



فرائض للعلوم الاقتصادية والإدارية
**KHAZAYIN OF ECONOMIC AND
ADMINISTRATIVE SCIENCES**

ISSN: 2960-1363 (Print)

ISSN: 3007-9020 (Online)



Comparison between B-Spline Regression and Penalized B-Spline Using M-Estimator Robust Method for Estimation Non-Parametric Regression Model

Gheadaa Ibrahim Sheab¹, Arshad Hameed Hasan²

^{1,2} College of Administration and Economics, University of Diyala, 32001, Diyala, Iraq

gh.gheadaa@uodiyala.edu.iq

arshadhameed@uodiyala.edu.iq

Abstract. Bootstrapping methods in nonparametric regression have garnered significant attention in recent years, with numerous distinct bootstrapping approaches proposed for estimating regression models. These models can be causal, built upon the principle of cause and effect, termed parametric regression models, which adhere to specific assumptions and take a functional form. However, in situations where knowledge about the phenomena under study is lacking and establishing a causal function between variables is not feasible, another type of regression model emerges, known as nonparametric regression. regression and penalized spline regression using B-spline basis functions, to obtain estimates for the nonparametric regression model. The study aims to identify the optimal estimator among the aforementioned smoothing estimators for representing the studied data, based on the results of simulation experiments. Through the empirical analysis, it was concluded that the penalized spline smoothing method using B-spline basis functions was the most effective in estimating the nonparametric regression model. In this research, several smoothing techniques were employed, including spline

Keywords: Nonparametric Regression, B-Spline, Spline Regression, Penalized Spline Regression, M-estimator Immunization Method

DOI: 10.69938/Keas.Con2.250227

المقارنة بين طريقتي (Penalize B-Spline&Regression B-Spline) باستعمال طريقة التحسين (M -estimate) لتقدير أنموذج الانحدار اللامعلمي

م.م غيداء ابراهيم شهاب¹ ، م.م ارشد حميد حسن
جامعة ديالى كلية الادارة والاقتصاد - قسم الاحصاء

gh.gheadaa@uodiyala.edu.iq

arshadhameed@uodiyala.edu.iq

المخلص: جذبت طرائق التمهيد في الانحدار اللامعلمي قدرا كبيرا من الاهتمام في السنوات الاخيرة، حيث تم اقتراح العديد من طرائق التمهيد المختلفة لكي يتم تقدير نماذج الانحدار، 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 نوع اخر من نماذج الانحدار يسمى بالانحدار اللامعلمي. في هذا البحث تم 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 إيجاد 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-splaine كان الافضل في تقدير أنموذج الانحدار اللامعلمي.

الكلمات المفتاحية: الانحدار اللامعلمي، شريحة B-Spline ، انحدار الشريحة، انحدار الشريحة الجزائية، طريقة التحسين M-estimate

Corresponding Author: E-mail: gh.gheadaa@uodiyala.edu.iq

المقدمة

تظهر أهمية 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 نوع اخر من نماذج الانحدار يسمى بالانحدار اللامعلمي . وهذا النوع من نماذج الانحدار لا تأخذ قيمة المتغير المستقل شكلا محدداً 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 شائعاً جداً 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 طرق 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 نظراً للسلاسة والمرونة والدقة التي تمتلكها. وان هذا الاسلوب يتضمن تحديد عدد من النقاط لربط المنحنيات ويطلق عليها العقد ومعظم الاحيان تكون من البيانات المتوفرة .

2-هدف البحث

- 1- توظيف دوال الشرائح B - مع كل من الشرائح الجزائية والانحدار في الخوارزمية الحصينة M و S .
- 2- مقارنة الطرائق اللامعلمية الحصينة لتقدير انموذج الانحدار واختيار افضل طريقة تقدير من خلال الاسلوب التجريبي .
- 3- النمذجة اللامعلمية لتقلبات اسعار اسهم الشركة.

