

تقديرات المربعات الصغرى والمربعات الصغرى المشدبة للانحدار الخطي بأخطاء طبيعية ملتوية

** حسين كريم ناشور *

**م.د. ابتسام كريم عبد الله*

Abstract

المخلص

التوزيع الطبيعي الملتوي بمعلمة الالتواء ϵ بسلون (μ, σ, ϵ) ESN يمثل فئة معلمية من التوزيعات الاحتمالية التي توفر انموذجاً أكثر مرونة نظراً لما توفره هذه المعلمة من امكانية في التغير المستمر من الحالة الطبيعية الى الحالة الغير طبيعية. من الناحية النظرية ان عملية تقدير معلمات انموذج انحدار خطي بمتوسط خطأ ذو قيمة وليس صفرياً تعتبر تحدياً كبيراً لما تحويه هذه العملية من صعوبات حيث لايمكن الحصول على صيغة نهائية لحساب هذه التقديرات وبالاخص معلمة القياس σ والالتواء ϵ , اما من الناحية التطبيقية فانه لايمكن الحصول على قيم لهذه التقديرات الا من خلال اللجوء للطرائق العددية. تم في هذا البحث تقدير معلمات انموذج انحدار خطي عام طبيعي ملتوي (SNGLM) بطريقتي المربعات الصغرى (LS) والمربعات الصغرى المشدبة (LTS) واستعمال متوسط مربعات الخطأ (MSE) كمعيار للمقارنة بين افضلية هاتين الطريقتين وذلك من خلال استخدام دراسة المحاكاة وذلك بتوليد بيانات ذات توزيع طبيعي ملتوي الى اليمين واليسار وبحجوم عينات مختلفة. حيث تبين ان طريقة LS تعطي قيم تقديرية بالنسبة لمعلمات SNGLM أكثر دقة من طريقة LTS .

المصطلحات الرئيسية: التوزيع الطبيعي الملتوي بمعلمة الالتواء ϵ بسلون (ESN), انموذج خطي عام طبيعي ملتوي (SNGLM), تقديرات المربعات الصغرى.

Estimates of least squares and least squares trimmed for linear regression with natural twisted errors

Abstract

The twisted normal distribution of the ESN torsion parameter (ϵ, μ, σ) represents a parameter class of probability distributions that provide a more flexible model given the potential of this parameter to continuously change from normal to abnormal. In theory, the process of estimating the parameters of a linear regression model with an average error of value, not zero, is a major challenge because of the difficulties involved in this process. These estimates are evaluated only through numerical methods. In this paper, the parameters of the SNGLM in LS and LTS and MSE as a criterion for comparing the preference of these two methods were estimated by using the simulation study by generating relevant data. A normal twist to the right and left, with different sample sizes, shows that the LS method gives more accurate values for SNGLM parameters than the LTS method.

Keywords: Twisted normal distribution with Epsilon torsion parameter (ESN), SNGLM, Linear squares model.

1. المقدمة

Introduction

إن مبدأ التماثل Symmetric هو المسار الذي يسلكه العديد من الباحثين لدراسة ظاهرة ما، والذي يتلخص في افتراض توزيع بيانات عينة ما، توزيعاً طبيعياً. إن افتراض الطبيعية لا يتحقق في أغلب الظواهر المدروسة وذلك بسبب وجود الالتواء الذي قد يكون سبباً رئيسياً في نموذج انحدار ضعيف، وك محاولة لمعالجة هذا الضعف تستخدم أساليب عديدة منها أسلوب أخذ اللوغارتم للبيانات وتحويلات Box Cox وغيرها، إلا أن هذه الأساليب تكون غير كافية في كثير من التطبيقات خاصةً البايولوجية منها والأسلوب البديل في هذه الحالة يتمثل بتوفير توزيعات تتسجم مع البيانات للحصول على نماذج أكثر مرونة لتتكيف مع كل نسبة التواء معينة.^[12]

دراسة مسألة الالتواء تجعل العلماء و الباحثين أكثر إدراكاً لأهمية التحليل الإحصائي الدقيق، ومما لا شك فيه أن أغلب الظواهر تحتوي على نسبة معينة من الالتواء وظهور ما يسمى بـ "عدم التماثل" asymmetric، ومن هذه التسمية برزت أهمية العائلة الطبيعية الملتوية (SN) Skew Normal Family والتي تتضمن العائلة الطبيعية باعتبارها حالة خاصة منها. من الناحية النظرية فإن العائلة (SN) لديها ميزة في كونها سهلة الانقياد رياضياً وامتلاكها عدد جيد من الخصائص المشتركة مع التوزيع الطبيعي، أما من الناحية التطبيقية فتعتبر العائلة (SN) هي الخيار الأفضل بالنسبة لجميع الظواهر التي يظهر فيها الالتواء^[10]. وبالتالي يمكننا القول بأن العائلة (SN) تهتم بتمثيل البيانات بأكبر قدر ممكن من التناسق (إعطاء بيانات الظاهرة المدروسة التوزيع المناسب) والتخلص من الافتراضات غير الواقعية، وهذا ما يوضح مقدار الحصانة التي تمتلكها هذه العائلة من خلال امتلاكها معلمة شكل إضافية تسمى "معلمة الالتواء" تعمل على تنظيم الالتواء بحيث تسمح هذه المعلمة بالتغير المستمر من الحالة الطبيعية إلى الحالة الملتوية^{[8],[10],[16]}.

