

The General Maximum Entropy Method by Shannon and Tsallis Measures for Estimating Parameters of a Kink Regression Model

Saad Obaid Jameel^a , Basim Shlaibah Msallam^{bs}

^a Baghdad University, College of Administration and Economics, Iraq (saad.iraq100@gmail.com)

^b Al-Iraqia University, College of Administration and Economics, Iraq (drbasimstatistics@gmail.com)

Corresponding Author: Basim Shlaibah Msallam* email:drbasimstatistics@gmail.com
Al-Iraqia University, College of Administration and Economics, Iraq

Article History:Received:11 January 2021; Accepted: 27 February 2021; Published online: 5 April

Abstract: In this paper, the parameters of the Kink regression model were estimated using the general maximum entropy method with the Shannon and Tsallis measures And its application for the analysis of economic exposure data and the Debt/GDP ratio in the Iraqi economy. The results of the analysis showed the preference of the Tsallis measure of the order ($\alpha = 3$) compared with other estimation methods. The results show a decrease in the Debt/GDP ratio after seeing the cut, and this corresponds to the state of the economy in Iraq during the security conditions that Iraq went through for the period from 2014 to 2017.

Keywords: Kink regression, GME, Shannon, Tsallis, Entropy.

Introduction

The Kink Regression is an important technique in statistical analysis because it deals with real phenomena in which the regression function is continuous with its explanatory variables related to the slope of the boundary and the Fixed limit parameter, but where the marginal trends are not continuous when observations of the relevant variable are observed, and that observation becomes is kink (Kink or Cut), And that the slope becomes has two values before and after the cut due to the lack of equality of the limit on the right side with the limited on the left side, and that the discontinuity of the marginal slope does not allow the use of normal regression to analyze the data of any phenomenon that has a cutoff point due to the inaccuracy of the results obtained, and it is useful to employ the Kink pieces When performing statistical analysis of this phenomenon as a support force to the regression model clearly and accurately. The process of employing Kink to obtain efficient results starts from including the regression model that cut, so the Kink regression model appeared, which is an extended method for the Regression Discontinuity model that appeared at the beginning of the problem, whether the regression was a parametric or not, despite its importance in analyzing most of the economic, health and social phenomena ... Etc. However, most of the research on this topic is few and lacks practical application and most of them use traditional estimation methods, the most famous of which is the least squares method LS [see 7], while the two researchers (Tarkhamtham and Yamaka 2018) [9] [10] presented the general maximum entropy method based on the Shannon 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.

In this research, the parameters of the kink regression model will be estimated and applied to real data representing the phenomenon of economic exposure and the Debt/GDP ratio in the Iraqi economy by using the general maximum entropy method with the Shannon and Tsallis measures and different order.

The second topic of the research included the nature of the kink regression model and some of its characteristics as well as the general maximum entropy method by adopting the Shannon and Tsallis measures, while the third section included the results of analyzing real data on the phenomenon for economic exposure and its impact on the ratio of Debt/GDP in the Iraqi economy.

The Model

The kink regression model writes the existence of kink points whose number is K with the number of explanatory variables simultaneously as follows [10]:

$$Y_i = \beta_1^-(x_{1i} - r_1) + \beta_1^+(x_{1i} - r_1) + \dots + \beta_k^-(x_{ki} - r_k) + \beta_k^+(x_{ki} - r_k) + \beta_0 x_{0i} + \varepsilon_i \quad (1)$$

$$i = 1, 2, \dots, n \quad , j = 1, 2, \dots, k \quad \hat{\beta} = [\beta_1^- \ \beta_1^+ \ \beta_0]$$

Y_i : It is the view value (i) of the observation of the response variable. X_{ji} : It is the value of the observation (i) from the observations of the explanatory variable. X_{0i} : it is the value of the observation (i) from the observations of the explanatory variable associated with the fixed term, and its value is equal to one the same, (r_1, \dots, r_k) The parameters of the points (Kink Point) [6].

