استعمال طريقة التحصين (S-Estimate) مع شرائح (B-Spline) لتقدير أنموذج الانحدار اللامعلمي (دراسة تجريبية)

الباحثة / غيداء ابراهيم شهاب / جامعة بغداد / كلية الادارة والاقتصاد
Gheadaa.ahmed1201@coadec.uobaghdad.edu.iq
م.د. نازك جعفر صادق/جامعة بغداد / كلية الادارة والاقتصاد dr.nazik@coadec.uobaghdad.edu.iq

P: ISSN : 1813-6729
E : ISSN : 2707-1359
http://doi.org/10.31272/JAE.45.2022.132.20

تأريخ أستلام البحث : 2021/11/2
مقبول للنشر بتأريخ :2021/12/5

المستخلص

ان واحدة من اكثر طرائق التمهيد شيوعا هي شرائح **B-splines** أصبح استخدام هذه الشرائح شائعًا جدًا بين العديد من مجالات الرياضيات والهندسة ، وعلوم الكمبيوتر في السنوات الأخيرة. في الأصل تم استخدام شرائح **B** لأغراض التقريب لكن شعبيتها وسعت تطبيقاتها, بالإضافة الى ذلك ان هذه التقنية توضيحية ومرنة للغاية على عكس تقنية الشريحة العادية او طرائق التنعيم الاخرى.

في هذا البحث سيتم استعمال بعض اساليب التمهيد منها، ممهد انحدار الشريحة والشريحة الجزائية باستعمال شريحة B-spline،وذلك من اجل الحصول على تقدير أنموذج الانحدار اللامعلمي . ويهدف البحث الى أيجاد أفضل مقدر من بين مقدرات التمهيد التي ذكرت . لتمثيل البيانات المدروسة بناء على نتائج تجارب المحاكاة ،ومن خلال الجانب التجريبي تم التوصل الى إن اسلوب تمهيد الشريحة الجزائية باستعمال شريحة B-spline كان الافضل في تقدير أنموذج الانحدار اللامعلمي .

الكلمات المفتاحية : نماذج الانحدار اللامعلمي ، طريقة التقدير Regression B- Spline ، طريقة التقدير Penalize B-Spline، **شرائح** B-Spline،طريقة التحصين S-estimate.

مجلة الادارة والاقتصاد
العدد 132 /اذار/ 2022
الصفحات : 279 - 288

البحث مستل من رسالة ماجستير

1- المقدمة

تظهر اهمية الاحصاء في محاولة دراسة عدة ظواهر مختلفة بواسطة نماذج اقرب الى الواقع. قد تكون هذه النماذج سببية ويتم بناؤها على اساس السبب والنتيجة وهذا ما يسمى بنموذج الانحدار ويكون له شكل دالة وتستند الى افتراضات محددة ولكن في بعض الاحيان عند غياب المعرفة بالظواهر المدروسة وعدم امكانية تحديد الدالة السببية بين المتغيرات ، ينتج نوعاً اخر من نماذج الانحدار يسمى بالانحدار اللامعلمي . وهذا النوع من نماذج الانحدار لا تأخذ قيمة المتغير المستقل شكلا محددًا ولكنها مبنية من المعلومات المأخوذة من البيانات وهذا يتطلب عينة حجمها كبير اكثر من المعتاد. ولكي يتم تقدير دالة الانحدار اللامعلمي فلا بد من استخدام اسلوب التمهيد وان هذا الاسلوب ينص على تقريب دالة الانحدار التقريبية الى دالة الانحدار الحقيقية ، وان واحدة من اكثر طرائق التمهيد شيوعا هي شرائح B-splines أصبح استخدام هذه الشرائح شائعًا جدًا بين العديد من مجالات الرياضيات والهندسة ، وعلوم الكمبيوتر في السنوات الأخيرة. في الأصل تم استخدام شرائح **B** لأغراض التقريب لكن شعبيتها وسعت تطبيقاتها ، بالإضافة الى ذلك ان هذه التقنية توضيحية ومرنة للغاية على عكس تقنية الشريحة العادية او طرائق التنعيم الاخرى ، فهي ليست ضرورية لحساب مشتقات او وضع افتراضات خاصة لحل المعادلات بسبب هذه الخصائص ، يمكن ان يكون وقت الحساب اسرع بشكل كبير مما كان عليه ان هذه الشرائح تحظى بشعبية في الرسوم البيانية للكمبيوتر نظرًا للسلاسة والمرونة والدقة التي تمتلكها. وان هذا الاسلوب يتضمن تحديد عدد من النقاط لربط المنحنيات ويطلق عليها العقد ومعظم الاحيان تكون من البيانات المتوفرة .

ان الهدف من هذا البحث هو توظيف دوال الشرائح (B-spline) مع كل من الشرائح الجزائية والانحدار في الخوارزمية الحصينة (S-estimate)، بالإضافة الى المقارنة بين طرائق التقدير هذه لاختيار افضل انموذج يمثل دالة الانحدار اللامعلمي ، وان هذه الطرائق تم تطبيقها بالجانب التجريبي.

