المجلة العراقية للعلوم الإحصائية (24) 2013
ص ص[17-34]

حول مواقع العقد لمقدر شريحة الانحدار

أ.د. محمد حبيب الشاروط*
حافظ محمد مطر**

الملخص

تعد شرائح الإنحدار Regression splines أحد الأساليب المستخدمة في تقدير منحنى الإنحدار اللامعلمي ومن أهم العناصر اللازمة لتطبيق هذا الأسلوب هو تحديد درجة الشريحة وعدد العقد (knots) المستخدمة في النموذج ومواقع هذه العقد . ومازالت مسألة اختيار عدد العقد ومواقعها هي المعضلة الرئيسية في تقدير منحنى الانحدار اللامعلمي باستخدام شرائح الانحدار ومتى ما تم إختيارها بعناية فإن كمية التمهيد في المنحنى المقدر سوف تكون في حالتها المثلى . جاء هذا البحث ليسلط الضوء على أسلوبين من أساليب نشر العقد في شرائح الإنحدار : الأسلوب الأول يتضمن نشر العقد بحيث تمثل مجزءات للبيانات واستخدام أحد معايير الإختيار بين النماذج مثل GCV (Generalized Cross Validation) لإختيار عدد هذه العقد ، أما الأسلوب الآخر فيعتمد على نشر العقد بمسافات متساوية واستخدام معيار GCV لاختيار المواقع النهائية للعقد وعددها . وتمت المقارنة بين الأسلوبين باستخدام أحد معايير الأخطاء وهو $MAAE$،وبالاعتماد على بيانات تجريبية باستخدام أسلوب المحاكاة توصلنا إلى أنَ تجارب المحاكاة بأن أسلوب نشر العقد بمسافات متساوية كان أفضل من أسلوب نشر العقد على شكل مجزءات للبيانات.

* استاذ مساعد/ قسم الإحصاء والمعلوماتية / كلية علوم الحاسوب والرياضيات/ جامعة القادسية.

** باحث / المديرية العامة للتربية / محافظة القادسية.

تاريخ استلام البحث 2013/12/6 ——————— تأريخ القبول 2013/5/5

On Knots locations for Regression Spline Estimator

Abstract

Regression splines is one of the methods that are used to estimate the regression curve non parametrically. One of the most important elements that contribute to the application of the method is to determine the degree of the spline function and the number of knots and their locations Choosing the number of knots and their locations is the main problem in estimating the non parametric regression using regression splines and, when carefully selected, the amount of smoothing in the fitted curve will be in optimal conditions. The research is to shed light on two methods of knots locations; the first method includes place of the knots which represents the data quintile and use one of models selection criteria such as GCV (Generalized Cross Validation) criterion to select the number of these knots. The other method depends on placing the knots on equal spaces and the use of GCV criterion to select the final locations of knots and its number , A comparison is made between the two methods using one of the errors criteria which is $MAAE$ (Mean Average Absolute Error),depending on the experimental data using the simulation method . Through simulation experiments the method of place knots in equal spaces , is better than the method of placing knots in the form of quintile of data .

1- المقدمة

إذا افترضنا نموذج الانحدار بالشكل الآتي:

$$y_i = g(x_i) + \varepsilon_i \quad , i = 1,...,n \qquad .... \qquad (1)$$

حيث ε_i هي الأخطاء العشوائية ونفرض بأن لها توزيعاً توقعه صفر وتباينه ثابت (s^2) ، g هي دالة الانحدار وهي دالة غير معلومة ،وإن إحدى الأدوات المستخدمة في إيجاد تقريب للدالة g هي دالة الشريحة Spline Function والتي تعرف كالآتي :

تعريف (1) [8] دالة الشريحة Spline Function S من الدرجة r مع متتابعة مواقع العقد $\{t_i\}_{i=1}^{k}$ هي دالة حقيقية $S : R \to R$ وتحقق الشروط الآتية :

1- الدالة S هي متعددة حدود من الدرجة r في كل فترة جزئية $[t_i , t_{i+1}]$.

2- الدوال $S^{(r-1)},...,S'',S',S$ مستمرة عند العقد t_i .

وتظهر الشرائح في صيغتين [9] الأولى يطلق عليها TPF-Spline وتكتب الشريحة كالآتي:

$$S(x)=\sum_{j=0}^{r}\beta_j x^j+\sum_{j=1}^{k}\beta_{r+j}(x-t_j)^r \text{...}(2)$$

حيث الصيغة $(x-t_j)_+^r$ تسمى متعددة الحدود المبتورة ويرمز لها بالرمز Truncated TPF (Polynomial) Function وتسمى الشريحة السابقة بالشريحة TPF (TPF- Spline).

أما الصيغة الأخرى فيطلق عليها الشريحة-B (B-Spline) وهي أكثر ملاءمة من الناحية العددية من الشرائح –TPF ، ووضع[6] صيغتين لحساب الشرائح-B باستخدام عقد بمسافات غير متساوية [4] وأخرى باستخدام عقد بمسافات متساوية ويمكن توضيح الأخيرة كالآتي[2] :

تعرف الشريحة j من الدرجة صفر كالآتي:

$$B_j(x,0)=\begin{cases}1 \ , \ (j-1)dx\le x-x_{\min} \ jdx\\ 0 \ , \ 0w\end{cases} \text{.......................}(3)$$

حيث $B_j(x,0)$ تعني قيمة الشريحة من الدرجة صفر عند النقطة x ، $dx=(x_{\max}-x_{\min})/n¢$ ، $n¢$ عدد الفترات بين العقد . وبعد حساب الشريحة من الدرجة صفر تحسب الشرائح من الدرجات الأعلى وفق الصيغة الآتية:

$$B_j(x,r)=\frac{r+p-j+1}{r}B_{j-1}(x,r-1)+\frac{j-p}{r}B_j(x,r-1) \ \ ... \qquad (4)$$

حيث $p=x-x_{\min}$ ، $B_0(x,r)=0 \ , "x,"r$.