an details of Model (1) is that it has two types of relationships, the first of which is a linear relationship between the response variable and the explanatory variable associated with the constant parameter X_{0i} , Which is often neglected for ease in the estimation process and the regression function in it is continuous, and the second type is non-linear between the response variable and the explanatory variables that have points (r_1, \dots, r_k) Kink Points represent where the response variable is continuous with the variables X_{ki} , but the marginal slope are not continuous at the cutoff point that represents an observation from the observations of the explanatory variable, so the marginal slope associated with the explanatory variables that have a kink cutoff have two values as well (Hansen:2017), The first symbol is denoted by the symbol (-), meaning the value of the marginal slope at the values of the explanatory variable observations before Kink, and by the symbol (+) for the value of the marginal slope at the values of the observations of the explanatory variable after cutting [7] [4].

Consistent with the nature of the phenomenon data that will be analyzed in the application side, which includes only one explanatory variable, we assume that the model (1) has one explanatory variable and one kink (r) only, as follows:

$$Y_i = \beta_1^- (x_1 - r) + \beta_1^+ (x_1 - r) + \beta_0 x_0 + \varepsilon_i \quad \dots (2)$$

It is noted from Model (2) that it cannot be written in the general form as in the traditional regression when it is written for n of the equations and the reason is that it is not possible to determine which of the observations of the explanatory variable that has Kink is to the right and left of it. Therefore, (Hansen:2017) method was adopted in writing Model (2) in terms of matrices as follows [7]:

$$Y_i = \beta' X_{i(r)} + \varepsilon_i \quad i = 1 \dots n \ , \quad \dots (3)$$

whereas

$$\hat{\beta} = [\beta_1^- \ \beta_1^+ \ \beta_0]$$

$$\hat{X}_{i(r)} = [(X_{1i} - r)^- \ (X_{1i} - r)^+ \ X_{0i}]$$

The researcher (Hansen: 2017) stated that testing the continuity hypothesis of the Kink regression model or not is one of the main steps that must be taken before estimating the parameters of the model (3), and because this step is difficult to implement, it was assumed that there is no continuity of the boundary slope function in the regression model according to the variables. Explanation associated with it, without going through testing this hypothesis and this approach will be adopted in this research also when estimating the parameters of the model (3) and determining the value of the Kink cutoff point.

The Generalized Maximum Entropy Method

The general maximum entropy method (GME) for estimating the parameters of the regression model in both linear and nonlinear types is based on directly reparametrized the regression model by reparametrized the parameters and errors in it as well as the Kink parameter in the form of predictions for random variables or in the form of a convex combination in terms of support points (S) whose limits are $(2 \leq S \leq 7)$ and their number is determined according to the researcher's point of view or the nature of the data [7] [3] For short, we will deal here with the case of $(S = 3)$ in the theoretical side and the practical side as follows [5]:

$$z_j = \left[-\frac{s_1}{2} \ , 0 \ , \frac{s_1}{2} \right] \ , \ q = \left[-\frac{s_2}{2} \ , 0 \ , \frac{s_2}{2} \right] \ v_i = \left[-\frac{s_3}{2} \ , 0 \ , \frac{s_3}{2} \right] \quad \dots \dots (4)$$

Where the values of (s_1, s_2, s_3) are a positive hypothetical numerical value between $(0, \infty)$, and z_j is directed from the order of $(1xs)$ representing the support points for each parameter of the model β as well as the fixed term,

\mathbf{q} is directed from the order (1xs)) Represents the support points of the parameter cutoff parameter Kink, \mathbf{v}_i , a vector of order (1xs), represents the support points for each of the random error items.

Several measures can be used when estimating model parameters (3) by the general maximum entropy method (GME), including the Shannon measure, the Tsallis measure, and the Renyi measure [10]. Here, the Shannon and Tsallis measures were chosen when estimating the GME method as follows:

1- Shannon General Maximum Entropy

Shannon has created a mathematical model that equates the amount of information while removing the ambiguity (uncertainty) [9]. Therefore, the Shannon Entropy function is generally written as follows:

$$H(x) = - \sum_i p_i \log p_i \quad \dots (5)$$

Whereas: $p_i(X = x_i) = p_i$ and $0 \leq p_i \leq 1$, p_i , It is the probability mass function of the Shannon measure, $\sum_i p_i = 1$.