إن طرائق تقدير الانحدار الخطي كثيرة ومتنوعة لعل من أبرزها والتي تعتبر الحجر الأساس في علم الإحصاء طريقة المربعات الصغرى الاعتيادية (O.L.S) Ordinary Least Squares، والتي تعمل وفق فروض يجب توفرها للخطأ العشوائي U_i ، من أهم هذه الفروض يجب أن يكون توزيع U_i توزيعاً طبيعياً بمتوسط 0 وتباين σ^2 . وبما أن طرائق التقدير التقليدية تكون حساسة تجاه القيم الشاذة (outliers) التي يعالجها الكثير من خلال حذفها أو تقديرها استناداً على مشاهدات أخرى، وبطبيعة الحال فإن هذا يقلل من مصداقية وصف الظاهرة المدروسة، أما في حالة الإبقاء على هذه القيم فعندها يجب استخدام طرائق حصينة للتقدير تمتاز بامتلاكها القدرة الكافية على تحمل وجود القيم الشاذة بالبيانات وعدم التأثير بها وبالتالي الحصول على نتائج أكثر كفاءة، دقة، واقعية، ومنطقية.

المشكلة الأساسية لهذا البحث تكمن في حال افتراض أن الخطأ العشوائي لا يتوزع توزيعاً طبيعياً، أي أن نموذج الانحدار الخطي يعاني من مشكلة الابتعاد عن التوزيع الطبيعي، Linear Regression Model Departure from Normality، بمعنى أدق أن الخطأ يتوزع توزيعاً طبيعياً ملتوي بمتوسط غير صفري وهذا ما يسبب صعوبة في حساب تقديرات معاملات النموذج وبالأخص معلمة الالتواء (ε) ومعلمة القياس (σ).

إن الغرض من هذا البحث هو تقدير معاملات نموذج انحدار خطي طبيعي ملتوي بطريقتي المربعات الصغرى (LS) والمربعات الصغرى المشدبة (LTS) والمقارنة بين افضلية هاتين الطريقتين باستخدام المحاكاة من خلال استعمال المعيار الإحصائي متوسط مربعات الخطأ (MSE).

لقد ظهرت التوزيعات الملتوية لأول مرة في عام 1897 من قبل Fechner^[9] الذي بين أنه يمكن الحصول على التوزيع الملتوي من خلال توصيل (ربط) اثنين من التوزيعات النصف الطبيعية Half Normal Distributions المختلفة من حيث معلمة القياس (σ). في عام 1976 قدم Leonard و O'Hagan^[13] لأول مرة التوزيع الطبيعي الملتوي (SN) Skew Normal Distribution في التحليل البيزي. وفي عام 1985 عمل Azzalini^[2] على التحقق بشكل دقيق من التوزيع الطبيعي الملتوي (SN) ودراسة خصائص دالة الكثافة الاحتمالية (pdf) له والتي كانت كالتالي:

$$f(x) = 2\phi(x)\Phi(\alpha x), \quad -\infty < x < \infty \quad (1.1)$$

حيث أن $\phi(x)$ تمثل pdf للتوزيع الطبيعي القياسي و $\Phi(x)$ تمثل الدالة التوزيعية التراكمية (cdf) للتوزيع الطبيعي القياسي، أما بالنسبة لـ α فهي تمثل معلمة الالتواء والتي تكون قيمتها بين $-\infty$ و ∞ . وفي عام 2000 قدم Epsilon Skew-Normal^[11] Mudholkar و Huston التوزيع الطبيعي الملتوي بمعلمة الالتواء أبسولون

ESN) Distribution. في عام 2005 قام Elsalloukh وآخرين^[7] بتقديم عائلة توزيع جديدة أطلقوا عليها اسم عائلة ايبسلون الملتوية القوة الاسية Epsilon-Skew Exponential Power (ESEP) والتي تتضمن التوزيع ESN كحالة خاصة منها, حيث بينوا ان هذه العائلة مفيدة في نمذجة البيانات الملتوية. في عام 2010 قام Reinaldo و آخرين^[14] بتوسيع معلمي للتوزيع ESN المقدم من قبل Mudholkar و Huston^[11] أطلقوا عليه العائلة الطبيعية الملتوية المرنة Flexible Epsilon-Skew-Normal Family (FESN) حيث ان هذه العائلة تحتوي على معلمة شكل اضافية (δ) تزيد من مرونة التوزيع ESN, هذا وان $FESN(\varepsilon, \delta)$ يختزل الى $ESN(\varepsilon)$ عندما $\delta = 0$. وفي عام 2013 استخدم Abdulah و Elsalloukh^[3] توزيع كما الملتوي بمعلمة الالتواء ايبسلون Epsilon Skew Gamma ($ES\Gamma$)^[4] على بيانات ذات توزيع قريب من توزيع كما حيث قاموا بتعريف التوزيع $ES\Gamma$, ذكر خصائصه ومميزاته, وتقدير معلماته باستعمال طريقتي الامكان الاعظم والعزوم, وتطبيق هذه التقديرات على بيانات تتوزع توزيع $ES\Gamma$. في عام 2018 قام Abdullah^[17] وزملائه باستخدام الخوارزمية الوراثية Genetic Algorithm (GA) لايجاد مقدرات دالة الامكان بالنسبة لمعلمات التوزيع SN ومقارنتها مع تقنيات تكرارية اخرى مثل (NR) Newton-Raphson, (NM) Nelder-Mead, والخوارزمية المعادة الوزن التكرارية Iteratively Re-Weighting Algorithm (IRA) حيث اثبتوا ان الخوارزمية (GA) هي الخوارزمية الاكثر كفاءة مقارنة مع باقي الخوارزميات من خلال حساب متوسط مربعات الخطأ Mean Squares Errors (MSE) لكل مقدر في حال تطبيقه في كل خوارزمية.