3-الجانب النظري:-

نماذج الانحدار اللامعلمي [1][7]

ان نماذج الانحدار اللامعلمي تمتلك مرونة أكبر 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:

$$y_i = g(x_i) + \varepsilon_i \quad i = 1, 2, \dots, n \dots \dots \dots (1)$$

حيث ان :-

y_i : يشير الى قيم متغير الاستجابة عند i .
 $g(x_i)$: تمثل دالة الانحدار المجهولة والمطلوب تقديرها .
 ε_i : يشير الى قيمة الخطأ العشوائي الذي يتوزع توزيعاً طبيعياً بمتوسط صفر وتباين ثابت وتتصف هذه الاخطاء بكونها مستقلة ومتماثلة.

والهدف الاساسي في هذا النموذج هو تقدير دالة الانحدار g والتي يشار إليها بدالة الانحدار أو منحني الانحدار عند النقاط $x_1, \dots, x_n$. وأن اسلوب منحني التقريب هذا يسمى بالتمهيد (Smoothing) .

1-3 مهاد انحدار الشريحة [3][14][17]

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

$$T_1, T_2, \dots, T_k, T_{k+1} \dots \dots$$

اي ان الفترة $[a, b]$ كان تكون ضمن مجال الدراسة وان $a = t_1, t_2, \dots, t_k, t_{k+1} \dots = b$ حيث تشير $T = 1, 2, \dots, K$ الى العقد (Kont) ان هذه العقد سوف تقسم المجال المهتمين بدراستها وكما اشرنا اليها سابقا $[a, b]$ الى k من الجوار الموضعي ، وان T_i تسمى بالعقد الداخلية ، $i = 1, 2, \dots, K$ 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, x_j) حيث ان

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

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

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

حيث ان $N_{i,j}(x)$: تشير الى دالة اساس (الشريحة -B) من الدرجة j , وان β_i تشير الى معاملات دالة الاساس .
ومن الجدير بالملاحظة ان $N_{i,j}$ هي متعددة حدود مقطعية من درجة اقل او يساوي j ولتكن p , ومجموع العقد يمكن ان يعبر عنها $T = [t_1, t_2, \dots, t_{N+1}]$ وان $N \geq 0$.
ان النموذج في الصيغة (2-1) يمكن اعاده كتابته بدلالة انحدار الشريحة -B وكما ياتي :

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

Q : حيث يمكن تعريفها على انها مجموع مربعات البواقي.
 β_j : تشير الى معاملات دوال الاساس ويمكن ان تكتب بالشكل الاتي :

$$\beta = [\beta_0, \beta_1, \dots, \beta_j]^T$$

ويمكن اعادة كتابة الانموذج الموضح بالصيغة (2-1) بالشكل الاتي:

$$Y = X\beta + \varepsilon \quad (4)$$

حيث ان

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

$$x = (N_{1,j}(x_1), N_{2,k}(x_2), \dots, N_{n,j}(x_n))^T$$

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

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

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

$$\hat{\beta} = (X^T X)^{-1} X^T y \dots \dots \dots (5)$$

ينتج من خلال الصيغة اعلاه منحنى ملائم لمنحنى لدالة لانحدار الشريحة m وكما موضح بالصيغة الاتية

$$\hat{m}_R = X^T (X^T X)^{-1} X^T y \dots \dots \dots (6)$$

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

$$\hat{y} = (X^T X)^{-1} X^T y = A_R Y \dots \dots \dots (7)$$

اي ان

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

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

وان

$$A_R = X(X^T X)^{-1} X^T y \dots \dots \dots (9)$$

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

6-2 ممد انحدار شريحة B-Spline الجزائية [6] [43]

ان هذا الأنموذج قدم من قبل *Eilers&marks* سنة 1986 واطلق عليه 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 بسبب العدد الكبير للعقد وذلك من خلال اضافة حد الجزاء لمعيار المربعات الصغرى ، ومن خلال ذلك يمكن توضيح الشريحة الجزائية من خلال ايجاد المقدّر $\hat{m}$ وكما يأتي:-

$$\sum_{i=1}^n \left[y_i - \sum_{k=1}^{n+1} B_k N_{i,K}(x) \right]^2 + \lambda \int_{x_{\min}}^{x_{\max}} \left\{ \sum_{k=1}^{n+1} B_k N_{i,K}(x) \right\}^2 dx \dots \dots \dots (10)$$