2- الجانب النظري: -

2-1 نماذج الانحدار اللامعلمي Models Nonparametric Regression [3][7]

ان نماذج الانحدار اللامعلمي تمتلك مرونة اكبر من نماذج الانحدار المعلمي ، لهذا يقوم الباحثين بإعطاء وصف للعلاقة بشكل عام ، بدلا من دراسة التفاصيل الدقيقة للعلاقة . وان اسلوب الانحدار اللامعلمي يمثل الخطوة الاستكشافية في عملية النمذجة او يمثل المرحلة الاخيرة في اجراء عملية تحليل البيانات .يمكن تمثيل نموذج الانحدار اللامعلمي بالشكل التالي:

$$y_i = m(x_i) + \varepsilon_i \qquad i = 1,2,\cdots,n \ldots\ldots\ldots\ldots (1)$$

حيث ان :-

y_i : يشير الى قيم متغير الاستجابة عند i .

$m(x_i)$: تمثل دالة الانحدار المجهولة والمطلوب تقديرها.

ε_i : يشير الى قيمة الخطأ العشوائي الذي يتوزع توزيعا طبيعياً بمتوسط صفر وتباين ثابت وتتصف هذه الاخطاء بكونها مستقلة ومتماثلة.

2-2 الشرائح من النوع (B) (B-Spline) [4] [5]

يمكن تعريف شرائح (B-spline) على انها دالة متعددة الحدود من الدرجة k في متغير x. لقد تم التوصل اليها من قبل (Nikolai Lobachevsky) في مطلع القرن التاسع عشر ، وتمت اعادة صياغتها من قبل اسحاق يعقوب وهي تمثل اختصار لـ Basis-spline. ان هذا النوع من الشرائح يكون متعدد الحدود تربط في ما بينها نقاط ربط تسمى العقد ، ويجب ان تكون هذه العقد اقل من حجم العينة ويكون عددها صحيحا ، وان مفهوم الشريحة هو منحنى متعدد الحدود مستمر يستعمل لتقريب الحل لمشكلة رياضية. وان منحنى هذه الشريحة يعتمد على العلاقة بين دالة الاساس ونقاط التحكم. ان لهذا المنحنى دالة تعرف بـ دالة الاساس ويمكن كتابتها بالشكل الاتي $m(x) = \sum_{i=1}^{n+1} B_i N_{i,j}(x)$ حيث تشير B_i الى نقاط التحكم ، $N_{i,j}(x)$ (j) تمثل دالة الاساس، j تمثل درجة الاساس لدالة متعددة الحدود حيث ان n=j-1 . تتمثل خاصية منحنى B-spline في أنه يجب أن يقع بالكامل داخل الهيكل المحدب لـ تحديد المضلع .

2-3 ممهد انحدار شريحة B B- Spline Regression Smoothing [2] [9] [12]

يتم في انحدار شريحة b ، تقسيم مجال قيم المتغير التوضيحي.((x)) والذي يمكن تمثيله بالفترة [a,b] بواسطة عدد من المواقع التي تسمى بالعقد ويمكن التعبير عنها بالرموز الاتية:

$T_1, T_2, \ldots\ldots T_N, T_{N+1}$

اي ان الفترة [a,b] كان تكون ضمن مجال الدراسة وان $a = t_1, t_2, \ldots t_k, t_{k+1} = b$

حيث تشير $T = 1,2 \ldots \ldots N$ الى العقد(Konte) ان هذه العقد سوف تقسم المجال المهتمين بدراسته وكما اشرنا اليه سابقا[a,b] الى n من الجوار الموضعي ، وانT_i تسمى بالعقد الداخلية $i = 1,2 \ldots \ldots, N,$ اي ان اي عقدتين متجاورتين يستخدم متعدد حدود من رتبة معينة محدد مسبقاً يتم بناء انحدار الشريحة باستعمال ما يسمى بدوال الاساس من الدرجة $(i \times j)$ حيث ان

$$N_{i,0} = \begin{cases} 1 & t_i \le x \le t_{i+1} \\ 0 & \text{other wise} \end{cases}$$

$$m(x) = \sum_{i=0}^{n} \beta_i N_{i,j}(x) \ldots \ldots \ldots \ldots \ldots t_i \le x \le t_{i+1} \ldots \ldots \ldots (2)$$

$$N_{i,j}(x) = \frac{x - t_i}{t_{i+j} - t_i} N_{i,j-1}(x) + \frac{t_{i+j+1} - x}{t_{i+j+1} - t_{i+1}} N_{i+1,j-1}(x) \ldots \ldots \ldots (3)$$

حيث ان$N_{i,j}(x)$: تشير الى دالة اساس (الشريحة –B)i من الدرجة j , وان β_i تشير الى معاملات دالة الاساس .

