statistics for the period from 1996 to 2018 By using the R program and according to equation (1.3), the general maximum entropy method is applied on the slope and Renyi measure with high order ($a = 6, a = 5, a = 4, a = 3$) and the Kink value of the explanatory variable is equal to (5), which is the point that divides the slope the limit (β) to two values (β^+_1, β^-_1) , the first before Kink and the second after Kink, as shown in the following figure:



After knowing the Kink point, the general maximum entropy methods of the higher order of Tsallis and the general maximum entropy of Renyi are applied in order to estimate the parameters of the Kink regression model, as well as extracting an absolute mean error standard for comparison between the methods used to determine the best of them. The results were obtained:

GME	$\hat{\beta}_0$	$\hat{\beta}_1^-$	$\hat{\beta}_1^+$	***Kink point***	***MAE***
Tsallis $\alpha = 3$	***0.976850***	***-0.131201***	***0.002497***	***-0.682631***	***2.366713***
Tsallis $\alpha = 4$	***0.976850***	***-0.131201***	***0.002497***	***-0.682631***	***2.366713***
Tsallis $\alpha = 5$	***0.987150***	***-0.091913***	***0.002496***	***1.992749***	***2.402055***
Tsallis $\alpha = 6$	***0.985978***	***-0.091912***	***0.002496***	***2.327672***	***2.407291***
Renyi $\alpha = 3$	***0.997019***	***-0.091916***	***0.002494***	***4.17462***	***2.390919***
Renyi $\alpha = 4$	***0.997035***	***-0.091914***	***0.002497***	***4.16988***	***2.390917***
Renyi $\alpha = 5$	***0.997070***	***-0.091914***	***0.002497***	***4.16010***	***2.390925***
Renyi $\alpha = 6$	***0.997101***	***-0.091914***	***0.002497***	***4.15127***	***2.390932***

From the results of the previous table, it is noticed that all (β_1^-) estimates have negative effects before the Kink point and then turn into positive effects after Kink. For all the methods used in the estimation, it is noticed that the estimate (β_1^+) was more positive in the Renyi measure when the value is Rank ($\alpha = 3$) after the Kink point and that the value of Kink is equal to (5), and in order to find the most efficient estimation method for the parameters of the Kink regression model among the methods used in this research, we use criterion the mean error absolute MAE , which showed that the Tsallis method is at the order. ($\alpha = 4$) is preferable to other methods.
Depending on the estimates of Tsallis of order ($\alpha = 4$), the Debt/GDP ratio was affected by inflation significantly before the value of Kink and then began to decline after the value of Kink.

5. Conclusion

The general maximum entropy method was applied in this research on real phenomena data representing inflation and the Debt/GDP ratio that suffers from the existence of Kink. The purpose of that is to estimate the parameters of the Kink regression model that represents this phenomenon and then choose the best estimation method to estimate its parameters and then Comparison of the methods used for estimation, based on MAE standard, to choose which method is the best estimation method, taking into account the idea of a higher order effect of the Tsallis measure and the Renyi measure.

References

[1] E. Ciavolino, *Modelling GME and PLS estimation methods for evaluating the job satisfaction in the public sector*, Linköping Elect. Conf. Proc. 149(14) (2008) 65–75.

[2] E. Ciavolino and A. Al-Nasser, *Information theoretic estimation improvement to the nonlinear gompertz's model based on ranked set sampling*, J. Appl. Quant. Meth. 5(2)(2010) 317–330.

[3] P. Ganong, & S. Jäger, *A permutation test and estimation alternatives for the regression kink design*, IZA Discussion Paper, 8282 (2014) 1–33.

[4] A. Golan and J. Perloff, *Comparison of maximum entropy and higher-order entropy estimators*, J. Econ. 107(1-2) (2002) 195–211.

[5] B. Hansen, *Regression kink with an unknown threshold*, J. Bus. Econ. Stat. 35(2) (2017) 228–240.

[6] S. Kamar and B. Msallam, *Comparative study between generalized maximum entropy and Bayes methods to estimate the four parameter Weibull growth model*, J. Prob. Stat. (2020).

[7] R. Sarma and T. Redd, *A Nonparametric Plugin Entropy Estimator based Renyi Entropy in Construction of Decision Trees*, Sri Krishnadevaraya University Anantapuramu, 2016.

[8] P. Tarkhamtham, W. Yamaka and S. Sriboonchitta, *The generalize maximum Tsallis entropy estimator in kink regression model*, J. Phys. Conf. Series, 1053(1) (2018) 012103.

[9] P. Tarkhamtham and W. Yamaka, *High-order generalized maximum entropy estimator in Kink regression model*, Thai J. Math. Special Issue, (2019) 185–200.

[10] P. Tibprasorn, P. Maneejuk and S. Sriboochitta, *Generalized information theoretical approach to panel regression kink model*, Thai J. Math. Special Issue, (2017) 133–145.