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 ممد الشريحة الجزائية الحصينة من خلال ايجاد مقدار $\hat{m}$ الذي يقلل المقدّر التالي:

$$\frac{1}{n} \sum_{i=1}^n p \left[y_i - \sum_{k=1}^{n+1} B_k N_{i,K}(x) \right]^2 + \lambda \int_{x_{\min}}^{x_{\max}} \left\{ \sum_{k=1}^{n+1} B_k N_{i,K}(x) \right\}^2 dx$$

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 خلال المعادلة الآتية والتي يطلق عليها معادلة هوبر:

$$p_c(x) \begin{cases} x^2 & IF |X| \leq C \\ 2C|x| - C^2 & IF |X| > C \end{cases}$$

حيث ان C تمثل عامل قياس ويمكن ايجادها وفق الصيغة الآتية :

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

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

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

ومن خلال هذه المصفوفة اعلاه ومصفوفة القطرية يمكن القول انه ليس هنالك اي معاملات سوى معاملات الشريحة الجزائية. ويمكن كتابة المصفوفة القطرية بالشكل الآتية :

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

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

$$\hat{\beta}_{ls} = (X^T X + \lambda D)^{-1} X^T \dots \dots \dots (2 - 14)$$

تشير $\hat{\beta}$ الى مقدرات المربعات الصغرى لدوال اساس القطع. وعليه فان تقدير المتجه المقابل يمكن الحصول عليه باستعمال شرائح انحدار المربعات الصغرى الجزائية، لذا يمكن كتابة الصيغة الخاصة لمقدّر المتجه كما يأتي :

$$\hat{m}_{ls} = X(X^T X + \lambda D)^{-1} X^T Y \dots \dots \dots (11)$$

ولغرض حساب معلمة الجزاء λ في الصيغة اعلاه تم استعمال معيار العبور الشرعي المعمم كما في الصيغة الآتية :

$$GCV(\lambda) = \sum_{i=1}^n \left(\frac{y_i - \hat{m}_i}{n - \text{tr}(H)} \right)$$

$$H = X(X^T X)^{-1} X^T \quad \text{حيث ان :}$$

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

$$\sum_{i=1}^n \psi_c \left(\frac{y_i - m(x_i, \beta)}{\hat{\sigma}_\varepsilon} \right)$$

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

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

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

إذ أن M_n يمثل وسيط العينة الاعتيادي للبواقي الملائمة، وأن MAD هو الوسيط للعينة ولكن n ($|r_1 - M_1|, \dots, |r_n - M_n|$) 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 $\hat{\sigma}_\varepsilon$ باستخدام الانحراف المطلق للوسيط ولكن بدلاً عن ذلك يتم استخدام تقديرات $z_{0.75} \hat{\sigma}_\varepsilon$ (0.75) تمثل مجزئات التوزيع الطبيعي القياسي)

3-3 الشرائح من النوع (B) [4] [1] [16]

يمكن تعريف شرائح ($B-spline$) على أنها دالة متعددة الحدود من الدرجة k في متغير x . لقد تم التوصل إليها من قبل نيكولا لوباتشيفسكي في مطلع القرن التاسع عشر ، وتمت إعادة صياغتها من قبل اسحاق يعقوب وهي تمثل اختصار ل $Basis-spline$. 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 . ويجب ان تكون عدد صحيح ، وان مفهوم الشريحة هو منحنى متعدد الحدود مستمر يستخدم لتقريب الحل لمشكلة رياضية. وان منحنى هذه الشريحة يعتمد على العلاقة بين دالة الاساس ونقاط التحكم. ان لهذا المنحنى دالة تعرف ب دالة الاساس ويمكن كتابتها بالشكل الآتي $p(x) = \sum_{i=1}^{n+1} B_i N_{i,K}(x)$ حيث تشير B_i الى نقاط التحكم التحكم ، $N_{i,K}(x)$ تمثل دالة الاساس، K تمثل درجة الاساس لدالة متعددة الحدود حيث ان $n=k-1$. تتمثل خاصية منحنى $B-spline$ في أنه يجب أن يقع بالكامل داخل الهيكل المحدب لـ تحديد المضلع .