ومن الجدير بالملاحظة ان N_{ij} هي متعددة حدود مقطعية من درجة اقل او يساوي j ولتكن p, ومجموع العقد يمكن ان يعبر عنها $T = [t_1, t_2, \ldots \ldots, t_{N+1}]$

ان النموذج في الصيغة (1) يمكن اعادة كتابته بدلالة انحدار الشريحة –B وكما ياتي :

$$Q = \sum_{i=1}^{N} \left(y_i - m(x_i)\right)^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{n} \beta_j N_{i,j}(x_i)\right)^2$$

Q : حيث يمكن تعريفها على انها مجموع مربعات البواقي.

β_j: تشير الى معاملات دوال الاساس ويمكن ان تكتب بالشكل الاتي :

$$\beta = \left[\beta_0, \beta_1, \ldots \ldots \beta_j\right]^T$$

ويمكن اعادة كتابة الانموذج الموضح بالصغية (1) بالشكل الاتي:

$$Y = X\beta + \varepsilon$$

حيث ان

$$Y = (Y_1, Y_2, \ldots \ldots Y_n)^T$$

$$x = \left(N_{0j}(x_1), N_{1,j}(x_2), \ldots \ldots \ldots N_{i,j}(x_m).\right)^T$$

اي ان $Y = [y_1, y_2 \ldots \ldots y_n]^T$ تشير الى قيم متغير المعتمد (الاستجابة) من درجة nx1 وان x تشير الى مصفوفة التصميم من درجة(i xj) ويمكن كتابتها بالشكل الاتي:-

$$X = \begin{bmatrix} N_{0,j}(x_1) & N_{1,j}(x_1) \ldots \ldots & N_{i,j}(x_1) \\ N_{0,j}(x_2) & N_{1,j}(x_2) \ldots \ldots & N_{i,j}(x_2) \\ \vdots & \vdots & \vdots \\ N_{0,j}(x_m) & N_{1,j}(x_m) \ldots \ldots \ldots & N_{i,j}(x_m) \end{bmatrix}$$

بما ان $N_{i,j}(x)$ تمثل دوال الاساس ، وان xهي من رتبة كاملة . لهذا يمكننا ايجاد معكوس المصفوفة (X^TX) عندما تكون $n \ge k$ اذا يمكن ان نوضح المقدر الطبيعي للمعاملات باستخدام طريقة المربعات الصغرى (ols)حيث يمكن كتابتها بالصيغة الاتية :

$$\widehat{\beta} = (X^TX)^{-1}X^Ty \ldots \ldots \ldots \ldots \ldots (4)$$

ينتج من الصيغة اعلاه منحنى ملائم لمنحنى لدالة الانحدار m وكما موضح بالصيغة الاتية

$$\widehat{m_R} = X^T(X^TX)^{-1}X^Ty \ldots \ldots \ldots \ldots \ldots (5)$$

ويسمى هذا عادة بتمهيد انحدار شريحة b للدالة m . لهذا ان قيم $\widehat{m}_p(x)$ المحتسبة عند نقاط التصميم(x_i) $i = 1,2,3, \ldots \ldots \ldots n$تكون مرتبطه مع متغير الاستجابة وكما موضع بالصيغة الاتية :

$$\hat{y} = (X^TX)^{-1}X^Ty = A_Ry \ldots \ldots \ldots \ldots (6)$$

اي ان

$$\hat{y}_R = (\hat{y}_1, \hat{y}_2, \ldots \ldots \hat{y}_n.)^T \ldots \ldots \ldots \ldots (7).$$

$$\hat{y}_i = \widehat{m}_R(x_i)$$

وان

$$A_R = X(X^TX)^{-1}X^T \ldots \ldots \ldots \ldots (8)$$

وهذا ما يسمى بمصفوفة انحدار الشريحة b تتطلب طريقة انحدار الشريحة تحديد مسبق لمواقع العقد والتي عددها (k+1)وهذا يؤدي الى ان استعمال مقدر انحدار الشريحة يتأثر باختيار مواقع هذه العقد وعددها. ان هذه العقد يتم وضعها عادةً في المواقع التي يكون عندها انحناءات المنحنى واضحة وتكون ذات تغيرات كبيرة . هناك عدة طرائق من الممكن اتباعها في تحديد عدد ومواقع العقد.

2-4 ممهد انحدار شريحة B الجزائية Penalize B- Spline Smoothing [12] [8]

ان ذا الأنموذج قدم من قبل Eilers&marks سنة 1986 واطلق علية تسمية P-Spline.ويكون مع عقد بمسافات تكون متساوية ولكن عددها في الغالب يكون اقل من البيانات. وكما بينا سابقا في تمهيد الشريحة من الممكن معالجة الاخفاق في التمهيد ، والذي حصل بسب العدد الكبير للعقد وذلك اضافة حد الجزاء لمعيار المربعات الصغرى ، ومن ذلك يمكن توضيح الشريحة الجزائية من خلال ايجاد المقدر^m وكما ياتي:-

$$\sum_{i=1}^{n}\left[y_i-\sum_{i=1}^{n+1}B_i\,N_{i,j}(x)\right]^2+\lambda\int_{x\_min}^{x_{mak}}\left\{\sum_{i=1}^{n+1}B_i\,N_{i,j}(x)\right\}^2 dx\,\ldots\ldots\ldots(9)$$