إن استخدام دالة الشريحة في تقدير منحنى الانحدار يعتبر أحدى الأدوات المفيدة خصوصا في حالة البيانات غير الخطية حيث يتم تقدير متعددات حدود من درجة معينة ليس بصورة كلية (Globally) ولكن بصورة مستقلة في كل فترة جزئية بين عقدتين، وظهرت عدة نماذج تستخدم دالة الشريحة في الإنحدار اللامعلمي (Non Parametric Regression) منها شرائح الانحدار Regression Splines [6],[3],[9] ، وفيها نحتاج إلى تحديد عدد العقد وأماكن انتشارها ، أما النماذج الأخرى فتستخدم حد الجزاء غير الممهد (Roughness Penalty) للتحكم بكمية التمهيد بدلا

من الاعتماد على أسلوب اختيار عدد العقد وأماكن انتشارها ويطلق عليها الشرائح الجزائية Penalized Splines ومن أنواعها الشريحة الممهدة Smoothing Spline [8] والشريحة الجزائية P-Spline [7] ,[2]، وقد تناول البحث دراسة شرائح الانحدار Regression Splines واستخدمنا في هذه الدراسة الشرائح –B التكعيبية Cubic B-Spline لما لها من خصائص الاستمرارية الجيدة حيث تمتلك مشتقة ثانية مستمرة عند العقد ، وتمت المقارنة بين أسلوبين لنشر العقد الأول يتم فيه وضع العقد باعتبارها مجزءات للبيانات ويتم استخدام معيار GCV لتحديد عدد العقد في النموذج ،أما الأسلوب الآخر فيعتمد على نشر العقد بمسافات متساوية ويتم استخدام معيار GCV لاختيار المتتابعة النهائية من العقد.

2 هدف البحث

يهدف هذا البحث إلى المقارنة بين أسلوبين من أساليب نشر العقد في شرائح الإنحدار ، الأسلوب الأول والذي يتضمن نشر العقد بحيث تمثل مجزءات للبيانات (Quintile Data Knots) ويتم فيه اختيار عدد العقد باستخدام معيار GCV ، وأما الأسلوب الثاني فيعتمد على نشر العقد بمسافات متساوية واستخدام معيار GCV لاختيار المتتابعة النهائية من العقد .

3 شريحة الانحدار التكعيبية Cubic Regression Spline

إذا افترضنا أن دالة الانحدار g في نموذج الانحدار (1) هي الشريحة-B التكعيبية[1] فإن الافتراض يكون كالآتي : $g(x)=\sum_{j=1}^{k+4} a_j B_J(x,3)$ وبصيغة المصفوفات $g = Ba$ حيث $g = (g(x_1), \dots, g(x_n))^T$ ، B هي مصفوفة من الرتبة $n \times (k+4)$ بمدخلات كالآتي : $B = (B_{ij})$ وأن $B_{ij} = B_j(x_i, 3)$ ، $a = (a_1, \dots, a_{k+4})$ متجه معاملات الشرائح .

ولتقدير متجه المعاملات a نستخدم المربعات الصغرى الاعتيادية بتصغير المعيار الآتية :

$$RSS = \sum_{i=1}^{n}\left[y - \sum_{j=1}^{k+4} a_j B_J(x)\right]^2 \qquad \dots\dots\dots\dots\dots\dots\dots\dots(5)$$

وبصيغة المصفوفات

$$RSS = (Y - Ba)^T (Y - Ba) \qquad \dots\dots \qquad (6)$$

حيث $Y = (y_1, \dots, y_n)^T$

وبإيجاد التفاضل الجزئي بالنسبة إلى a ومساواة الناتج للصفر نحصل على متجه المعاملات المقدرة

$$\hat{a} = (B^T B)^{-1} B^T Y \qquad \ldots\ldots \qquad (7)$$

ومتجه القيم المقدرة

$$\hat{g} = B\hat{a} = B(B^T B)^{-1} B^T Y \qquad \ldots\ldots \qquad (8)$$

4 اختيار عدد العقد ومواقعها :

إن كمية التمهيد في المنحنى المقدر عند استخدام شرائح الانحدار Regression Splines تعتمد على عدد العقد المستخدمة في النموذج ومواقعها، وسوف نقارن في هذه الدراسة بين أسلوبين لنشر العقد ، الأول يتم فيه نشر العقد باعتبارها مجزءات للبيانات وسوف نرمز له بالرمز Q.D.K (Quintile Data Knots) ويكون كالآتي :

i. إذا تم اختيار عقدة واحدة فإنها توضع بحيث **50%** من البيانات تقع قبلها و **50%** تقع بعدها (الوسيط) .

ii. إذا تم اختيار عقدتين فإنهما توضعان بحيث تقسمان البيانات إل ثلاثة أقسام متساوية .

iii. إذا تم اختيار ثلاث عقد فإنها توضع بحيث تمثل ربيعات أي تقسم البيانات إلى أربعة أجزاء متساوية

وهكذا....