(Tarkhamtham and Yamaka) stated that he prefers to deal with the Kink regression model without the variable X_{0i} to facilitate the estimation in GME method, so the parameters $(\varepsilon_k, r_1, \beta_1^+, \beta_1^-)$ will be modeled only by writing them In the form of a convex combination for a number of support points $S = 3$ specified [5] [7] as follows:

$$\beta_1^- = \sum_m p_{1m}^- z_{1m}^-, \quad x_{1,t} \leq r_1 \quad \dots (6)$$

$$\beta_1^+ = \sum_m p_{1m}^+ z_{1m}^+, \quad x_{1,t} > r_1 \quad \dots (7)$$

Where p_{1m}^- represents a probability distribution of the S dimension of the region before the cutoff point ($\sum_{m=1}^3 p_{1m}^- = 1$, p_{1m}^+), represents a probability distribution of the S dimension of the region after the kink point ($\sum_{m=1}^3 p_{1m}^+ = 1$), ($m=1, \dots, (S=3)$).

As for the kink point, it can be written:

$$r = \sum_m h_m q_m, \quad \dots (8)$$

Whereas: $\sum_{m=1}^3 h_m = 1$

Also, we rewrite the error ε_i as a convex formula [12] [8]:

$$\varepsilon_i = \sum_m \omega_{im} v_{im}, \quad i = 1, 2, \dots, n \quad \dots (9)$$

Whereas: $\sum_{m=1}^S \omega_{im} = 1$

Using equations (5) (6) (7) (8) (9), the general maximum entropy model of the Shannon measure can be constructed as follows [9] [10] [11] :

$$\begin{aligned} H(p, h, \omega) &= \operatorname{argmax}\{H(p) + H(h) + H(\omega)\} \\ &\equiv - \sum_k \sum_m p_{km}^- \log p_{km}^- - \sum_k \sum_m p_{km}^+ \log p_{km}^+ - \sum_k \sum_m h_{km} \log h_{km} - \sum_k \sum_m \omega_{tm} \log \omega_{tm} \quad (10) \end{aligned}$$

Whereas formula (10) is subject to limitations (consistency and standardization) [4]:

$$Y_i = \sum_m p_{1m}^- z_{1m}^- \left(x_{1i} - \sum_m h_{1m} q_{1m} \right) + \sum_m p_{1m}^+ z_{1m}^+ \left(x_{1i} - \sum_m h_{1m} q_{1m} \right) + \sum_m \omega_{im} v_{im} \quad \dots (11)$$

And

$$\sum_m p_{1m}^- = 1, \quad \sum_m p_{1m}^+ = 1, \quad \sum_m h_m = 1, \quad \sum_m \omega_{im} = 1$$

We rewrite the Lagrange function as follows:

$$L = H(p, h, \omega) + \gamma_1(\theta) + \gamma_2(1 - \sum_m p_{1m}^-) + \gamma_3(1 - \sum_m p_{1m}^+) + \gamma_4(1 - \sum_m h_{1m}) + \gamma_5(1 - \sum_m \omega_{tm}) \quad \dots (12)$$

Substituting (10) and (11) into (12) we get:

$$L = - \sum_k \sum_m p_{km}^- \log p_{km}^- - \sum_k \sum_m p_{km}^+ \log p_{km}^+ - \sum_k \sum_m h_{km} \log h_{km} - \sum_k \sum_m \omega_{im} \log \omega_{im} + \gamma_1(Y_i - \sum_m p_{1m}^- z_{1m}^- (x_{1i} - \sum_m h_{1m} q_{1m})) - \sum_m p_{1m}^+ q_{1m}^+ (x_{1i} - \sum_m h_{1m} q_{1m}) - \sum_m \omega_{im} v_{im}) + \gamma_2 - \gamma_2 \sum_m p_{1m}^- + \gamma_3 - \gamma_3 \sum_m p_{1m}^+ + \gamma_4 - \gamma_4 \sum_m h_{1m}) + \gamma_5 - \gamma_5 \sum_m \omega_{im}) \quad \dots (13)$$