تشير λ الى معلمة التمهيد وتكون قيمتها اكبر من الصفر وان الجزء الاول من الحد يمثل مجموع مربعات البواقي، والجزء الثاني من الحد يمثل جزاء الخشونة الذي يكون مفروضا على الجزء غير الممهد من المنحني

ويمكن تعريف ممهد الشريحة الجزائية الحصينة من ايجاد مقدر ^m الذي يقلل المقدرالتالي:

$$\frac{1}{n}\sum_{i=1}^{n}p\left[y_i-\sum_{i=1}^{n+1}B_i\,N_{i,j}(x)\right]^2+\lambda\int_{x\_min}^{x_{mak}}\left\{\sum_{i=1}^{n+1}B_i\,N_{i,j}(x)\right\}^2 dx$$

نلاحظ ان هذه الصيغة تكون مشابهة للصيغة (2-13)، ولكن مع وجود الدالة p بالإضافة الى انها مضروبة بمقلوب حجم العينة . وكما نعلم ان λ تمثل معلمة التمهيد وتكون قيمتها اكبر من الصفر . ويمكن ان نحصل على قيمة p من المعادلة الاتية والتي يطلق عليها معادلة هوبر:

$$p_c(x)=\begin{cases}x^2 & \mathrm{IF}|x|\le c\\ 2c|x|-c^2 & \mathrm{IF}|x|>c\end{cases}$$

حيث ان Cتمثل عامل قياس ويمكن ايجادها وفق الصيغة الاتية :

$$C=10^{-1}\sqrt{\mathrm{var}(x)}$$

وبالاعتماد على مصفوفة التصميم حيث ان$y=[y_1,y_2,\ldots\ldots\ldots..y_n]^T$ يشير الى متجه قيم المتغير الاستجابة ، وان X تشير الى مصفوفة (ixj). ويمكن كتابتها بالصيغة الاتية :

$$X=\begin{bmatrix}N_{0,j}(x_1) & N_{1,j}(x_1)\ldots\ldots.. & N_{i,j}(x_1)\\ N_{0,j}(x_2) & N_{1,j}(x_2)\ldots\ldots.. & N_{i,j}(x_2)\\ \vdots & \vdots & \vdots\\ N_{0,j}(x_m) & N_{1,j}(x_m)\ldots\ldots\ldots. & N_{i,j}(x_m)\end{bmatrix}$$

ومن المصفوفة اعلاه يمكن القول انه ليس هنالك اي معاملات سوى معاملات الشريحة الجزائية. ويمكن كتابة المصفوفة القطرية بالشكل الاتية :

$$D=\mathrm{diag}(0_{p+1},1_k)$$

وعليه فان مقدر دالة الهدف يمكن حسابه باستعمال الحسابات المباشرة وكالاتي :

$$\hat{\beta}_{ls}=(X^TX+\lambda D)^{-1}X^Ty\ldots\ldots\ldots.(10)$$

تشير ^β الى مقدرات المربعات الصغرى لدوال اساس القطع. وعلية فان تقدير المتجه المقابل يمكن الحصول عليه باستعمال شرائح انحدار المربعات الصغرى الجزائية، لذا يمكن كتابة الصيغة الخاصة لمقدر المتجه كما يأتي :

$$\hat{m}_{ls}=X(X^TX+\lambda D)^{-1}X^TY\ldots\ldots\ldots(11)$$

ولغرض حساب معلمة الجزاءλ في الصيغة اعلاه تم استعمال معيار العبور الشرعي المعمم كما في الصيغة الاتية :

$$GCV(\lambda)=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-g_i\hat{}(x_i,\lambda)\right)^2\ldots\ldots\ldots..(12)$$

$g_i\hat{}(x_i,\lambda)$تشير الى انحدار الشريحة الذي تم تقديره عن طريق ترك المشاهدة j خارجا، بالإضافة الى ذلك يمكن تقدير معلمة الجزاء المثلى عن طريق تقليل معيار (GCV) عندما تكون $0 \geq \lambda$. ويمكن كتابة مصفوفة التمهيد الخطية $H_i = X(X^TX+\lambda G)^{-1}X^T$ومن خلال هذه الصيغة يمكن كتابة الصيغة التالية :

$$g_i\hat{}(x_i,\lambda) = \sum_{j\neq i}\left(\frac{H^{ij}{}_{\lambda}}{1-H^{ij}{}_{\lambda}}y_i\right) \ldots\ldots\ldots\ldots\ldots (13)$$

ويمكن استبدال المقدر $g_i\hat{}(x_i,\lambda)$ بما يساويه بالصيغة (9):

$$GCV(\lambda) = \frac{1}{n}\sum_{j\neq i}\left(\frac{\left(y_i - g_i\hat{}(x_i,\lambda)\right)}{1-H^{ij}{}_{\lambda}}y_i\right)^2 \ldots\ldots\ldots\ldots (14)$$