أما عدد العقد فيتم اختياره بحيث نحصل على أقل قيمة للمعيار GCV Generalized Cross Validation [3] وصيغته كالآتي:

$$GCV = \frac{n^{-1}\sum_{i=1}^{n}\left[y_i - \hat{g}(x_i)\right]^2}{(1-df/n)^2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(9)$$

حيث $df = k + 4$ ، n عدد البيانات ، k عدد العقد في النموذج .

أما الأسلوب الآخر لنشر العقد فيتضمن نشر العقد بمسافات متساوية وسوف نرمز له بالرمز Equally Spaced Knots (E.S.K) واستخدام معيار GCV لاختيار المتتابعة النهائية من العقد

وقد وضعنا خوارزمية لاختيار عدد العقد ومواقعها بعد نشرها بمسافات متساوية وتتضمن أربعة خطوات كالآتي :

خوارزمية اختيار متتابعة العقد[1] :

1 – في الخطوة الأولى يتم نشر العقد بمسافات متساوية واستخدام المعيار GCV لاختيار عدد هذه العقد .

2- في هذه الخطوة يتم تحريك مواقع العقد في الاتجاه الأيمن من المحور الأفقي بمقدار ضئيل على أن لا تتجاوز المتتابعة حدود الفترة ويتم الأمر من خلال التحكم بعدد التكرارات وقيمة المقدار المضاف ، وفي كل تكرار يتم حساب قيمة المعيار GCV ونختار متتابعة العقد التي تعطي أقل قيمة له.

3- في الخطوة الثالثة يتم نفس الإجراء المتبع في الخطوة الثانية ولكن تكون حركة العقد في الاتجاه المعاكس ويتم اختيار متتابعة العقد التي تعطي أقل قيمة للمعيار GCV.

4- في الخطوة الرابعة يتم اختيار المتتابعة النهائية من العقد من بين المتتابعتين في الخطوتين الثانية والثالثة والتي تعطي أقل قيمة للمعيار GCV.

5 تجارب المحاكاة :

لغرض إجراء المقارنة بين أساليب نشر العقد في شرائح الانحدار التي تم توضيحها سابقاً ، لابد من وضع بعض الافتراضات المهمة للحصول على تحليل أكثر شمولية من خلال استخدام عدد كبير من العينات وبأحجام مختلفة أواختيار قيم مختلفة لتباين الأخطاء ونظرا لصعوبة تطبيق هذه الافتراضات والحصول على هذه العينات في الواقع العملي يتم استخدام الأسلوب التجريبي من خلال تطبيق أسلوب المحاكاة (Simulation)

المحاكاة Simulation

تعرف المحاكاة بأنها تقليد للواقع العملي بحيث تقوم بتوظيف نماذج تظهر فيها عدد كبير من الحالات الافتراضية لتكون نتائج التحليل أكثر شمولية ، وقد ظهرت الاستعانة بالمحاكاة أساسا بوصفها احد افرازات التقدم الحاصل في مجال الحاسبات الالكترونية من جانب ، ومن جانب أخر بسبب التطور

الحاصل في مجال التحليل العددي.أما مسوغات العمل بالمحاكاة فتكون في الغالب للتأكد من تحقق جانب تطبيقي موجود أصلا أو لصعوبة الحصول على بيانات توفر معلومات دقيقة عن ظاهرة معينة أو عندما يصعب إثبات البرهان الرياضي بشكله النظري لبيان أفضلية طرائق تقدير معينة على حساب أخرى.تم توليد المتغيرات التوضيحية X_i والتي تتبع التوزيع المنتظم القياسي باستخدام برنامج R وبطريقة (Mersenne – Twister) [5]، وباستخدام طريقة (Box – Muller) يتم تحويل هذه المتغيرات إلى متغيرات تتبع التوزيع الطبيعي القياسي ومن ثم باستخدام التحويل $\varepsilon=\sigma Z$ -حيث Z متغير يتبع التوزيع الطبيعي القياسي- يمكن الحصول على متغيرات عشوائية تتبع التوزيع الطبيعي الذي توقعه صفر وتباينه ثابت ويساوي σ^2، ولغرض استخدام أسلوب المحاكاة في بحثنا هذا تم استخدام عدة مستويات للإنحراف المعياري للأخطاء وهي :

1- $(s = 0.125)$ 2- $(s = 0.75)$ 3- $(s = 1)$ 4- $(s = 1.5)$ 5- $(s = 2)$

أما قيم المتغير المعتمد فيتم توليده باستخدام المعادلة الآتية :

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

، حيث x_i هي المتغيرات التوضيحية التي تم توليدها كما سبق ، و $g(x_i)$ قيمة دالة الإختبار عند النقطة x_i ، e_i هي الأخطاء العشوائية التي تم توليدها كما سبق.

وقد تم استخدام أربع قيم لحجوم العينات هي : $n = 20,50,150,250,500$.