حساب منحنيات $B-Spline$ [6] [8] [10] [11] [12]

لكي نحصل على منحنى ($B-Spline$). يجب ان يتوفر لدينا سلسلة من العقد ، سلسلة من نقاط التحكم . بحيث كل نقطة تحكم تتوافق مع معامل ، ولكي يتم تحقيق شرط الاستمرارية لابد من توصيل مقاطع منحنى ($B-Spline$).

من أجل حساب منحنيات $B-spline$ ، يلزم وجود ثلاثة أشياء

- ❖ متجه العقد.
- ❖ دالة الاساس.
- ❖ نقاط التحكم.

بمجرد تحديد متجه العقدة ودوال الاساس ونقاط التحكم ، يمكن اشتقاق منحنى $B-spline$. وبالتالي المنحنى الناتج هو المنحنى التقريبي لحل مشكلة تقدير دالة الانحدار . ويمكن كتابة دالة منحنى $b-spline$ بالشكل الآتي :

$$g(x) = \sum_{i=1}^{n+1} B_i N_{i,K}(x) \dots \dots \dots t_i \leq x \leq t_{i+1} \dots \dots (12)$$

❖ متجه العقد (Knot Vectors)

يعد اختيار متجه العقدة أمراً مهماً جداً لإنشاء منحنى $B-spline$. وان اختيار هذا المتجه يحدد وفق درجة متعدد الحدود وحجم الفترات، بالإضافة الى ذلك ان نعومه المنحنى تحدد تبعاً لاختيار متجه العقد. هناك ثلاثة أنواع مختلفة من متجهات العقد المستخدمة.

منتظم: تكون العقد متباعدة بشكل متساوي.
المنتظم المفتوح: يكون عدد قيم العقد مساوي الى درجة منحنى $b-spline$ ، وقيم العقد الداخلية تكون متباعدة بشكل متساوية.

غير المنتظم: وتكون عدد قيم العقد متباعدة بشكل غير متساوي.

في هذه الرسالة تم استخدام متجهات العقدة المفتوحة المنتظمة لحساب منحنيات $B-spline$. حيث متجه العقدة المستخدم يساوي $(k + m)$ ، حيث k هو درجة الاساس و m هو عدد النقاط الداخلية في متجه العقدة. عندما لا تكون هناك نقاط عقدة داخلية ، يكون منحنى $B-spline$ هو منحنى $Bezier$.

❖ دالة الاساس (Basis Functions)

بعد ان تم تحديد متجه العقد . يمكننا بعد ذلك حساب دالة اساس b -Spline حيث يمكن حساب وفق الصيغة الاتية :

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

بعد ان تم حساب ال B -Spline من الدرجة واحد يمكننا حساب اي درجة لهذه الدالة وفق هذه الصيغة .

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

حيث تشير k الى درجة الاساس , وقيم t_i هي العناصر لمتجه العقد المحدد مسبقا .

❖ نقاط التحكم (Control Points)

تتمثل الخطوة الاخيرة في تحديد نقاط التحكم من اجل الحصول على تقدير تقريبي دقيق الى المنحنى المطلوب. يمكننا حساب نقاط التحكم وفق صيغة (Greville Abcissa) التالية :

$$x_i = \frac{1}{n} (t_i + t_{i+1} + \dots + t_{i+n_1}) \dots \dots \dots i = 0, 1, \dots, m - k \dots \dots (13)$$

تشير n الى درجة الاساس او $(k-1)$ و m الى العدد الكلي للعقد في متجه العقد.

4-طريقة (M-estimate) [17]

تعتبر هذه الطريقة واحدة من الاكثر الطرائق استخداما وذلك بسبب كفاءة المقدرات المستحصل عليها مع طريقة المربعات الصغرى . ان هذه الطريقة احدى الطرائق الحصينة لوجود القيم الشواذ (الملوثة) في متغير الاستجابة (المتغير المعتمد)، تتميز هذه الطريقة بانها امتداد لمقدرات الامكان الاعظم ، وان الحرف M يشير الى الامكان الاعظم . وان مقدر M يمكن كتابته كالآتي :

1-4 خوارزمية M-estimate لتقدير (Regression B-Spline) [17]