ويمكن استبدال المصفوفة $H^{ij}{}_{\lambda}$ بمعدل اثر المصفوفة ، لذا يمكن اعادة كتابة الصيغة اعلاه كما ياتي

$$GCV(\lambda) = n\sum_{j\neq i}\left(\frac{\left(y_i - g_i\hat{}(x_i,\lambda)\right)}{n-tr(H_{\lambda})}y_i\right)^2 \ldots\ldots\ldots (15)$$

عندما يكون لدينا مجموعة من البواقي الملائمة$r_i = y_i - m(x_i - \beta)$ من انحدار الشريحة الجزائية ، يمكننا توظيف مقدر معيار الحصانة لكي يتم تقدير الانحراف المعياري لتلك البواقي وكما يلي :

$$\sum_{i=1}^{n}\psi_c\left(\frac{y_i - m(x_{i,}\beta)}{\widehat{\sigma_{\varepsilon}}}\right)$$

$$p(|r - r_{0.5}| < \omega) = \frac{1}{2}$$

ωتمثل قيمة نقطة توزيع عينة محدودة، $r_{0.5}$ تمثل وسيط المجتمع ، اضافة الى ذلك فان ω يمكن ان يتم استعمالها قي حساب الانحراف المطلق للوسيط . ويمكن كتابته بالصيغة الاتية :

$$MAD = Median(|r_1 - M_1|, \ldots, |r_n - M_n|)$$

إذ أنM_r يمثل وسيط العينة الاعتيادي للبواقي الملائمة، وأنMAD هو الوسيط للعينة ولكنn $(|r_1 - M_1|, \ldots, |r_n - M_n|)$ مع نقطة توزيع العينة المحدودة وهو ما يقارب 0.5 وعلى افتراض أن البوقي يتم اختيارها بشكل عشوائي من التوزيع الطبيعي، مع ملاحظة أنه لا يتم تقدير الانحراف المعياري$\widehat{\sigma_{\varepsilon}}$ باستخدام الانحراف المطلق للوسيط ولكن بدلاً عن ذلك يتم استخدام تقديرات $\widehat{\sigma_{\varepsilon}}z_{0.75}$(0.75 تمثل مجزاءات التوزيع الطبيعي القياسي).

2-5 طريقة التحصين (S-estimate) [11]

وصفت هذه المقدرات من قبل (Rousseeum,yohai) عام 1984ويمكن كتابة الخوارزمية الخاصة بعمل هذه الطريقة وهي كالاتي :

1-2-5 خوارزمية S-estimate لتقدير (Regression B-Spline)

وتتلخص هذه الطريقة بالخطوات الاتية :

1- يتم تحديد قيمة اولية للدالة $\widehat{m_1}$ بالإضافة الى تحديد قيمة التكرارات ويمكن اعتماد تقدير انحدار الشريحة من خلال المعادلة:-

$$\widehat{m} = x(\acute{X}X)^{-1}\acute{X}y \ldots\ldots (16).$$

2- ايجاد البواقي (residuals) من خلال الصيغة الاتية:-

$$r_i = y_i - \widehat{m}_i \ldots\ldots\ldots (17).$$

3- تقدير الانحراف المعياري $\widehat{\sigma_{\varepsilon}}$ وحسب الصيغة الاتية:-

$$\widehat{\sigma_{\varepsilon}} = MADN + median(r_i) \ldots\ldots (18)$$

حيث ان :-$MADN = 1.4836MAD$

و ان MAD يمكن ايجادة من خلال الصيغة الاتية :-

$$MAD = median|r_i - median(r_i)|$$

4- حساب قيمة $\tilde{r}$ من المعادلة الاتية: -

$$\tilde{r} = \frac{r_i}{\widehat{\sigma_{\varepsilon}}} \ldots\ldots\ldots\ldots\ldots (19)$$

5- ايجاد قيمة Λ من خلال المعادلة الاتية: -

$$\Lambda = \mathrm{diag}(\acute{\rho}(\widetilde{r_i})/r_i)$$

6- حساب

$$\tau = (n\widehat{\sigma_\varepsilon}/r^T\ \Lambda\ r) \quad ....(20)$$

7- حساب $\widehat{\beta}$ من خلال المعادلة الاتية: -

$$\widehat{\beta} = \left(x^T\Lambda X + \frac{1}{\tau}\right)^{-1} X^T y$$

8- جعل $\widehat{m_i} = \widehat{\beta}$ ، ونقوم بتكرار الخطوات من الخطوة (2)الى (7) ونتوقف حتى نحصل على تقارب في التقدير

2-5-5 خوارزمية S-estimate لتقدير (Penajized B-Spline)

تتلخص هذه الطريقة بالخطوات الاتية :

1- يتم تحديد قيمة اولية للدالة $\widehat{m_i}$ بالإضافة الى تحديد قيمة التكرارات ويمكن اعتماد تقدير انحدار الشريحة من المعادلة: -

$$\widehat{m} = x(\acute{X}X + \lambda D)^{-1}\acute{X}y \ldots\ldots(21).$$

2- ايجاد البواقي (residuals) من الصيغة الاتية: -

$$r_i = y_i - \widehat{m}_i \ldots\ldots\ldots(22).$$

3- تقدير الانحراف المعياري $\widehat{\sigma_\varepsilon}$ وحسب الصيغة الاتية: -

$$\widehat{\sigma_\varepsilon} = \mathrm{MADN} + \mathrm{median}(r_i) \ldots\ldots(23)$$