وتم تكرار كل تجربة من تجارب المحاكاة بواقع (300) تكرار لكل تجربة وحساب معدل متوسط الخطأ المطلق $MAAE$ (Mean Average Absolute Error) والذي تكون صيغته كالآتي :

$$MAAE = \frac{\sum_{i=1}^{N} AAE_i}{N} \quad \dots . \quad (10)$$

حيث N عدد التكرارات وأن AAE (Average Absolute Error) وصيغته :

$$AAE = \frac{\sum_{i=1}^{n} \left| g(x_i) - \hat{g}(x_i) \right|}{n} \quad \dots . \quad (11)$$

دوال الاختبار المستخدمة في تجارب المحاكاة

تم اختيار ثلاث دوال مختلفة للاختبار في تجارب المحاكاة يمكن توضيحها كالآتي:

1- دالة الانحدار اللاخطية باستخدام دالة أسية ودالة مثلثية وصيغتها :

$$f(x) = \exp(-400(x+0.4)^2) + \frac{2\exp(-500(x-0.65)^2)}{\sin(x-2)}$$

2- دالة الانحدار اللاخطية باستخدام دالة مثلثية ودالة متعددة حدود وصيغتها :

$$f(x) = \sin(2\pi x^3) + 0.5(x-0.75)^2$$

3- دالة الانحدار غير المتجانسة مكانياً Spatially Heterogeneous Function وصيغتها :

$$f(x) = \sqrt{x(1-x)}\,\sin\left(\frac{2\pi(1+(1.5)^{(9-4j)/5})}{x+(1.5)^{(9-4j)/5}}\right)\ ,\ j=6$$

4- دالة الانحدار اللاخطية باستخدام دالة أسية وأخرى مثلثية وصيغتها :

$$f(x) = \sin(2x) + 0.75\exp(-16x^2)$$

تجارب المحاكاة

بعد توليد المتغيرات العشوائية تم تنفيذ تجارب المحاكاة بواقع (300) تكرار لكل تجربة وكانت النتائج كالآتي:

أولاً : نتائج تجارب المحاكاة عند استخدام دالة الإختبار الأولى :

1. نلاحظ من خلال قيم $MAAE$ في الجدول (1) أن أسلوب نشر العقد بمسافات متساوية (E.S.K) قد

حقق تقدماً على أسلوب نشر العقد باعتبارها مجزءات للبيانات في الحالات الآتية :

i. عندما $\sigma = 0.125$ ولجميع حجوم العينات .

ii. عندما $\sigma = 0.75$ ولجميع حجوم العينات.

iii. عندما $\sigma = 1$ ولجميع حجوم العينات عدا حجم العينة $n = 150$.

iv. عندما $\sigma = 1.5$ ولجميع حجوم العينات عدا حجم العينة $n = 250$.

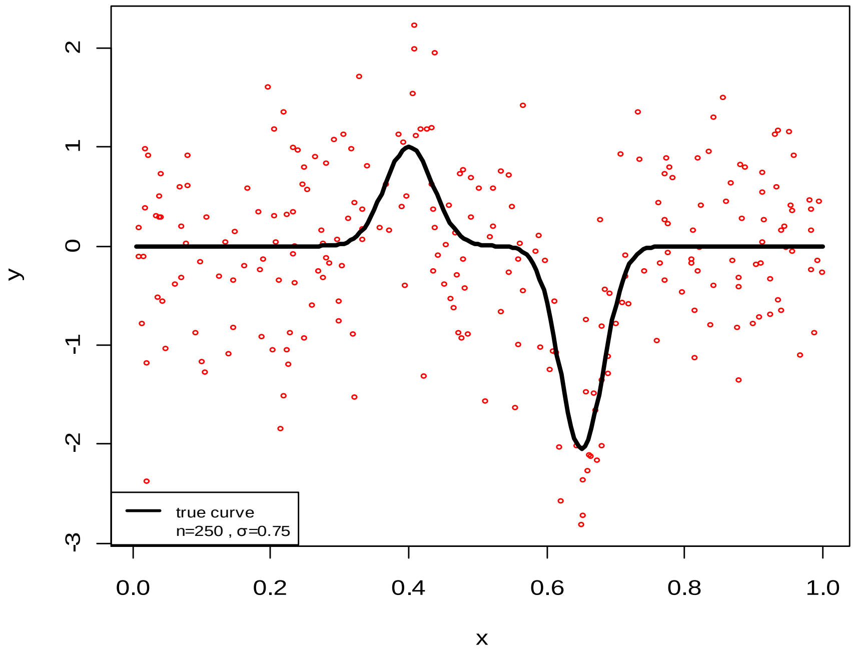
v. عندما $\sigma = 2$ ولجميع حجوم العينات عدا حجم العينة $n = 500$.

2- نلاحظ من خلال قيم الجدول (1) أن قيم $MAAE$ تقل كلما زاد حجم العينة عند ثبوت قيم الانحراف المعياري.

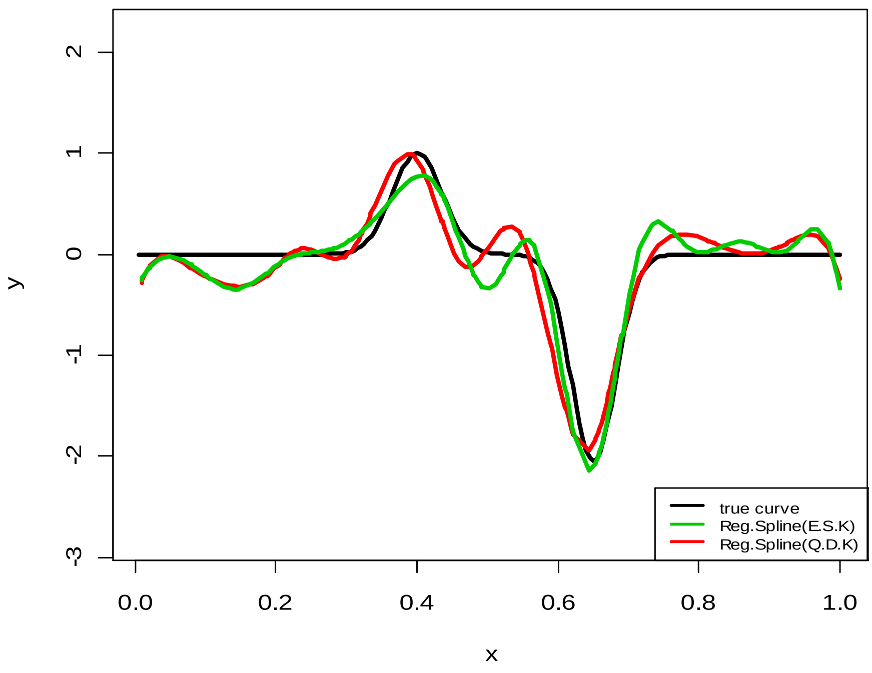
3- وأن قيم المعيار $MAAE$ تقل كلما قلت قيم الانحراف المعياري عند ثبوت حجم العينة .