وتتلخص هذه الطريقة بالخطوات الاتية :

1- يتم تحديد قيمة اولية للدالة $\hat{m}_i$ بالإضافة الى تحديد قيمة التكرارات ويمكن اعتماد تقدير انحدار الشريحة من خلال المعادلة

$$\hat{m} = x(X^T X)^{-1} X^T y \dots \dots \dots (11)$$

2- ايجاد البواقي (residuals) من خلال الصيغة الاتية

$$r_i = y_i - \hat{m}_i \dots \dots \dots (12).$$

3- تقدير الانحراف المعياري $\hat{\sigma}_e$ وحسب الصيغة الاتية

$$MADN = \frac{MAD}{Z_{0.75}} = 1.4836MAD \dots \dots (13)$$

حيث ان MAD يمكن ايجاده من خلال الصيغة الاتية

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

4- حساب قيمة s من المعادلة الاتية

$$s = \frac{y_i - \hat{m}_i}{\hat{\sigma}_e} \dots \dots \dots (14)$$

5- ايجاد القيم الزائفة (pseudo data) حسب الصيغة الاتية

$$z = \hat{m}_i + \frac{\Psi_c(s)}{2} \dots \dots (15)$$

حيث ان $c=1.345$ ، $\Psi_c(s) = \max[-c, \min(c, s)]$
6- حساب تقدير جديد للدالة $\hat{m}$ بالاعتماد على القيم الزائفة وكالاتي :

$$\dots \dots (16). \hat{m}_p = x(x^T X)^{-1} X^T Z$$

7- جعل $\hat{m}_p = \hat{m}$ ، ونقوم بتكرار الخطوات من الخطوة (2) الى (6) ونتوقف حتى نحصل على تقارب في التقدير

2-4 خوارزمية M-estimate لتقدير (Penalized B-Spline) [17]

تتلخص هذه الطريقة بالخطوات الاتية :

1- يتم تحديد قيمة اولية للدالة $\hat{m}_i$ بالإضافة الى تحديد قيمة التكرارات ويمكن اعتماد تقدير الشريحة الجزئية من خلال المعادلة

$$\dots \dots (17). \hat{m} = x(X^T X + \lambda D)^{-1} X^T y$$

2- ايجاد البواقي (residuals) من خلال الصيغة الاتية

$$\dots\dots\dots(18)r_i = y_i - \hat{g}_i$$

3- تقدير الانحراف المعياري $\hat{\sigma}_e$ وحسب الصيغة الاتية

$$MADN = \frac{MAD}{Z_{0.75}} = 1.4836MAD \dots\dots\dots(19)$$

حيث ان MAD يمكن ايجادة من خلال الصيغة الاتية

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

4- حساب قيمة S من المعادلة الاتية

$$\dots\dots\dots(20)s = \frac{y_i - \hat{m}_i}{\hat{\sigma}_e}$$

5- ايجاد القيم الزائفة (*pseudo data*) حسب الصيغة الاتية

$$\dots\dots\dots(21)z = \hat{m}_i + \frac{\psi_{c(s)}}{2}$$

حيث ان $c=1.345$ ، $\psi_c(s) = \max[-c, \min(c, s)]$
6- حساب تقدير جديد للدالة $\hat{m}$ بالاعتماد على القيم الزائفة وكالاتي :

$$\dots\dots\dots(22)\hat{m}_p = x(X^T X + \lambda D)^{-1} X^T y$$

7- جعل $\hat{m}_p = \hat{m}$ ، ونقوم بتكرار الخطوات من الخطوة (2) الى (6) ونتوقف حتى نحصل على تقارب في التقدير

5-الجانب التجريبي[8]

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 دوال لتمثيل النموذج الاصلي وهي كالاتي:-

1-دالة لاقطية[2]

$$f(x) = x + 2\exp(-16x^2)$$

2- دالة من الدرجة الثانية [15]

$$f(x) = 8(x - 0.5)^2$$

3- دالة من الدرجة لاقطية[11]

$$f(x) = \sin(2x) + 2 \exp(-2x^2)$$

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

جدول (1) يبين المعدل لقيم معيار المقارنة (MAE) ولجميع حجوم العينات ومستويات التلويث لمقارنة طريقة B- Spline في انحدار الشريحة (SB_{RS}) و B-Spline لانحدار الشريحة الجزائية (SB_{PS}) بالنسبة لطريق التحسين M للأنموذج الاول