حيث ان :- $\mathrm{MADN} = 1.4836\mathrm{MAD}$

و ان MAD يمكن ايجاد من خلال الصيغة الاتية :-

$$\mathrm{MAD} = \mathrm{median}|r_i - \mathrm{median}(r_i)|$$

4- حساب قيمة $\tilde{r}$ من المعادلة الاتية: -

$$\tilde{r} = \frac{r_i}{\widehat{\sigma_\varepsilon}} \ldots\ldots\ldots(24)$$

5- ايجاد قيمة Λ من خلال المعادلة الاتية :-

$$\Lambda = \mathrm{diag}(\acute{\rho}(\widetilde{r_i})/r_i)$$

6- حساب

$$\tau = (n\widehat{\sigma_\varepsilon}/r^T\ \Lambda\ r) \quad ....(25)$$

7- حساب $\widehat{\beta}$ من المعادلة الاتية: -

$$\widehat{\beta} = \left(x^T\Lambda X + \frac{\lambda}{\tau}D\right)^{-1} X^T y$$

8- جعل $\widehat{m_i} = \widehat{\beta}$ ، ونقوم بتكرار الخطوات من الخطوة (2)الى (7) ونتوقف حتى نحصل على تقارب في التقدير

3- الجانب التجريبي

يمكن تعريف المحاكاة على انها عملية تقليد للواقع العملي حيث يمكن من خلالها ايجاد الحلول للمشكلات الرياضية ، حيث يتم بناء انموذج مشابه للنموذج الاصلي ومن ثم تطبيق المعاينة عليه . وقد تم استعمال هذا الاسلوب كثيرا نتيجة للتقدم الحاصل في مجال الحاسبات الالكترونية بالإضافة الى ذلك تم استعمالها في مجالات الاحصاء المختلفة لتطوير ودراسة العديد من الطرائق الاحصائية المختلفة.

الدوال التي تم استعمالها في تجارب المحاكاة:

في هذا البحث تم استعمال ثلاث دول لتمثيل النموذج الاصلي وهي كالاتي:-

1- دالة لاخطية[1] [6]

$$\mathbf{m(x) = x + 2exp(-16x^2)}$$

2- دالة من الدرجة الثانية [10]

$$\mathbf{m(x) = 8(x - 0.5)^2}$$

3- دالة من لاخطية [7]

$\mathbf{m}\ (\mathbf{x}) = \ \mathbf{sin}(2\mathbf{x}) + 2\ \mathbf{exp}(-2\mathbf{x}^2)$

تم تنفيذ تجارب المحاكاة باستعمال ثلاث حجوم للعينات ($\mathbf{n_1}$=50, $\mathbf{n_2}$=100, $\mathbf{n_3}$=150) كما تم تلويث البيانات بنسب (20% , 16%, 10%)

جدول (3) يبين المعدل لقيم معيار المقارنة (MAE) ولجميع حجوم العينات ومستويات التلويث لمقارنة طريقة B- Spline في انحدار الشريحة($\mathbf{B_{RS}S}$) وB- Splineلانحدار الشريحة الجزائية($\mathbf{B_{PS}S}$) بالنسبة لطريق التحصين S للأنموذج الاول

Sample size	طرائق التقدير	مستويات التلويث		
		10%	16%	20%
$\mathbf{n_1}$=50	$\mathbf{B_{RS}S}$	0.86657	0.85149	0.84380
	$\mathbf{B_{PS}S}$	0.30368	0.28741	0.28581
$\mathbf{n_2}$=100	$\mathbf{B_{RS}S}$	0.77856	0.75957	0.74920
	$\mathbf{B_{PS}S}$	0.43643	0.42076	0.41220
$\mathbf{n_3}$=150	$\mathbf{B_{RS}S}$	0.73229	0.71619	0.70303
	$\mathbf{B_{PS}S}$	0.42622	0.41386	0.40408

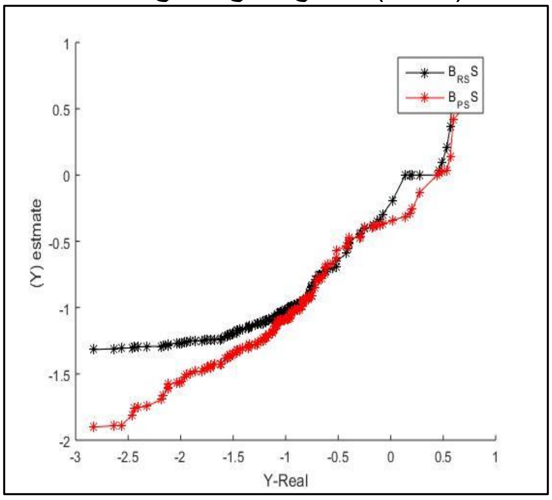
جدول (3) يبين المعدل لقيم معيار المقارنة (MAE) ولجميع حجوم العينات ومستويات التلويث لمقارنة طريقة B- Spline في انحدار الشريحة($\mathbf{B_{RS}S}$) وB- Splineلانحدار الشريحة الجزائية($\mathbf{B_{PS}S}$) بالنسبة لطريق التحصين S للأنموذج الثاني

