

مقارنة بعض طرائق تمهيد الانحدار اللامعلمي باستخدام المحاكاة

A Comparison of Non-Parametric Smoothing Regression

Models Using Simulation .

أ . م . د محمد حبيب الشاروط

قسم الإحصاء

كلية الإدارة والاقتصاد

جامعة القادسية

الديوانية - العراق

E-Mail:- malsharood@yahoo.com

حافظ محمد مطير

قسم الرياضيات

كلية علوم الحاسوب والرياضيات

جامعة القادسية

الديوانية - العراق

hafedmotair@yahoo.com

الخلاصة

تناول البحث دراسة مقارنة بين بعض طرائق تمهيد الانحدار اللامعلمية (Non Parametric Smoothing Regression Models) باستخدام المحاكاة، حيث تم اختيار أربعة أنواع منها وهي مقدر N-W ، مقدر الانحدار الخطي الموضعي Local Linear Estimator ، الشريحة الممهدة Smoothing Spline ، انحدار الشرائح الجزئية Penalized Spline Regression .

فالنوعين الأولين منها يتبعان طرائق تمهيد Kernel (Kernel Estimators) ، أما الآخرين يتبعان طرائق التمهيد الشرائحية (Spline Models) . يهدف البحث الى إيجاد أفضل مقدر من بين مقدرات التمهيد المذكورة لتمثيل البيانات المدروسة بناء على نتائج تجارب المحاكاة وقد تمكنا من خلال الجانب التجريبي التوصل الى إن الطرائق الشرائحية أفضل من طرائق Kernel في تمهيد منحنى الانحدار اللامعلمي.

1- المقدمة

إن نظرية وطرائق التمهيد اللامعلمية تطورت خلال العقدين الماضيين وأصبحت موضوعا لدراسة ضخمة لسببين أولهما إدراك الإحصائيين بأن التقدير المعلمي الصرف في تقدير منحنى الانحدار لا يتوافق مع الحاجة الى المرونة في تحليل البيانات والسبب الآخر يعود الى إن تطور أجهزة الحواسيب من الناحيتين المادية والبرمجية (Hardware&Software) مما قد مكن الباحثين من تطوير الجانب النظري لحساب المقدرات اللامعلمية وبذلك فقد أصبح من الممكن استخدام طرائق التمهيد في عدة اختصاصات منها علوم الحاسبات، الهندسة، البيولوجيا، الإحصاء القياسي والرياضيات وغيرها.

ولأنموذج الانحدار:

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

حيث الأخطاء العشوائية e_i تكون غير مرتبطة وتتوزع بوسط حسابي صفر وتباين ثابت s^2 تهدف طرائق الانحدار اللامعلمي الى تقدير الدالة غير المعلومة g حيث تفترض إن الدالة تنتمي الى فضاء غير منتهي البعد من الدوال المستمرة في المشتقة الأولى والثانية مع مربع مشتقة ثانية قابل للتكامل [8] ولقد صدرت في هذا المجال العديد من الدراسات والكتب تناول فيها الباحثين طرائق تمهيد الانحدار اللامعلمي منها طرائق Kernel حيث تناولتها كتب (Hurdel 1990) [12]، (Bowman, A.W & Azzalini 1997) [3]، (Looder 1999) [14] وبحوث (Cleveland 1979) [4]، (Fan 1992a) [9]، (Fan 1992b) [10] أما الطرائق الشرائحية فتناولتها كتب مثل (Wahba, G., 1990) [17]، (Eubank, R.L 1999) [8] (Ruppert & Wand & Carroll 2003) [15] وبحوث (Silverman 1985) [16]، (Eiler & Marx 1996) [6]، (Eiler & Marx 2004) [7].

تضمن البحث اربعة من طرائق التقدير للأنموذج في (1) منها مقدر NW، الانحدار الخطي الموضوعي، الشريحة الممهدة والشرائح الجزائية (B). وتمت المقارنة والمفاضلة فيما بينها اعتمادا على معايير البواقي MSE، MAE للوصول الى أفضل مقدر لتمثيل البيانات المولدة بأسلوب المحاكاة.

2- الجانب النظري

1-2 دالة الشريحة Spline function [15],[18],[2]

الدالة $S : R \otimes R$ تسمى دالة الشريحة من الدرجة r عند العقد (Knots) $\{x_i\}_{i=1}^k$ حيث $-\infty < x_0 < x_1 < \dots < x_k < x_{k+1} < \infty$ - إذا كانت :

1- s متعددة حدود من درجة لا تزيد عن r في الفترة $[x_i, x_{i+1}]$.

2- $s \in S^{(r-1)}, \dots, S \notin S$ هي دوال مستمرة عند العقد x_i . [15]

ومن أكثر الشرائح شيوعا الشرائح التكعيبية (Cubic Spline) بسبب خصائص الاستمرارية التي تمتلكها و تكتب الشريحة التكعيبية بالشكل:-

$$s(x) = b_0 + b_1x + b_2x^2 + b_3x^3 + \sum_{j=1}^k g_j (x - x_j)_+^3 \quad (2)$$

ويطلق عليها أيضا متعددة الحدود المبتورة (Truncated Polynomial Function) (TPF) [18]

حيث $(x - x_j)_+^3 = (x - x_j)^3$ عندما $x > x_j$ وتساوي صفر للقيم الأخرى.