Sample size حجوم العينات	طرائق التقدير	مستويات التلويت الثلاث		
		10%	16%	20%
$n_1=50$	$B_{RS}S$	0.859185845375766	0.842545086970516	0.835102747705256
	$B_{PS}S$	0.302688952249448	0.286476156221006	0.284871454859448
$n_2=100$	$B_{RS}S$	0.772527284688903	0.753273310455836	0.74269244720281
	$B_{PS}S$	0.435502461638852	0.419882054805362	0.411384431676352
$n_3=150$	$B_{RS}S$	0.727996290706994	0.711515683255502	0.698247512499867
	$B_{PS}S$	0.425407711125869	0.413096579712921	0.403354036795963

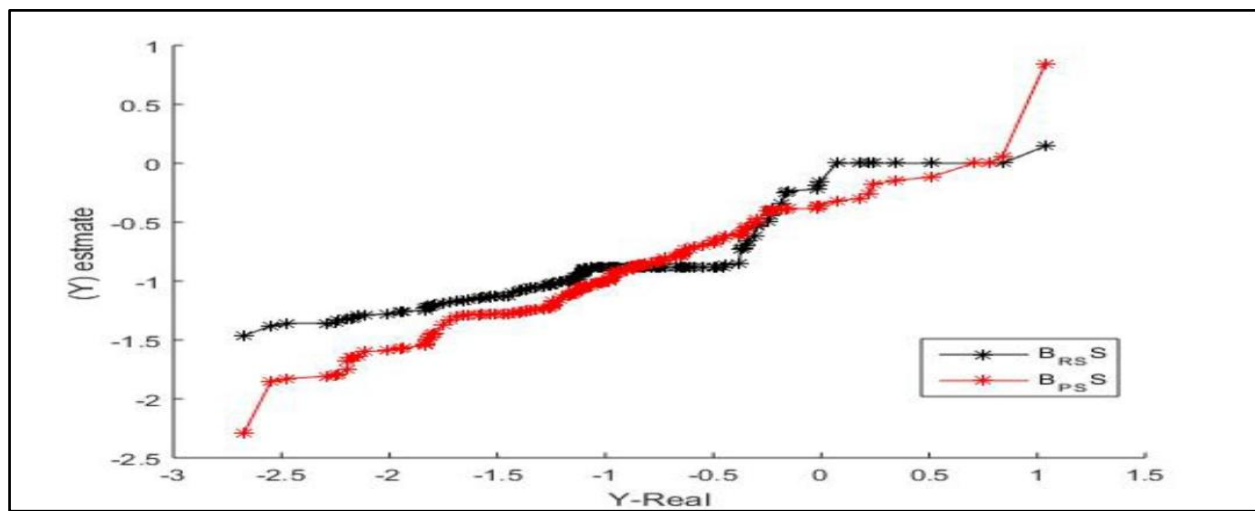
جدول (2) يبين المعدل لقيم معيار المقارنة (MAE) ولجميع حجوم العينات ومستويات التلويت لمقارنة طريقة B- Spline في انحدار الشريحة (SB_{RS}) و B- Spline لانحدار الشريحة الجزائية (SB_{PS}) بالنسبة لطريق التحصين M للأنموذج الثاني

Sample size حجوم العينات	طرائق التقدير	مستويات التلويت الثلاث		
		10%	16%	20%
$n_1=50$	$B_{RS}S$	0.847681517356411	0.838768348466307	0.827979888429032
	$B_{PS}S$	0.30786908446748	0.302487960805792	0.292180517516567
$n_2=100$	$B_{RS}S$	0.772287325507211	0.753340502016683	0.749001249649594
	$B_{PS}S$	0.438801672757033	0.423901115070321	0.418244068041991
$n_3=150$	$B_{RS}S$	0.734709607268037	0.714465369664029	0.700686147483468
	$B_{PS}S$	0.431209327832542	0.416545504131353	0.406585459190554