Sample size	طرائق التقدير	مستويات التلويث		
		10%	16%	20%
$\mathbf{n_1}$=50	$\mathbf{B_{RS}S}$	0.85383	0.84527	0.83456
	$\mathbf{B_{PS}S}$	0.30867	0.30341	0.29308
$\mathbf{n_2}$=100	$\mathbf{B_{RS}S}$	0.77665	0.75806	0.75369
	$\mathbf{B_{PS}S}$	0.43964	0.42465	0.41901
$\mathbf{n_3}$=150	$\mathbf{B_{RS}S}$	0.73788	0.71801	0.70407
	$\mathbf{B_{PS}S}$	0.43198	0.41728	0.40729

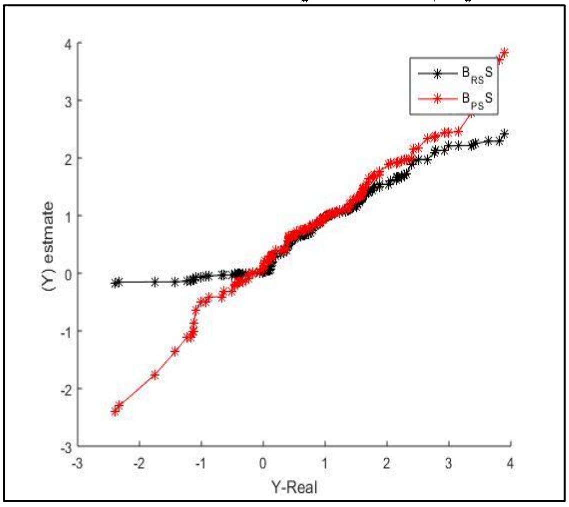
جدول (3) يبين المعدل لقيم معيار المقارنة (MAE) ولجميع حجوم العينات ومستويات التلويث لمقارنة طريقة B- Spline في انحدار الشريحة($\mathbf{B_{RS}S}$) وB- Splineلانحدار الشريحة الجزائية($\mathbf{B_{PS}S}$) بالنسبة لطريق التحصين S للأنموذج الثالث

Sample size	طرائق التقدير	مستويات التلويث		
		10%	16%	20%
$\mathbf{n_1}$=50	$\mathbf{B_{RS}S}$	0.82119	0.80586	0.79726
	$\mathbf{B_{PS}S}$	0.27895	0.26349	0.26243
$\mathbf{n_2}$=100	$\mathbf{B_{RS}S}$	0.73218	0.72513	0.70859
	$\mathbf{B_{PS}S}$	0.42095	0.40924	0.39986
$\mathbf{n_3}$=150	$\mathbf{B_{RS}S}$	0.69819	0.68017	0.66051
	$\mathbf{B_{PS}S}$	0.41920	0.40540	0.39290

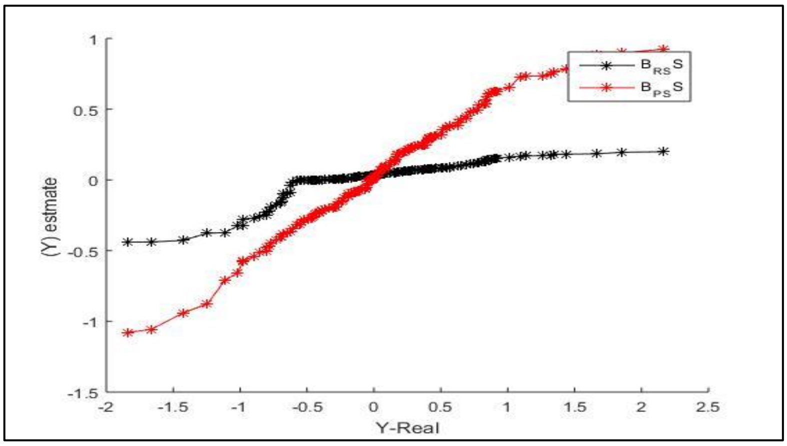
يمكن ملاحظة ان طريقة (B-Spline Penalize) باستعمال طريقة التحصين (S-estimate) كانت الافضل في التقدير و سجلت اقل قيمة بالنسبة لمعيار المقارنة معدل لمتوسط الخطأ المطلق (MAE) ولجميع نماذج المحاكاة ومستويات التلويث التي تم استعمالها في هذا الجانب .