Skew Normal Distribution

2. التوزيع الطبيعي الملتوي

تحتل عائلات الموقع والقياس مكاناً مهيماً في النمذجة والتحليل الإحصائي, وهذا يعكس أهمية التوزيعات الملتوية في هذا المجال, فلقد كان الدافع وراء تقديم التوزيعات الملتوية هو الحصول على توزيعات تكون قادرة على التكيف بأكثر قدر ممكن مع البيانات^[1], وبصورة عامة فإن التوزيع الطبيعي الملتوي (SN) لديه الكثير من الإمكانات لكن غياب أسلوب كفوء لتقدير المعلمات هو بمثابة عائق رئيسي لاستخدامه على نطاق أوسع^[6], ويشير مصطلح طبيعي ملتوي (skew-normal) إلى فئة معلمية من التوزيعات الاحتمالية والتي تكون بمثابة توسعة للتوزيع الطبيعي من خلال معلمة شكل إضافية تنظم الالتواء, بحيث تسمح بالتغير المستمر من الحالة الطبيعية إلى الحالة الغير طبيعية^[6] ففي المجال التطبيقي فإن التوزيع SN يعتبر الحل الأمثل بالنسبة لجميع الحالات التي يوجد فيها بعض الالتواء وان الخاصية القيمة والتي تظهر بشكل واضح في هذه الحالة هي استمرارية التحول من الحالة الطبيعية إلى الملتوية, اما من الناحية النظرية فإن فئة التوزيعات الطبيعية الملتوية (SN) لديها ميزة في كونها سهلة الانقياد رياضياً وامتلاكها عدد جيد من الخصائص المشتركة مع التوزيع الطبيعي القياسي.

بين (Mudholkar و Huston)^[11] التوزيع (SN) بصيغة مختلفة عما جاء به Azzalini^[2], وأطلقوا عليها اسم التوزيع الطبيعي الملتوي بمعلمة الالتواء ايبسلون Epsilon Skew Normal (ESN) والذي تكون دالة الكثافة الاحتمالية له بالشكل التالي:

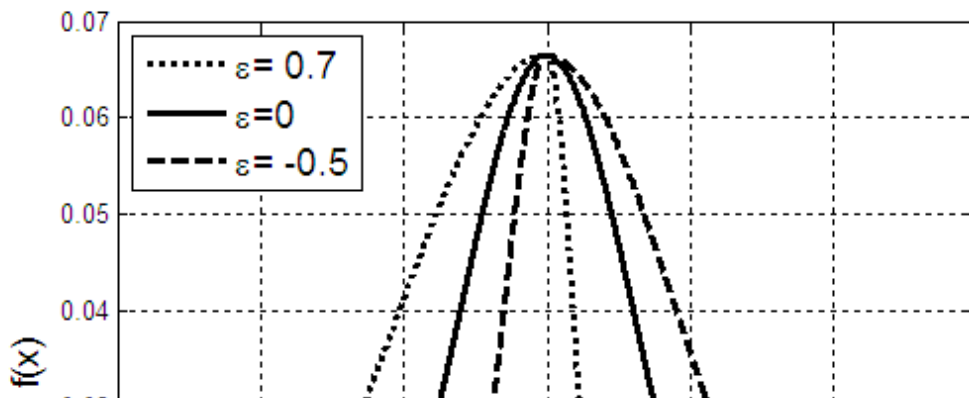
$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2(1+\varepsilon)^2}} & \text{if } x < \mu \\ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2(1-\varepsilon)^2}} & \text{if } x \geq \mu \end{cases}, \quad -1 < \varepsilon < 1 \quad (2.1)$$

حيث ان كل من μ , σ , و ε تمثل معلمة الموقع Location, القياس Scale, الالتواء Skewness على التوالي. وسنستخدم المعادلة (2.1) في هذا البحث والتي تمتلك متوسط وتباين يكونان كالتالي:

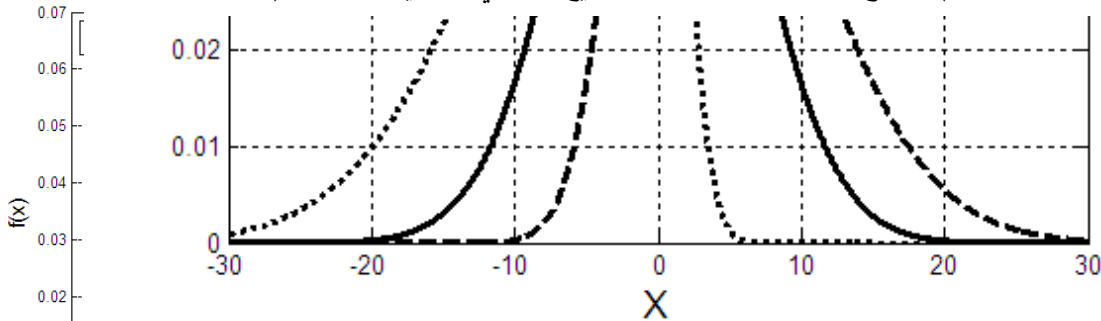
$$E(x) = \mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}} \quad (2.2)$$