الجدول(1) نتائج تجارب المحاكاة عند استخدام دالة الاختبار الأولى

Models	حجم العينة	$\sigma = 0.125$	$\sigma = 0.75$	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$
		MAAE	MAAE	MAAE	MAAE	MAAE
Regression Spline with Q.D.K	20	0.09625758	0.5335138	0.7098396	1.048778	1.421016
	50	0.09788548	0.3457806	0.4531055	0.6643574	0.8810042
	150	0.09831009	0.2163562	0.2786648	0.3954826	0.5168291
	250	0.09594464	0.1810787	0.230804	0.3178467	0.3996757
	500	0.1056698	0.1534023	0.1815139	0.2397104	0.3015433
Regression Spline with E.S.K	20	0.09193173	0.4711824	0.5884561	0.8310854	1.089463
	50	0.07184761	0.3428681	0.04162623	0.5517577	0.6862512
	150	0.06189328	0.215811	0.2882169	0.3920627	0.4739122
	250	0.06025209	0.1678671	0.2272049	0.3253438	0.3927666
	500	0.05805759	0.1263854	0.160673	0.236289	0.3116285



الشكل(1) المنحنى الحقيقي لدالة الإختبار الأولى على افتراض أن $n = 250$, $\sigma = 0.75$



الشكل (2) منحنى مقدر شريحة الإنحدار لدالة الإختبار الأولى عند استخدام

أسلوب Q.D.K (المنحنى الأحمر) ، وعند استخدام أسلوب E.S.K (المنحنى الأخضر) على افتراض أن

$n = 250$, $\sigma = 0.75$.

ثانياً : نتائج تجارب المحاكاة عند استخدام دالة الإختبار الثانية :

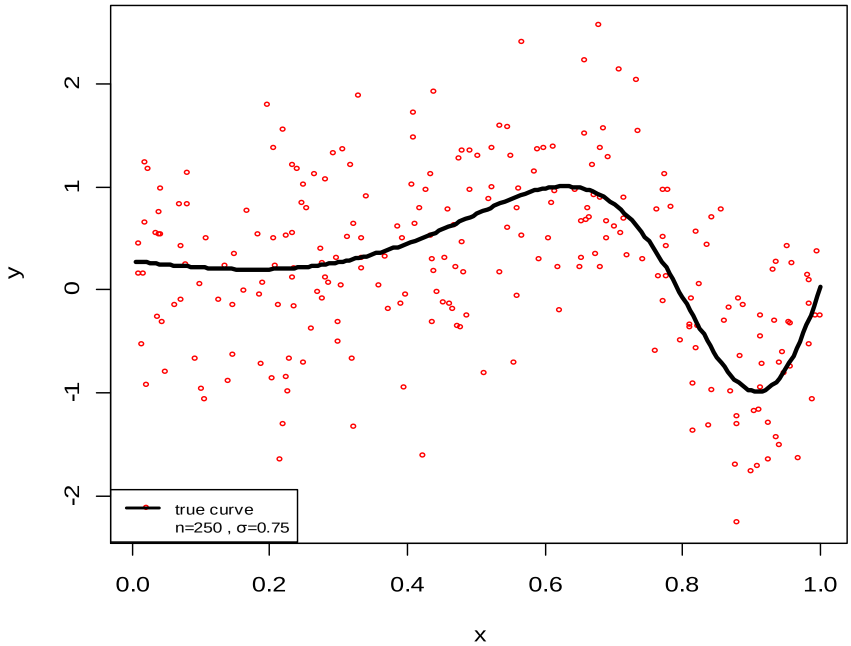
1- نلاحظ من خلال قيم $MAAE$ في الجدول (2) أن أسلوب نشر العقد بمسافات متساوية (E.S.K) قد حقق تقدماً على أسلوب نشر العقد باعتبارها مجزءات للبيانات في كل قيمة من قيم الانحراف المعياري ولكل حجم عينة .

2- نلاحظ من خلال قيم الجدول (2) أن قيم $MAAE$ تقل كلما زاد حجم العينة عند ثبوت قيم الانحراف المعياري.