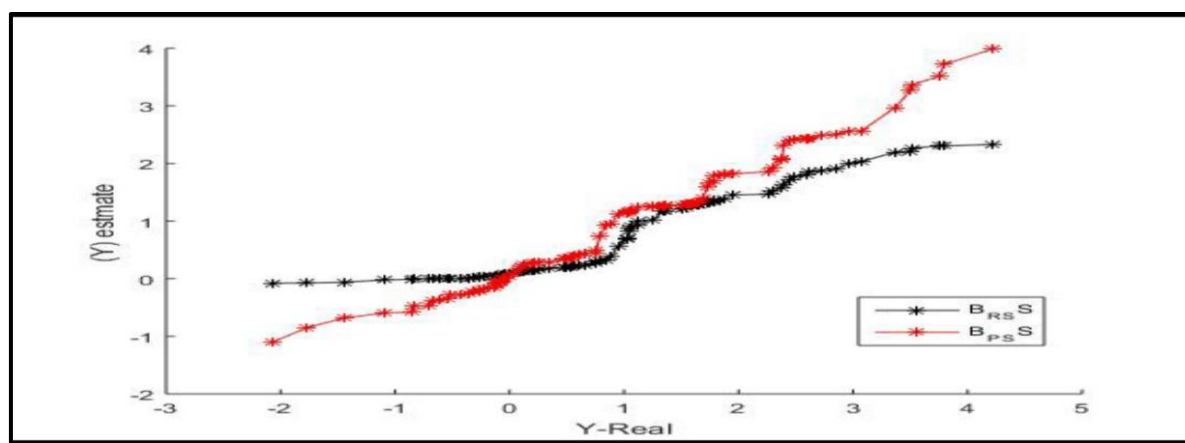
جدول (3) يبين المعدل لقيم معيار المقارنة (MAE) ولجميع حجوم العينات ومستويات التلويت لمقارنة طريقة B- Spline في انحدار الشريحة (SB_{RS}) و B- Spline لانحدار الشريحة الجزائية (SB_{PS}) بالنسبة لطريق التحصين M للأنموذج الثالث

Sample size حجوم العينات	طرائق التقدير	مستويات التلويت الثلاث		
		10%	16%	20%
$n_1=50$	$B_{RS}S$	0.812047015154894	0.796206007522084	0.787338057892965
	$B_{PS}S$	0.262490067482313	0.262490067482313	0.261431471718643
$n_2=100$	$B_{RS}S$	0.725153321004964	0.71750701070267	0.700778786567903
	$B_{PS}S$	0.41999711574937	0.408460815726807	0.398964222494532
$n_3=150$	$B_{RS}S$	0.692936465560763	0.674811390196406	0.654913853427359
	$B_{PS}S$	0.418391290031671	0.404660524319443	0.392136163354314