$$V(x) = \frac{\sigma^2[(3\pi - 8)\varepsilon^2 + \pi]}{\pi} \quad (2.3)$$

الشكل (2.1) يوضح دالة الكثافة الاحتمالية للتوزيع (ESN) ذلك باختيار اكثر من قيمة لـ ε , حيث ان المنحني ذو الخط المنقط يوضح انه "إذا كانت ($\varepsilon > 0$) فهذا يعني لدينا التواء في الجانب الأيمن" [3] (Right Skewed) وفي هذه الحالة يكون الالتواء موجب اي إن البيانات تتركز معظمها في جهة اليمين ويكون المتوسط اكبر من الوسيط اكبر من المنوال (المتوسط يقع يمين المنوال). أما المنحني ذو الخط المتقطع فإنه يوضح "إذا كانت ($\varepsilon < 0$) فهذا يعني لدينا التواء في الجانب الأيسر" [3] (Left Skewed)، وفي هذه الحالة يعتبر الالتواء سالب اي إن البيانات تتركز معظمها في جهة اليسار ويكون المنوال اكبر من الوسيط اكبر من المتوسط (المتوسط يقع إلى يسار المنوال). ويكون التوزيع متماثل في حال كانت ($\varepsilon = 0$). الشكل (2.2) يوضح وبصورة ادق توزيع ESN ملتوي لليمين ولل اليسار بالنسبة لمعلمتي التواء موجبة وسالبة.



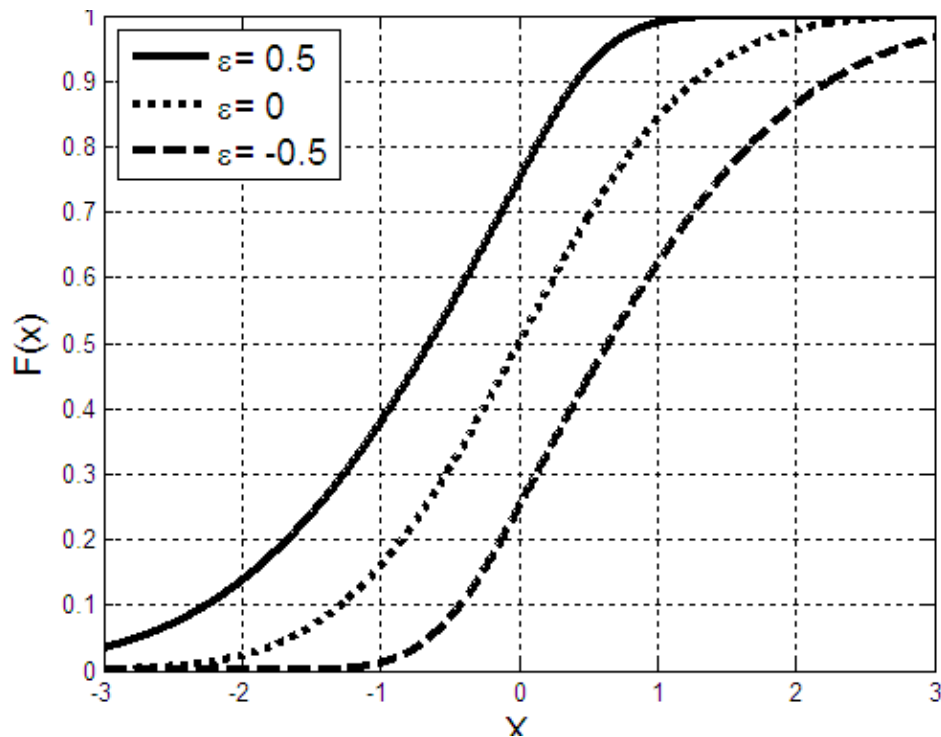
(يوضح دالة الكثافة الاحتمالية للتوزيع الطبيعي الملتوي 2.1 الشكل)



(يوضح الالتواء الموجب (الايمن) والسالب (الايسر) للتوزيع الطبيعي الملتوي 2.2 الشكل)

$$F(x) = \begin{cases} (1 + \varepsilon)\Phi\left(\frac{x}{1+\varepsilon}\right) & , x < 0 \\ \varepsilon + (1 - \varepsilon)\Phi\left(\frac{x}{1-\varepsilon}\right) & , x \geq 0 \end{cases} \quad (2.4)$$

والشكل (2.3) يوضح تأثير ε على شكل دالة الـ cdf للتوزيع ESN باختيار اكثر من قيمة لـ ε . حيث ان المنحني ذو الخط المنقط يوضح شكل دالة الـ cdf للتوزيع الطبيعي ($\varepsilon = 0$), المنحني ذو الخط المتصل يوضح دالة الـ cdf لتوزيع ESN ملتوي لليمين وهذا واضح من خلال تأثير قيمة ε المساوية لـ 0.5 في كون المنحني من الجانب الايمن اصبح ذو ذيل اطول من الجانب الاخر, والمنحني ذو الخط المتقطع يوضح دالة الـ cdf لتوزيع ESN ملتوي لليسار وهذا واضح من خلال تأثير قيمة ε المساوية لـ -0.5 في كون المنحني من الجانب الايسر اصبح ذو ذيل اطول من الجانب الاخر.



3. الانموذج الخطي العام الطبيعي-الملتوي Skew-Normal General Linear Model (SNGLM)

يمكن تمثيل العلاقة بين المتغير المعتمد (Y_i) وعدد من المتغيرات المستقلة (X_i) وحد الخطأ العشوائي (U_i) بالمعادلة الخطية الآتية:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_K X_{iK} + U_i \quad \begin{cases} i = 1, 2, \dots, n \\ j = 0, 1, 2, \dots, K \end{cases} \quad (3.1)$$

هذا ويمكن كتابة (3.1) بصورة مختصرة وباستخدام المصفوفات والمتجهات كالآتي:

$$Y = X\beta + U \quad (3.2)$$

حيث أن Y يمثل متجه لمشاهدات المتغير المعتمد ذو بعد $(n \times 1)$, X مصفوفة ذات أبعاد $(n \times (K+1))$ لمشاهدات المتغيرات المستقلة, β متجه $((K+1) \times 1)$ لمعاملات الانحدار علماً بأن العنصر الأول منه يمثل الحد الثابت, و U هو متجه $(n \times 1)$ للخطأ العشوائي والذي يتوزع توزيعاً طبيعياً ملتوي $U \sim \text{ESN}(\mu, \sigma, \varepsilon)$.