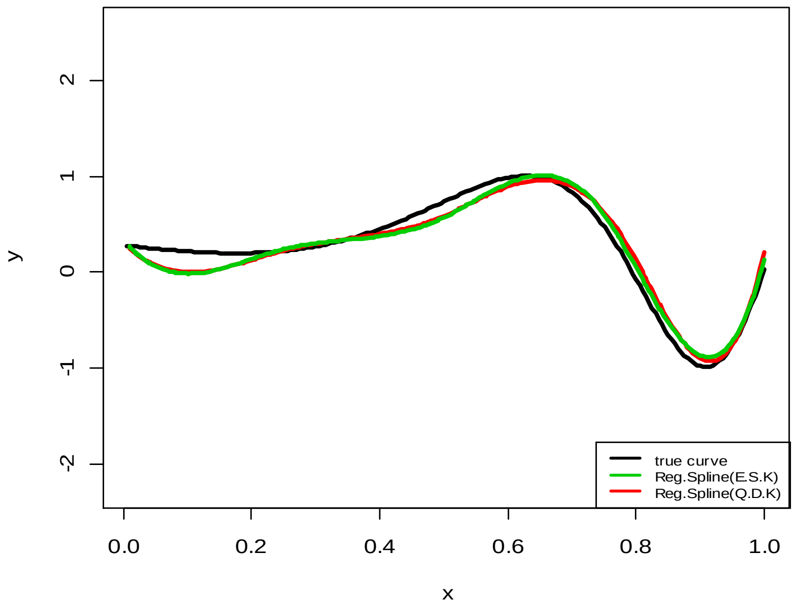
3- وأن قيم المعيار $MAAE$ تقل كلما قلت قيم الانحراف المعياري عند ثبوت حجم العينة .

الجدول(2) نتائج تجارب المحاكاة عند استخدام دالة الإختبار الثانية.

Models	حجم العينة	$\sigma = 0.125$	$\sigma = 0.75$	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$
		MAAE	MAAE	MAAE	MAAE	MAAE
Regression Spline with Q.D.K	20	0.08711958	0.5316548	0.7082808	1.048569	1.420581
	50	0.05494526	0.3316553	0.4418268	0.657449	0.8750401
	150	0.03202458	0.187058	0.2551828	0.3791485	0.5030589
	250	0.02434222	0.1439107	0.1999671	0.2942375	0.3820619
	500	0.02089087	0.1057342	0.1386106	0.2048615	0.2749947
Regression Spline with E.S.K	20	0.06958483	0.4137322	0.5337017	0.7845859	1.045831
	50	0.0431271	0.2562451	0.32876	0.462122	0.6052375
	150	0.02591476	0.1386392	0.1966484	0.2811482	0.363011
	250	0.02089087	0.110388	0.152141	0.220482	0.2805366
	500	0.01575161	0.08205351	0.1078197	0.1545503	0.2097205



الشكل(3) المنحنى الحقيقي لدالة الإختبار الثانية على افتراض أن $n = 250$, $\sigma = 0.75$.



الشكل (4) منحنى مقدر شريحة الإنحدار لدالة الإختبار الثانية عند استخدام أسلوب Q.D.K (المنحنى الأحمر) ، وعند استخدام أسلوب E.S.K (المنحنى الأخضر) على افتراض أن $n = 250$, $\sigma = 0.75$.

ثالثاً : نتائج تجارب المحاكاة عند استخدام دالة الإختبار الثالثة :

1- نلاحظ من خلال قيم $MAAE$ في الجدول (3) أن أسلوب نشر العقد بمسافات متساوية (E.S.K) و أسلوب نشر العقد باعتبارها مجزءات للبيانات حققا نتائج متقاربة حيث حقق أسلوب E.S.K تفوقاً في الحالات الآتية :

i. عندما $\sigma = 0.125$ ولجميع حجوم العينات .

ii. عندما $\sigma = 0.75$ ولجميع حجوم العينات عدا حجم العينة $n = 250$.

iii. عندما $\sigma = 1$ ولجميع حجوم العينات عدا حجم العينة $n = 500$.

iv. عندما $\sigma = 1.5$ ولجميع حجوم العينات.

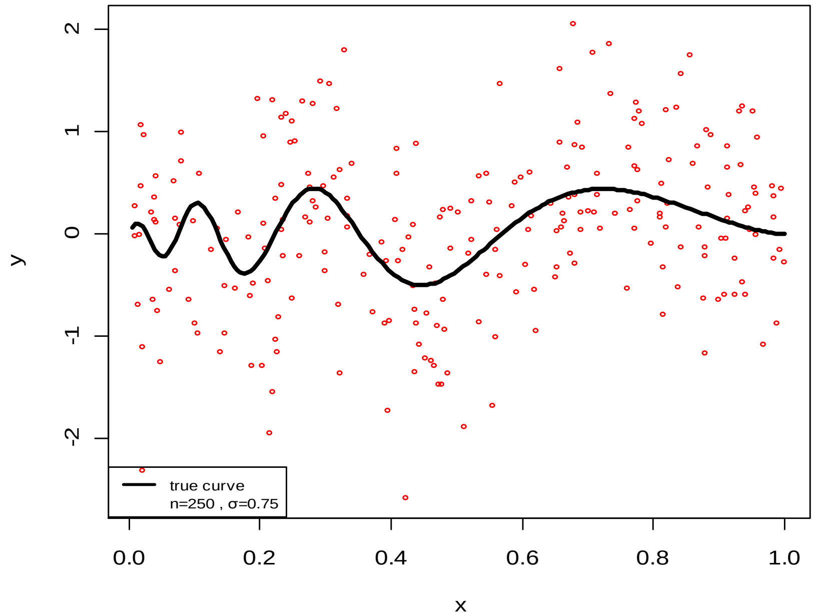
v. عندما $\sigma = 2$ ولجميع حجوم العينات .

2- نلاحظ من خلال قيم الجدول (3) أن قيم $MAAE$ تقل كلما زاد حجم العينة عند ثبوت قيم الانحراف المعياري.