By partially deriving equation (12) and finding the solution, we can find \hat{p}_{1m}^- , \hat{p}_{1m}^+ , $\hat{\omega}_{tm}$, and \hat{h}_{tm} as follows

$$\hat{p}_{1m}^- = \frac{\exp[-z_{1m}^- \sum_i \hat{\gamma}_{1i} (x_{1i} - \sum_m h_{1m} q_{1m})]}{\sum \exp[-z_{1m}^- \sum_i \hat{\gamma}_{1i} (x_{1i} - \sum_m h_{1m} q_{1m})]} \quad \dots (14)$$

$$\hat{p}_{1m}^+ = \frac{\exp[-z_{1m}^+ \sum_i \hat{\gamma}_{1i} (x_{1i} - \sum_m h_{1m} q_{1m})]}{\sum \exp[-z_{1m}^+ \sum_i \hat{\gamma}_{1i} (x_{1i} - \sum_m h_{1m} q_{1m})]} \quad \dots (15)$$

The same method is calculated $\hat{\omega}_{tm}$ and \hat{h}_{1m} :

$$\hat{\omega}_{im} = \frac{\exp[-\hat{\gamma}_{1i} v_{1m}]}{\sum \exp[-\hat{\gamma}_{1i} v_{1m}]} \quad \dots (16)$$

$$\hat{h}_{1m} = \frac{\exp[-\sum_i \hat{\gamma}_{1i} p_{1m}^- z_{1m}^- (x_{1i} - \sum_m q_{1m}) + \sum_i \hat{\gamma}_{1i} p_{1m}^+ z_{1m}^+ (x_{1i} - \sum_m q_{1m})]}{\sum \exp[-\sum_i \hat{\gamma}_{1i} p_{1m}^- z_{1m}^- (x_{1i} - \sum_m q_{1m}) + \sum_i \hat{\gamma}_{1i} p_{1m}^+ z_{1m}^+ (x_{1i} - \sum_m q_{1m})]} \quad \dots (17)$$

2- Tsallis General Maximum Entropy

The entropy function is defined according to the Tsallis measure as follows [3] [9]:

$$H_\alpha^T(x) = g \frac{\sum_k p_k^\alpha - 1}{1 - \alpha} \quad \dots (18)$$

Since α is order 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 formula No. (26) [10]. It should be noted that when $\alpha \rightarrow 1$ approaches, the measure of Tsallis turns into a Shannon is defined by the formula (9), [12].

As is the case in Shannon Measure, the parameters $(\epsilon_i, r, \beta_1^+, \beta_1^-)$ are reparametrized by also writing them in the form of Convex Combination for a number of support points S and as it was written with equations (6), (7) and (8) and (9)

Accordingly, the entropy function of the Tsallis measure is as follows [11] [5]:

$$H(p, h, \omega) = \text{argmax}\{H(p) + H(h) + H(\omega)\} \equiv \frac{1}{1-\alpha} (\sum_k \sum_m p_k^{\alpha-} - 1) + \frac{1}{1-\alpha} (\sum_k \sum_m p_k^{\alpha+} - 1) + \frac{1}{1-\alpha} (\sum_k \sum_m h_{km}^\alpha - 1) + \frac{1}{1-\alpha} (\sum_k \sum_m \omega_{km}^\alpha - 1) \quad \dots (19)$$

In the same manner as was used in the Shannon scale, we find \hat{p}_{1m}^- , \hat{p}_{1m}^+ , $\hat{\omega}_{tm}$ and \hat{h}_{tm} .