عندما يكون توزيع الخطأ توزيعاً طبيعياً ملتوي (ESN) فإن متوسطه يكون كما هو مذكور في المعادلة (2.2). وبما أن توقع الخطأ عبارة عن حد ثابت لذا لا بد من ضربه بمتجه الوحدة $I_{n \times 1}$ عند التقدير, وذلك لأننا سنتعامل مع متجه من الأخطاء العشوائية, لذلك فإن الصيغة التقديرية للانموذج (3.2) ستكون بالشكل التالي:

$$\hat{Y} = EY = X\beta + E(U)$$

$$EY = X\beta + \left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}} \right) I \quad (3.3)$$

Lest Square Estimates

4. تقديرات المربعات الصغرى

بما ان حد الخطأ بالنسبة للمتغير المعتمد (Y) هو عبارة عن مقدار الفرق بين القيمة الحقيقية و القيمة التقديرية, لذا فإن حد الخطأ بالنسبة للانموذج الخطي العام الطبيعي الملتوي (SNGLM) يكون كالآتي:

$$U = Y - EY \quad (4.1)$$

وبتعويض المعادلة (3.3) بالمعادلة (4.1) نحصل على

$$U = Y - X\beta - \left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right) I \quad (4.2)$$

لذلك فإن مجموع مربعات الاخطاء سيكون بالشكل الآتي:

$$\sum_{i=1}^n (U_i)^2 = U'U \quad (4.3)$$

فبعد تعويض المعادلة (4.2) بما يساويها في المعادلة (4.3) وبالتبسيط حيث ان

$(X\beta)'I = I'X\beta$, $Y'I = I'Y$ و $(X\beta)'Y = Y'(X\beta)$ سنحصل على الصيغة النهائية لمجموع مربعات الاخطاء وهي كالآتي:

$$\begin{aligned} U'U &= Y'Y - Y'(X\beta) - \left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right) Y'I - (X\beta)'Y + (X\beta)'(X\beta) + \left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right) (X\beta)'I \\ &\quad - \left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right) I'Y + \left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right) I'(X\beta) + \left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right)^2 I'I \\ U'U &= Y'Y - 2Y'(X\beta) - 2\left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right) Y'I + (X\beta)'(X\beta) + 2\left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right) (X\beta)'I \\ &\quad + n\left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right)^2 \end{aligned} \quad (4.4)$$

بأخذ التفاضل للمعادلة (4.4) بالنسبة لمتجه معلمات الانحدار (β), وبالمساواة للصفر نحصل على الصيغة التقديرية لـ β .

$$\begin{aligned} \frac{\partial U'U}{\partial \beta} &= -2Y'X + 2X'X\beta + 2\left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right) X'I \\ &\quad - X'Y + X'Xb + \left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right) X'I = 0 \\ b &= (X'X)^{-1} \left(X'Y - \left(\mu - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right) X'I \right) \end{aligned} \quad (4.5)$$

وبأخذ التفاضل للمعادلة (4.4) بالنسبة لـ μ , σ , و ε على التوالي ومساواتها للصفر سنحصل على صيغ تقديرية لكل من μ , σ , ε كما موضح ادناه :

$$\frac{\partial U'U}{\partial \mu} = -Y'I + (Xb)'I + n\left(\hat{\mu} - \frac{4\sigma\varepsilon}{\sqrt{2\pi}}\right) = 0$$

$$\hat{\mu} = \frac{Y'I - (Xb)'I + \left(\frac{4\sigma\epsilon}{\sqrt{2\pi}}\right)n}{n} \quad (4.6)$$

$$\frac{\partial U'U}{\partial \sigma} = \left(\frac{8\epsilon}{\sqrt{2\pi}}\right)Y'I - \left(\frac{8\epsilon}{\sqrt{2\pi}}\right)(Xb)'I - \left(\mu - \frac{4\hat{\sigma}\epsilon}{\sqrt{2\pi}}\right)\left(\frac{8\epsilon}{\sqrt{2\pi}}\right)n = 0$$

$$\hat{\sigma} = \frac{((Xb)'I - Y'I + n\mu)\sqrt{2\pi}}{4n\epsilon} \quad (4.7)$$

$$\frac{\partial U'U}{\partial \epsilon} = \left(\frac{8\hat{\sigma}}{\sqrt{2\pi}}\right)Y'I - \left(\frac{8\hat{\sigma}}{\sqrt{2\pi}}\right)(Xb)'I - \left(\mu - \frac{4\hat{\sigma}\hat{\epsilon}}{\sqrt{2\pi}}\right)\left(\frac{8\hat{\sigma}}{\sqrt{2\pi}}\right)n = 0$$

$$\hat{\epsilon} = \frac{((Xb)'I - Y'I + n\mu)\sqrt{2\pi}}{4n\hat{\sigma}} \quad (4.8)$$

بأعادة النظر الى ماذكر من صيغ تقديرية بالنسبة لمعاملات SNGLM ونخص بالذكر الصيغ التقديرية للمعاملات μ , σ , ϵ سنجد ان كل واحدة من هذه التقديرات عبارة عن صيغة مغلقة (Closed form) بمعنى انه لايمكن الحصول على صيغة نهائية لها وحلها الا عن طريق استخدام طرائق عددية Numerical Methods للتقدير للتقدير [17] وهذا ماسيتم العمل عليه لاحقاً.