3- وأن قيم المعيار $MAAE$ تقل كلما قلت قيم الانحراف المعياري عند ثبوت حجم العينة .

الجدول(3) نتائج تجارب المحاكاة عند استخدام دالة الإختبار الثالثة

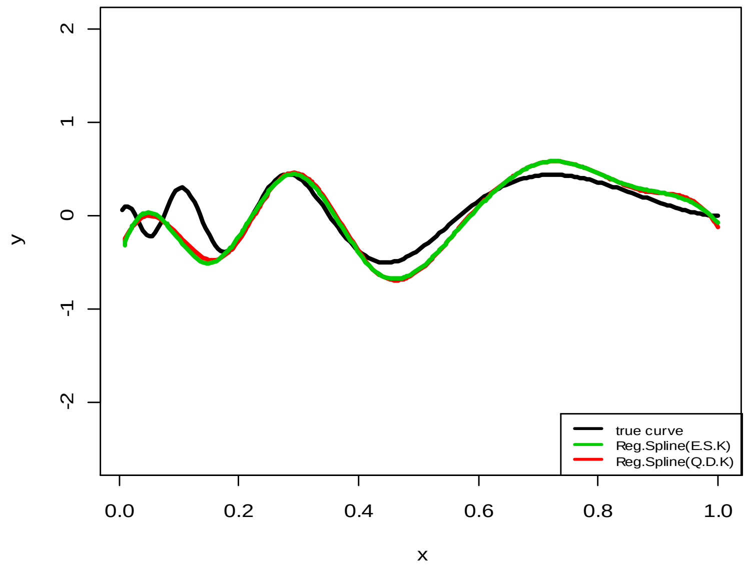
Models	حجم العينة	$\sigma = 0.125$	$\sigma = 0.75$	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$
		MAAE	MAAE	MAAE	MAAE	MAAE
Regression Spline with Q.D.K	20	0.08745416	0.5316844	0.7084242	1.048617	1.420495
	50	0.0620338	0.3329684	0.4431266	0.6582125	0.875231
	150	0.04658117	0.1912144	0.2587003	0.3814512	0.5045855
	250	0.0431232	0.1492789	0.2041639	0.2971904	0.384046
	500	0.04041738	0.1141975	0.1454006	0.2096818	0.278844
Regression Spline with E.S.K	20	0.0815601	0.4268211	0.5381524	0.7883528	1.049466
	50	0.05784057	0.2817187	0.3485646	0.4728326	0.605492
	150	0.03822959	0.1883691	0.2358746	0.3131288	0.3723204
	250	0.03194441	0.1524575	0.1982578	0.2631742	0.3105209
	500	0.02418778	0.1135558	0.1466881	0.2094643	0.2542414



الشكل(5) المنحنى الحقيقي لدالة الإختبار الثالثة على افتراض أن

$.\sigma = 0.75 \;\; , \;\; n = 250$

.



الشكل (6) منحنى مقدر شريحة الإنحدار لدالة الإختبار الثالثة عند استخدام

أسلوب Q.D.K (المنحنى الأحمر) ، وعند استخدام أسلوب E.S.K (المنحنى الأخضر)

على افتراض أن $\sigma = 0.75 \;\; , \;\; n = 250$

رابعاً : نتائج تجارب المحاكاة عند استخدام دالة الإختبار الرابعة :

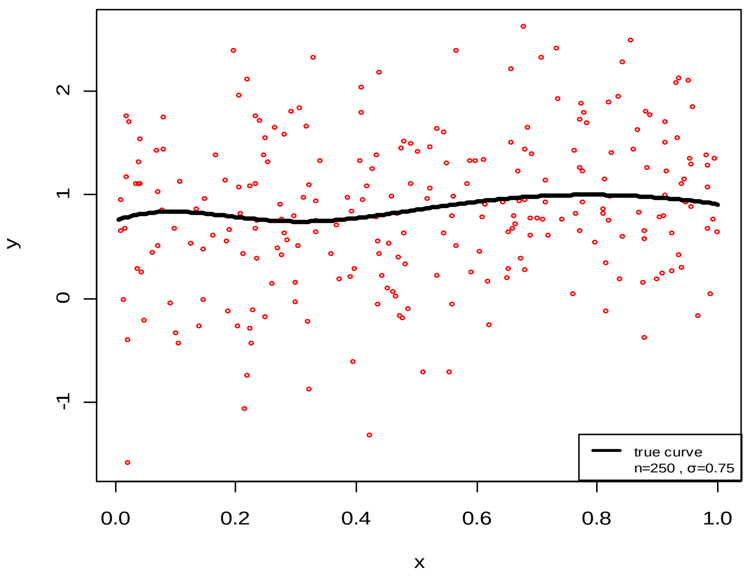
1- نلاحظ من خلال قيم $MAAE$ في الجدول (4) أن أسلوب نشر العقد بمسافات متساوية (E.S.K) قد حقق تقدماً على أسلوب نشر العقد باعتبارها مجزءات للبيانات في لكل قيمة من قيم الانحراف المعياري ولكل حجم عينة .

2- نلاحظ من خلال قيم الجدول (4) أن قيم $MAAE$ تقل كلما زاد حجم العينة عند ثبوت قيم الانحراف المعياري.

3- وأن قيم المعيار $MAAE$ تقل كلما قلت قيم الانحراف المعياري عند ثبوت حجم العينة .