$$\hat{p}_{1m}^- = \left(\frac{1-\alpha}{\alpha}\right) \left[\sum_m \hat{y}_{1m} z_{1m}^- \left(\hat{x}_{1i} - \sum_m h_{1m} q_{1m} \right) + \gamma_{2k} \right] \quad \dots (20)$$

$$\hat{p}_{1m}^+ = \left(\frac{1-\alpha}{\alpha}\right) \left[\sum_m \hat{y}_{1m} z_{1m}^+ \left(\hat{x}_{1i} - \sum_m h_{1m} q_{1m} \right) + \gamma_{3k} \right] \quad \dots (21)$$

$$\hat{\omega}_{im} = \left(\frac{1-\alpha}{\alpha}\right) \left[\sum_m \hat{y}_{1m} v_{1m} + \gamma_{5k} \right] \quad \dots (22)$$

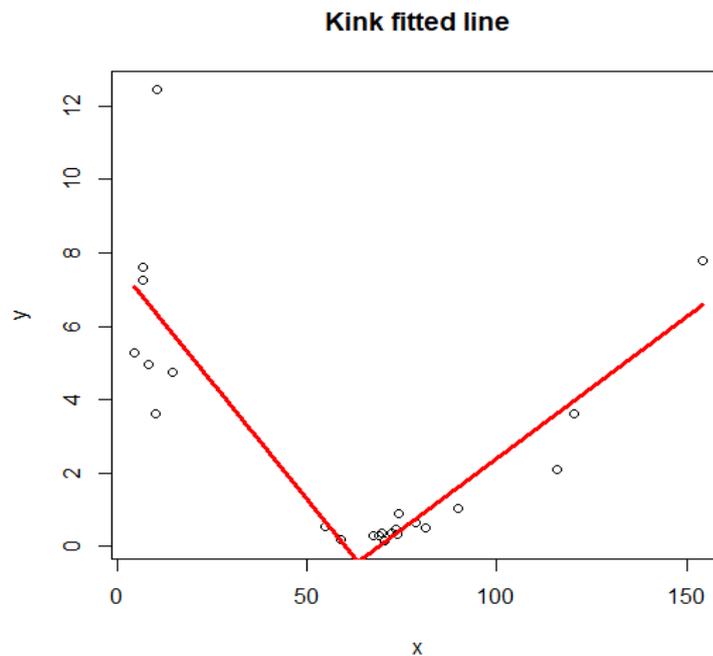
$$\hat{h}_{1m} = \left(\frac{1-\alpha}{\alpha}\right) \left[\sum_m \gamma_{1m} p_{1m}^- z_{1m}^- \left(\hat{x}_{1i} - \sum_m q_{1m} \right) - \sum_m \gamma_{1m} p_{1m}^- z_{1m}^- \left(\hat{x}_{1i} - \sum_m q_{1m} \right) - \gamma_{4k} \right] \quad \dots (23)$$

Where γ represents the Lagrang coefficient.

Case Study

In this topic, real data will be analyzed that represent economic exposure (an explanatory variable) and the ratio of DEBT/GDP (response variable) for the Iraqi economy for the period (1996-2018) and data obtained from (Ministry of Planning / Iraqi Central Bureau of Statistics) and according to the model (3) By using the R program, the GME method was applied to the general maximum entropy of the two scales of Shannon and Tsallis with high order ($\alpha = 5, \alpha = 4, \alpha = 3, \alpha = 2$). Figure (1) shows the value of Kink, which is equal to (69.6), which is the standard point that makes the marginal slope (β) divided into two values (β^-, β^+):

Figure (1) shows the value of Kink



By applying SGME and TSGME estimation methods to estimate the kink regression model, the results were obtained:

GME	$\hat{\beta}_0$	$\hat{\beta}_1^-$	$\hat{\beta}_1^+$	Kink point	MAE
-----	-----------------	-------------------	-------------------	------------	-----

Shannon	1.4063494	-0.0073778	-0.006045599	69.20053	2.390737
Tsallis $\alpha = 2$	1.395894	-2.34E-12	-0.006044578	69.62567	2.390408
Tsallis $\alpha = 3$	1.393115	2.66E-15	-0.006044465	69.46187	2.390245
Tsallis $\alpha = 4$	1.394039	-1.95E-13	-0.006044591	69.62206	2.390326
Tsallis $\alpha = 5$	1.396655	6.72E-13	-0.006044439	69.63927	2.390446