Least Trimmed Squares

5. طريقة المربعات الصغرى المشذبة

تعتبر طريقة المربعات الصغرى المشذبة (LTS) التي تم تقديمها من قبل Rousseeuw عام 1984 [15] احد طرائق الانحدار الحصين المهمة والجديرة بالثقة بالنسبة لتحليل البيانات وذلك لاسلوبها الذي يعتمد على تجاهل القيم الشاذة. تعمل هذه الطريقة على تصغير مجموع مربعات الاخطاء بعد ترتيبها تصاعدياً ويتم حساب مقدر L.T.S من خلال الاتي:

$$D(r) = \sum_{i=1}^h (r^2)_{i:n} \quad (5.1)$$

حيث ان r_1^2 تمثل البواقي المربعة والمرتبطة تصاعدياً $(r^2)_{1:n} \leq (r^2)_{2:n} \leq \dots \leq (r^2)_{n:n}$. h يمثل عدد المشاهدات التي يتم اعتمادها بعد حذف القيم الشاذة، ويعرف بثابت القطع الذي يحدد درجة الحصانة المطلوبة، والذي تكون "قيمته مساوية لـ $[n/2] + [(p+1)/2]$ حيث ان p يمثل عدد المعاملات [5]."

طريقة LTS وكما موضح في المعادلة (5.1) تستبعد البواقي المربعة والكبيرة من المجموع، وهذا مايسمح باستبعاد النقاط الشاذة للبيانات بشكل كامل، فبالاعتماد على قيمة h ، وترتيب البيانات الشاذة يمكن ان تكون هذه الطريقة كفوءة للغاية [5]. من ناحية الية عمل LTS فهي تعمل بنفس الية عمل LS لكن الاختلاف الوحيد يتمثل بأن طريقة LTS قبل الشروع بعملية التقدير تعمل على حذف الشواذ من العينة. اما من ناحية تكيف هذه الطريقة بالنسبة لتوزيع خطأ ESN فهذا يتم من خلال ابدال البواقي الاعتيادية ببواقي الانحراف، ومن ثم تقدير معاملات النموذج الانحدار حسب طريقة LS (المذكورة مسبقاً) وبالاغتماد على نفس المقدرات.

Simulation

6. المحاكاة

لتوليد انموذج SNGLM بمتغيرين مستقلين كما موضح في (6.1)، واجراء مقارنة بين طريقتين كان لابد من عمل دراسة المحاكاة والتي تضمنت هنا توليد أربع حجومات لعينات (Y_i) تتوزع التوزيع $ESN(\mu, \sigma, \epsilon)$ وهي 20, 50, 100, 200 وكذلك توليد متغيرين مستقلين (X_1) و (X_2) ، حيث تم التوليد بالاعتماد على دالة معكوس الـ CDF (Inverse CDF) للتوزيع $ESN(\mu, \sigma, \epsilon)$ [11] وكما موضح في (6.2) - كما تم تحديد قيم افتراضية (اولية) Initial

values لكل من معلمات الانموذج المطلوب تقديرها $\beta_0, \beta_1, \beta_2, \mu, \sigma$ وهي -1, 3, 5, 7, 1 على التوالي بحيث تكون هذه القيم ثابتة لجميع أحجام العينات.

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U \quad (6.1)$$

$$F^{-1}(u) = \begin{cases} \mu + \sigma \left((1 + \varepsilon) \Phi^{-1} \left(\frac{u}{1 + \varepsilon} \right) \right) & , 0 < u < \frac{(1 + \varepsilon)}{2} \\ \mu + \sigma \left((1 - \varepsilon) \Phi^{-1} \left(\frac{u - \varepsilon}{1 - \varepsilon} \right) \right) & , \frac{(1 + \varepsilon)}{2} \leq u < 1 \end{cases} \quad (6.2)$$

بالنسبة لمعلمة الالتواء ε فلقد تم افتراض قيمتين لها وهي -0.5 و 0.7 حيث ان مقدرات SNGLM الناتجة من القيمة الاولى كما موضح في الجدول (1) تمثل مقدرات ناتجة من توزيع طبيعي ملتوي لليسا Left Epsilon Skew Normal (LESN), بينما مقدرات SNGLM الناتجة من القيمة الثانية كما موضح في الجدول (2) تمثل مقدرات ناتجة من توزيع طبيعي ملتوي لليمين Right Epsilon Skew Normal (RESN). كان لابد من اللجوء لطريقة عددية Numerical Method المتمثلة بخوارزمية Trust Region لايجاد قيم لمقدرات SNGLM.

من خلال ملاحظة الجداول (1) و(2) لوحظ ان طريقة LS افضل من طريقة LTS في حساب المقدرات وهذا واضح من خلال امتلاك طريقة LS قيمة MSE اقل من LTS ولجميع أحجام العينات. كما لوحظ ان القيم التقديرية بالنسبة لمعلمة القياس ($\hat{\sigma}$) والالتواء ($\hat{\varepsilon}$) بكلتا الطريقتين تقترب بنسبة كبيرة من القيم الحقيقية، وهذا واضح من خلال امتلاكها MSE يقترب بنسبة كبيرة من الصفر ولجميع أحجام العينات. بزيادة حجم العينة فإن القيم التقديرية لكل من معلمات الانحدار b_1 و b_2 وبكلتا الطريقتين تقترب من القيم الحقيقية، فعلى سبيل المثال ومن خلال ملاحظة قيم MSE للقيمة التقديرية لـ b_2 في الجدول (2) وجد انها تقل بنسبة 93% بطريقة LS عند زيادة حجم العينة من 20 الى 200.