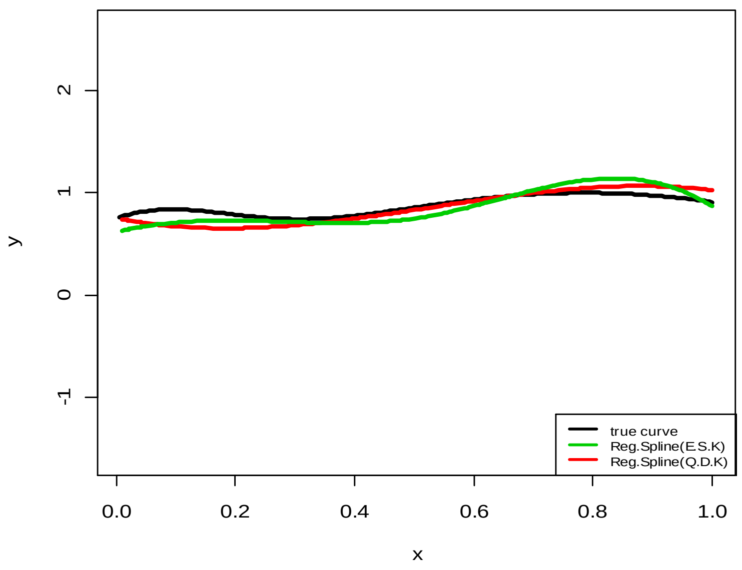
الجدول(4) نتائج تجارب المحاكاة عند استخدام دالة الإختبار الرابعة.

Models	حجم العينة	$\sigma = 0.125$	$\sigma = 0.75$	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$
		MAAE	MAAE	MAAE	MAAE	MAAE
Regression Spline with Q.D.K	20	0.08712704	0.5316501	0.7082813	1.048562	1.42058
	50	0.05491297	0.3316544	0.4418121	0.6574194	0.8750368
	150	0.03198354	0.1870482	0.2551656	0.3791379	0.5030642
	250	0.02432241	0.1439107	0.1999602	0.2942353	0.3820606
	500	0.01737735	0.1057294	0.13861	0.2048619	0.2750018
Regression Spline with E.S.K	20	0.0657062	0.4002905	0.5209285	0.7781133	1.039643
	50	0.03954483	0.2294867	0.3038812	0.4314743	0.5811009
	150	0.02343913	0.1208921	0.1754008	0.2523994	0.3334587
	250	0.01811317	0.09817588	0.1326719	0.1972625	0.2482182
	500	0.01285914	0.07331026	0.09554421	0.134996	0.1861153



الشكل(7) المنحنى الحقيقي لدالة الإختبار الرابعة على افتراض أن

$. \sigma = 0.75 \quad , \quad n = 250$



الشكل (8) منحنى مقدر شريحة الإنحدار لدالة الإختبار الرابعة عند استخدام

أسلوب Q.D.K (المنحنى الأحمر) ، وعند استخدام أسلوب E.S.K (المنحنى الأخضر)

على افتراض أن $\sigma = 0.75 \quad , \quad n = 250$

6- الاستنتاجات : -

نستنتج من خلال ما تم مناقشته من نتائج في تجارب المحاكاة الثلاث ما يأتي:

1- أن أسلوب نشر العقد بمسافات متساوية E.S.K أفضل من أسلوب نشر العقد بمسافات تمثل مجزءات للبيانات Q.D.K .

2. أن قيم المعيار $MAAE$ تقل كلما زاد حجم العينة وكذلك كلما قلت قيم الإنحراف المعياري للأخطاء .

3. أظهر المعيار GCV قدرة جيدة في الوصول إلى متتابعة العقد التي تعطي كمية جيدة من التمهيد في المنحنى المقدر .

7- التوصيات :

1- يوصي الباحثان باستخدام أسلوب نشر العقد بمسافات متساوية كونه أثبت كفاءة في تقدير منحنى الإنحدار اللامعلمي مقارنة بأسلوب نشر العقد باعتبارها مجزءات للبيانات .

2- استخدام معيار GCV في اختيار المتتابعة النهائية من العقد ومن ثم تحديد كمية التمهيد في المنحنى المقدر .

8- المصادر

1- مطير، حافظ محمد (2011) " مقارنة بعض أساليب تمهيد الإنحدار اللامعلمي باستخدام المحاكاة " رسالة ماجستير علوم في الرياضيات – كلية علوم الحاسوب والرياضيات –جامعة القادسية .

2. Eilers, P.H.C. and Marx, B.D. (1996)." Flexible smoothing using B-splines and penalizedlikelihood (with Comments and Rejoinder)". Statistical Science 11(2): 89-121

3. Eubank, R. L (1999)"Nonparametric Regression and Spline Smoothing" Marcel Dekker, NewYork, NY.
4. Hastie.T,Tibsherani. R and Friedman. J(2009) "The Element of Statistical Learning".Springer.ISBIN:978-387-84857-0.
5. Jones, O , Maillardet,R and Robinson,A (2009) "Introduction To Scientific Programming and Simulation Using R"CRC Press Taylor&Francis Group Boca Raton London New York.
6. Keele, L (2008)"Smoothing and Semiparametric Regression For The Social Sciences", Ohio State University, 2008 John Wiley and Sons.Ltd.
7. Ruppert, D and Carroll, R. J (1997) "Penalized Regression Splines" , http://www.citeulike.org/user/hyndman/article/2577705
8. Silverman, B. W. (1985). "Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion)". Journal of the Royal Statistical Society, Series B,47:1–52.
9. Wand, M. P. (2000), “A Comparison of Regression Spline Smoothing Procedures,” Computational Statistics, 15,443–462.