We note from the results of the above table that the estimators (β^- , β^+) have negative effects, and it is noticed that the estimate (β^-) has a negative effect except for the Tsallis estimator when the value of the order ($\alpha = 3$) is before the kink point (69.6) and this continues Negative effect in any region after kink and for all methods Shannon and Tsallis for different orders (α), Therefore, we find that the most efficient method for estimating the parameters of the Kink regression model is the Tsallis method when the value of the order is ($\alpha = 3$) using the MAE standard, which showed that it is the lowest of the rest of the methods. A positive effect before kink, then transformed into a negative effect after kink, which gives a comfortable impression that it is the most efficient estimate compared to other methods, and its MAE value has confirmed this.

Depending on Tsallis estimates of the rank ($\alpha = 3$), the Debt/GDP ratio was affected by economic exposure before the value of Kink, and this increased during the year (2015), and this confirms the conditions that Iraq was exposed to during the war with terrorist groups. The decrease is confirmed by the negative sign (-) for the parameter estimator (β^+), since the Iraqi economy activity began to rise gradually.

Conclusion

In this research, the general maximum entropy method was applied to real data representing economic exposure and the Debt/GDP ratio that suffer from the presence of kink. The purpose of that was to estimate the parameters of the Kink regression model representing that phenomenon and to determine the optimal representation method to estimate its parameters. Absolute error MAE to determine the best estimation method, taking into account the notion that there is a higher order effect of the Tsallis in the presence of the Shannon method.

Acknowledgements: To Professor dr Woraphon Yamaka, Centre of Excellence in Econometrics, Faculty of Economics Chiang Mai University, Thailand and To Professor dr Ahmed Al-waeli University of wasit, College of Administration and Economics, Iraq

The References

1. Böckerman, P., Kanninen, O., & Suoniemi, I. (2018). A kink that makes you sick: The effect of sick pay on absence. *Journal of Applied Econometrics*, 33(4), 568-579.
2. Ciavolino, E. (2008). Modelling GME and PLS estimation methods for evaluating the job satisfaction in the public sector.
3. Ciavolino, E., & Al-Nasser, A. D. (2010). Information theoretic estimation improvement to the nonlinear gompertz's model based on ranked set sampling. *Journal of Applied Quantitative Methods*, 5(2).
4. Ganong, P., & Jäger, S. (2014). A permutation test and estimation alternatives for the regression kink design.
5. Golan, A., & Perloff, J. M. (2002). Comparison of maximum entropy and higher-order entropy estimators. *Journal of Econometrics*, 107(1-2), 195-211.
6. Hansen, B. E. (2017). Regression kink with an unknown threshold. *Journal of Business & Economic Statistics*, 35(2), 228-240.
7. Kamar, S. H., & Msallam, B. S. (2020). Comparative Study between Generalized Maximum Entropy and Bayes Methods to Estimate the Four Parameter Weibull Growth Model. *Journal of Probability and Statistics*, 2020.
8. Sarma. R. Divakara & Reddy T. Bhaskara,(2016), A Nonparametric Plugin Entropy Estimator based Renyi Entropy in Construction of Decision Trees, Sri Krishnadevaraya University Anantapuramu.

9. Tarkhamtham, P., Yamaka, W., & Sriboonchitta, S. (2018, July). The generalize maximum Tsallis entropy estimator in kink regression model. In *Journal of Physics: Conference Series* (Vol. 1053, No. 1, p. 012103). IOP Publishing.
10. Tarkhamtham, P., Yamaka, W., Yamaka, W., & Yamaka, W. (2019). High-Order Generalized Maximum Entropy Estimator in Kink Regression Model. *Thai Journal of Mathematics*, 185-200.
11. Tibprasorn, P., Maneejuk, P., & Sriboochitta, S. (2017). Generalized information theoretical approach to panel regression kink model. *Thai Journal of Mathematics*, 133-145.
12. Xu, D., & Erdogmun, D. (2010). Renyi's entropy, divergence and their nonparametric estimators. In *Information Theoretic Learning* (pp. 47-102). Springer, New York, NY.