شكل رقم (1) يبين القيم الحقيقية والقيم التقديرية لمقدرات النموذج الاول $(B_{RS}S)(B_{pS}S)$ عند حجم العينة n=150



شكل رقم (2) يبين القيم الحقيقية والقيم التقديرية لمقدرات النموذج الثاني $(B_{RS}S)(B_{pS}S)$ عند حجم العينة n=150



شكل رقم (3) يبين القيم الحقيقية والقيم التقديرية لمقدرات النموذج الثالث $(B_{RS}S)(B_{pS}S)$

عند حجم العينة n=150

4 - الاستنتاجات والتوصيات

يمكن عرض اهم الاستنتاجات التي تم التوصل اليها الباحث في الجانب التجريبي وهي كما يلي :-

1- بالنسبة للجانب التجريبي ومن خلال مقارنة النتائج للنماذج الثلاثة تبين ان طريقة التمهيد الافضل كانت تمهيد الشريحة الجزائية لحجوم العينات الثلاثة ولجميع مستويات التلويث

2- من خلال نتائج المحاكاة يمكن ملاحظة القيمة المعيارية للنماذج الثلاثة حيث انها بدأت تقل عند ازدياد نسب التلويث

3- يمكن تطبيق شرائح (M-Spline , I-Spline , T-Spline) . في تقدير دوال الانحدار اللامعلمي .

4- يمكن استعمال طرائق تحصين اخرى (MM-estimate, L- estimate) في حاله هناك قيم ملوثة في البيانات (المتغير التوضيحي).

5 - المصادر

1- خمو، خلود يوسف (2004) " مقارنة أساليب بيز مع طرائق أخرى لتقدير منحنى الانحدار اللامعلمي " أطروحة دكتوراه في الإحصاء ، كلية الإدارة والاقتصاد ، جامعة بغداد.
2- رشيد ،حسام عبد الرزاق (2014) ، " الممهدات اللامعلمية لأنموذج المعاملات المتغيرة والمتغيرة جزئيا" اطروحة دكتوراه في الاحصاء. كلية الادارة والاقتصاد ، جامعة بغداد.
3- مجيد ،غياث حميد (2016) ، " تحديد افضل اسلوب تمهيدي لتقدير أنموذج انحدار لامعلمي " اطروحة دكتوراه في الاحصاء . كلية الادارة والاقتصاد ، جامعة بغداد.

4- Eilers, P.H.C. and Marx, B.D. (1996)." Flexible smoothing with B-splines and penalties (with comments and rejoinder)". Statistical Science 11(2): 89-121.

5- Gálvez, A & Iglesias, A (2013). "Firefly Algorithm for Explicit B-Spline Curve Fitting to Data Points." Applied Mathematics and Computations Volume 2013, Article ID 528215, 12 pages.

6- Härdle, Wolfgang, (1994), "Applied Nonparametric Regression". Cambridge: Cambridge University Press.

7- Indra, D . P (2020). "A Comparison between Nonparametric Approach: Smoothing Spline and B-Spline to Analyze The Total of Train Passengers in Sumatra Island." Applied Mathematics and Computations Volume 1, Issue 1, 73-800.

8- Jator, Samuel & Zachariah ,Sinkala(2007). "A Higher Order B-spline Collocation Method for Linear Boundary Value Problems." Applied Mathematics and Computations 191: 100-116.

9- Johnson, R.W., (2005). "A B-spline Collocation Method for solving the Incompressible Navier- Stokes Equations Using an ad hoc Method: the Boundary Residual Method," Computers& Fluids 34: 121-149

10- Wu, H. and Zhang, J., (2006), "Nonparametric regression methods for longitudinal data analysis: Mixed-Effects modeling approaches", John Wiley & Sons, New Jersey

11- Susanti, Y. &Pratiwi, H.,(2013)," M estimation, S estimation, and MM estimation in robust regression" International Journal of Pure and Applied Mathematics Volume 91 No. 3 2014, 349-360

12- Wang, B. &Miao, Z.(2014)." Comparative Analysis for Robust Penalized Spline Smoothing Method" Volume 2014, 11 pages

……………………………………………………………………………………….
....
....

Using the S-estimate method with B-Spline to estimate the nonparametric regression pattern (pilot study)

Researcher / Ghaida Ibrahim Shehab / University of Baghdad / College of Administration and Economics
Gheadaa.ahmed1201@coadec.uobaghdad.edu.iq
L. Dr. Nazik Jaafar Sadiq/University of Baghdad/College of Administration and Economics
dr.nazik@coadec.uobaghdad.edu.iq

Abstract

One of the most common bootstrap methods is B-splines The use of these slices has become very popular among many fields of mathematics, engineering, and computer science in recent years. Originally, B slices were used for rounding purposes, but their popularity and applications have expanded. In addition, this technique is illustrative and highly flexible, unlike the normal slice technique or other smoothing methods.

In this research, some smoothing methods will be used, including smoothing the regression of the segment and the penalty segment using the B-spline, in order to obtain an estimate of the non-parametric regression model. The research aims to find the best estimator from among the mentioned primer estimators. To represent the studied data based on the results of simulation experiments, and through the experimental side, it was concluded that the method of smoothing the penalty slide using the B-spline was the best in estimating the nonparametric regression model.

Keywords: Nonparametric regression models, Regression B-spline method, Penalize B-Spline estimation method, B-Spline segments, S-estimate method.
The research is extracted from a master's thesis.

****
****
****