عند زيادة حجم العينة بنسبة كبيرة ($N=200$) وللطريقتين فإن القيمة التقديرية بالنسبة لمعلمة الموقع ($\hat{\mu}$) ومعلمة الحد الثابت (b_0) تقترب من القيمة الحقيقية وكما موضح في الجدول (2)، فعلى سبيل المثال عند زيادة حجم العينة من 20 الى 200 فإن MSE لكل من $\hat{\mu}$ و b_0 بطريقة LS يقل بنسبة 97.8% و 91% على التوالي.

N	Method Parameter	LS	MSE_LS Parameter	LTS	MSE_LTS Parameter
20	b_0	-1.1794	0.0322	-1.7262	0.5274
	b_1	3.0569	0.0032	3.0234	0.0005
	b_2	4.9961	15.0624E-6	5.0135	0.0002
	$\hat{\mu}$	6.4446	0.3085	7.2099	0.0440
	$\hat{\sigma}$	0.9998	39.1188E-9	0.9998	42.3806E-9
	$\hat{\varepsilon}$	-0.4996	14.9713E-8	-0.4996	14.9215E-9
	MSE	29.3745E-17		0.1267	
50	b_0	-2.0612	1.1261	-0.7445	0.0653
	b_1	3.0074	0.0001	3.0142	0.0002
	b_2	5.0143	0.0002	5.0149	0.0002
	$\hat{\mu}$	7.5294	0.2803	6.3099	0.4762
	$\hat{\sigma}$	0.9996	18.6966E-9	1.0001	21.2000E-9
	$\hat{\varepsilon}$	-0.4999	21.2496E-9	-0.5004	19.1164E-8
	MSE	28.2716E-17		7.9269	
100	b_0	-0.0396	0.9224	0.9506	3.8049
	b_1	3.0003	73.1811E-9	2.9942	33.3896E-6
	b_2	4.9964	13.0618E-6	4.9981	3.6554E-6
	$\hat{\mu}$	6.2856	0.5104	5.6795	1.7437
	$\hat{\sigma}$	1.0000	16.791E-10	1.0002	38.4361E-9
	$\hat{\varepsilon}$	-0.5000	30.9288E-11	-0.5001	17.4485E-9
	MSE	81.3015E-17		4.0644	
200	b_0	-0.2877	0.5074	-1.7070	0.4998
	b_1	2.9988	13.788E-7	2.9984	24.289E-7
	b_2	5.0013	17.113E-7	5.0035	12.1232E-6
	$\hat{\mu}$	6.2930	0.4998	7.5123	0.2625

$\hat{\sigma}$	1.0001	79.280E-10	1.0024	57.183E-7
$\hat{\varepsilon}$	-0.5000	18.329E-10	-0.5012	14.715E-7
MSE	13.295E-19		0.0189	

جدول (1) تقديرات معلمات انموذج SNGLM و MSE عندما $\varepsilon = -0.5$

N	Method Parameter	LS	MSE_LS Parameter	LTS	MSE_LTS Parameter
20	b_0	0333-1.	0.0011	-0.9708	0.0009
	b_1	3.0742	0.0055	3.0326	0.0011
	b_2	5.0086	0.0001	5.0310	0.0010
	$\hat{\mu}$	6.0968	0.8159	6.3027	0.4862
	$\hat{\sigma}$	0.9997	80.4616E-9	0.9997	72.0856E-9
	$\hat{\varepsilon}$	0.6996	16.5109E-8	0.6996	14.8233E-8
	MSE	21.3257E-18		0.1763	
50	b_0	-1.7540	1.5686	-1.7209	0.5198
	b_1	3.0075	0.0001	3.0135	0.0002
	b_2	5.0124	0.0002	5.0108	0.0001
	$\hat{\mu}$	7.3654	0.1335	7.4755	0.2261
	$\hat{\sigma}$	1.0011	11.074E-7	0.9997	89.6127E-9
	$\hat{\varepsilon}$	0.7007	46.3102E-8	0.6996	15.1962E-8
	MSE	23.6049E-26		6.2454	
100	b_0	-1.6891	0.4749	-1.6907	0.4771
	b_1	3.0010	10.326E-7	2.9965	12.5844E-6
	b_2	4.9975	60.576E-7	4.9998	59.7998E-9
	$\hat{\mu}$	7.8681	0.7536	8.2135	1.4725
	$\hat{\sigma}$	0.9970	89.764E-7	1.0094	0.0001
	$\hat{\varepsilon}$	0.6979	45.003E-7	0.7065	42.2718E-6
	MSE	66.5081E-26		6.2126	
200	b_0	-0.9927	0.0001	-0.5704	0.1846
	b_1	2.9985	23.972E-7	2.9979	43.891E-7
	b_2	5.0027	71.340E-7	5.0055	29.7207E-6
	$\hat{\mu}$	6.8661	0.0179	6.1979	0.6434
	$\hat{\sigma}$	1.0000	96.7652E-14	1.0001	99.337E-10
	$\hat{\varepsilon}$	0.7000	10.6254E-15	0.6999	38.980E-10
	MSE	53.811E-25		0.0343	

جدول (2) تقديرات معلمات انموذج SNGLM و MSE عندما $\varepsilon = 0.7$

Conclusions and Recommendations

7. الاستنتاجات والتوصيات

في هذا البحث تم حساب مقدرات معلمات انموذج SNGLM باستخدام طريقتي LS و LTS, وبالاتماد على مبدأ تصغير مجموع مربعات الخطأ أقل ما يمكن بالنسبة لكل معلمة. كما تمت المقارنة بين الطريقتين باستخدام دراسة المحاكاة لتوليد بيانات ملتوية لليمين واليسار وبحجوم عينات مختلفة. من خلال برنامج الماتلاب وباستخدام خوارزمية تكرارية تمثلت بخوارزمية Trust-region-dogleg تم حساب المقدرات.