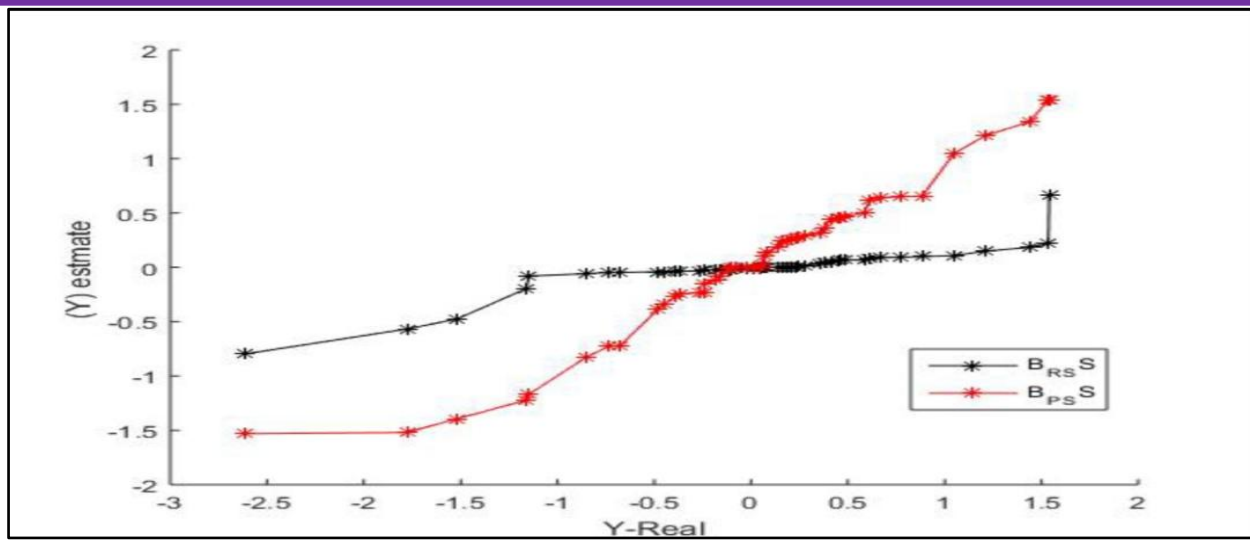
يمكن ملاحظة ان طريقة (PenalizeB-Spline) باستعمال طريقة التحصين (M-estimate) كانت الافضل في التقدير حيث سجلت اقل قيمة بالنسبة لمعيار المقارنة معدل لمتوسط الخطأ المطلق (MAE) ولجميع نماذج المحاكاة ومستويات التلوين التي تم استعمالها في هذا الجانب.



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



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



شكل رقم (3) يبين القيم الحقيقية والقيم التقديرية لمقدرات النموذج الثالث $(S B_{RS})(S B_{PS})$ عند حجم العينة $n=150$

الاستنتاجات

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

- 1- من خلال نتائج المحاكاة يمكن ملاحظة القيمة المعيارية للنماذج الثلاثة حيث انها بدأت تقل عند ازدياد نسب التلويث.
- 2- لوحظ من خلال النتائج للنماذج الثلاثة تبين ان طريقة التمهيد الافضل كانت تمهيد الشريحة الجزئية لحجوم العينات الثلاثة ولجميع مستويات التلويث.
- 3- من خلال نتائج المحاكاة يمكن ملاحظة انه كلما زاد حجم العينة ادى ذلك الى الحصول على نتائج اقل بالنسبة لمعيار المقارنة (MAE).

التوصيات

من خلال ما توصل اليه الباحث من استنتاجات يمكن عرض بعض التوصيات وهي كما يلي :

- 1- يمكن الاعتماد على طريقة التقدير (Penalized B-Spline) وذلك لان كون ادائها متفوقا عن باقي طرائق التقدير الاخرى.
- 2- يمكن استعمال طرائق تحسين اخرى (MM-estimate, L- estimate) في حاله هناك قيم ملوثة في البيانات (المتغير التوضيحي).
- 3- استعمال معيار متوسط الخطا المطلق (MAE) لمقارنة طرائق التقدير المختلفة بدلا من المعايير الاخرى .

المصادر

- 1-خمو، خلود يوسف (2004) " مقارنة أساليب بيز مع طرائق أخرى لتقدير منحني 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]-Johnson, R.W., (2005). "Higher Order B-spline Collocation at the Greville Abscissa." Applied Numerical Mathematics: 52:63-75.
 - [11]- Lee, E. T. Y. (1982). "A Simplified B-Spline Computation Routine". Computing. 29 (4): 365–371
 - [12]-Magoon, Jason,(2010). "Application of the B-spline Collocation Method Geometrically Non-Linear Beam Problem" .Statistical Science 2(1): 9-25.
 - [13]-Marsh, L. C., Cormier, D. R. (2002) "Spline regression models". Sage Publications, Inc. Londo
 - [14]- Wu, H. and Zhang, J., (2006), "Nonparametric regression methods for longitudinal data analysis: Mixed-Effects modeling approaches", John Wiley & Sons, New Jersey
 - [15]-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
 - [16]-Wahba, G. (1990). "Spline Models for Observational Data". SIAM Philadelphia. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59.
 - [17]-Wang, B. &Miao, Z.(2014). "Comparative Analysis for Robust Penalized Spline Smoothing Method" Volume 2014, 11 p