من اهم الاستنتاجات التي تم التوصل اليها هو ان طريقة LS تعطي تقديرات اكثر دقة من طريقة LTS وذلك لامتلاكها MSE أقل بالنسبة لجميع حجوم العينات. هذا وتقرب القيمة التقديرية من الحقيقية لكل من معلمة القياس (σ) واللتواء (ε) لكلا الطريقتين ولجميع حجوم العينات ايضاً. كما تمت الملاحظة على انه بزيادة حجم العينة فإن MSE يقل بالنسبة لمقدرات معلمات الانحدار b_1 و b_2 لكلا الطريقتين. بزيادة حجم العينة بنسبة كبيرة تزداد دقة مقدرات كل من معلمة الموقع ($\hat{\mu}$) ومعلمة الحد الثابت (b_0) وكذلك يقل MSE لها.

ان الافضلية التي تحققها طريقة LS المكيفة مع توزيع خطأ $ESN(\mu, \sigma, \varepsilon)$ على طريقة حصينة مثل طريقة LTS تظهر لنا التأثير الواضح بالنسبة لمعلمة اللتواء ε على سلوك هذه الطريقة وجعلها اكثر حصانة لذلك نوصي

باستخدامها. كما نوصي باستخدام التوزيع ESN وذلك نظراً لما يوفره هذا التوزيع من مرونة في حال توفر بيانات ملتوية Skewed data, وإضافة طرائق تقدير أخرى للمقارنة كطريقة الامكان الاعظم.

References

8. المصادر

- [1] Azzalini, A., & Capitanio, A. (2003) "Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t-Distribution" Journal of the Royal Statistical Society, Series B (Statistical Methodology), vol. 65, no. 2, pp. 367–389.
- [2] Azzalini, A. (1985) " A Class of Distributions Which Includes the Normal Ones" Scandinavian Journal of Statistics, vol.12, no.2, pp.171-178.
- [3] Abdulah, Ebtisam & Elsalloukh, Hassan (2013) "Analyzing Skewed Data with the Epsilon skew Gamma Distribution" Journal of Statistics Applications & Probability, vol.2, no.3, pp.195-202.
- [4] Abdulah, Ebtisam & Elsalloukh, Hassan (2012) "The Epsilon Skew Gamma Distribution" JSM Proceeding, Biometrics Section. San Diego California: American Statistical Association.
- [5] Alma, Ö.G. (2011) "Comparison of Robust Regression Methods in Linear Regression" Int. J. Contemp. Math. Sciences, vol.6, no.9, pp.409-421.
- [6] Cantoni & E. Ronchetti (2001) "Robust inference for generalized linear models" Journal of the American Statistical Association, 96, 1022-1930.
- [7] Elsalloukh, Hassan & H. Guardiola, Jose & Young, Dean. (2005) "The epsilon-skew exponential power distribution family" Far East Journal of Theoretical Statistics, vol.17, no.1, pp.97-112.
- [8] Figueiredo, F. & Gomes, M.I. (2013) "The skew-normal distribution in SPC" REVSTAT- Statistical Journal, vol.11, no.1, pp.83–104.
- [9] Fechner, G.T., 1897. Kollektivmasslehre. Leipzig, Engleman.
- [10] Genton, M.G (2004) "Skew-Elliptical distribution and their applications: A journey Beyond normality" Boca Raton: Chapman & Hall/CRC.
- [11] Mudholkar, G.S. & Huston , A.D. (2000) "The epsilon-skew-normal distribution for analyzing near-normal data" Journal of statistical planning and Inference, vol.83, no.2, pp. 291-309.
- [12] Mashtare, T., & Hutson, A. (2011) "Utilizing the Flexibility of the Epsilon-Skew-Normal Distribution for Tobit Regression Problems" Communications in statistics-Theory and methods, vol.40, no.3, pp.408-423.
- [13] O'HAGAN, A & LEONARD, TOM (1976) "Bayes estimation subject to uncertainty about parameter constraints" Biometrika, vol.63, no.1, pp.201-203.
- [14] Reinaldo B. Arellano-Valle, Milton A. Cortes,& Hector W. Gomez (2010) "An Extension of the Epsilon-Skew-Normal Distribution" Communications in statistics-theory and methods, vol.39, no.5, pp.912-922.

- [15] Rousseeuw, Peter j & Leroy, Annick M. (1987) "Robust regression and outlier detection" Wiley series in probability and mathematical statistics.
- [16] Sujit , K. Sahu & High, S. Chai (2009) "A new Skew-Normal distributiion and its Properties" Calcutta Statistical Association Bulletin, vol.61, no.1-4, pp.197-226.
- [17] Yalçinkaya , Abdullah. et al (2018) "Maximum likelihood estimation for the parameters of skew normal distribution using genetic algorithm" Swarm and Evolutionary Computation, vol.38, pp.127-138.