أما الشريحة التكعيبية الطبيعية (Natural Cubic Spline) فتتحقق إضافة للشرطين السابقين الشرط التالي

3- الدالة s خطية في الفترتين (x_k, ∞) , $(-\infty, x_1)$. [2]

2-2 الشرائح من النوع (B) (B-Spline basis) : [5],[6],[7]

يتكون هذا النوع من الشرائح من قطع من متعددات حدود تصل بينها نقاط ربط تسمى عقد (Knots) ولقد وضع (de Boor) [5] ، صيغة تكرارية لحساب الشرائح من أي درجة مع عقد بمسافات غير متساوية بعد تعريف الشريحة من الدرجة صفر.

كما وضع صيغة تستخدم المسافات المتساوية بين العقد وكما يلي

إذا كانت $B_j(x; q)$ تمثل قيمة الشريحة B_j من الدرجة q عند النقطة x لمتتابعة من العقد بمسافات متساوية فإن الشريحة من الدرجة صفر تعرف كما يلي:

$$B_j(x; 0) = \begin{cases} (j-1)dx & \text{if } x - x_{\min} < jdx \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

حيث $dx = (x_{\max} - x_{\min}) / n$ وان n تمثل عدد الفترات بين العقد. وبعد حساب الشريحة من الدرجة صفر تحسب بقية الشرائح من أي درجة بالصيغة التكرارية التالية:

$$B_j(x;q) = \frac{q+p-j+1}{q} B_{j-1}(x;q-1) + \frac{j-p}{q} B_j(x;q-1) \quad (4)$$

$$p = x - x_{\min}, \quad B_0(x,q) = 0 \quad \text{حيث } x, q$$

كما توجد صيغة أخرى لحساب الشرائح (B) كفروقات (differences) للشرائح (TBF) وهي

$$B_j(x;q) = -1^{q+1} D^{q+1} f_j(x;q) / (h^p p!) \quad (5)$$

$$f_j(x;q) = (x - x_j)_+^q \quad \text{حيث}$$

إن الشرائح (B) أكثر استقراراً من الناحية العددية من الشرائح (TPF) ولها العديد من الخصائص التي تجعلها أكثر ملائمة منها:-

$$\begin{aligned} 1- \quad & \int_{t_j}^{t_{j+q}} B_j(x;q) dx = 1 \\ & \text{لأي درجة } q \\ 2- \quad & B_j(x;q) > 0 \quad \text{حيث } t_j < x < t_{j+q} \end{aligned}$$

3-2 انحدار الشريحة الممهدة [12],[8],[13],[1],[11] Smoothing Spline Regression

إن القياس الشائع لمدى مطابقة الدالة g لمجموعة من البيانات هو مجموع مربعات البواقي $\sum_{i=1}^n (y_i - g(x_i))^2$ ، ولكن استخدام هذا المعيار وحده فقط أمر غير مقبول من قبل الإحصائيين حيث ينتج تقدير متذبذب لاستكمال البيانات (Interpolate Data). لذلك يتم إضافة ما يسمى الجزء غير الممهدة (Roughness Penalty) إلى المعيار السابق لينتج معيار المربعات الصغرى لجزئية $PRSS$ وكلائي:

$$PRSS = \sum_{i=1}^n (y_i - g(x_i))^2 + \int_a^b (g^{(m)}(x))^2 dx \quad (6)$$

وبتصغير المعيار السابق أمام صف من الدوال g حيث $g \in W_2^m[a,b]$ في فضاء Sobolev فإن المقدر الناتج يسمى الشريحة الممهدة من الدرجة $2m - 1$ وكحالة خاصة عندما $m = 2$ فإن الشريحة الممهدة تسمى شريحة ممهدة تكعيبية Cubic Smoothing Spline وعلى فرض إن $a = 0, b = 1$ وموافقة الدالة g لشروط الحد الطبيعية Natural Boundary Conditions فإن المقدر الناتج يسمى شريحة تكعيبية طبيعية Natural Cubic Spline بعقد عند x_i .

وللحصول على عمليات حسابية $O(n)$ قام Reinsch بوضع صيغة تربيعية للجزء غير الممهدة كما في الصيغة الآتية:

$$\int_a^b (g^{(m)}(x))^2 dx = g^T K g \quad (7)$$

حيث K مصفوفة من الدرجة $n \times n$ وان $K = QR^{-1}Q^T$ حيث المصفوفة Q هي مصفوفة Tridiagonal من الدرجة $(n-2) \times (n-2)$ والمصفوفة R هي مصفوفة Tridiagonal متماثلة من الدرجة $(n-2) \times (n-2)$

اما I تسمى المعلمة التمهيدية Smoothing Parameter والتي تلعب دورا هاما في التوازن بين جودة التطابق (Goodness of Fit) المقاس بواسطة مجموع مربعات البواقي وبين كمية التمهيد المقاسة بواسطة الجزاء غير الممهيد

فعندما $I \in \mathbb{R}^+$ سوف يقل تأثير الجزاء غير الممهيد حتى يتلاشى ($I = 0$) وعندها لا توجد منطقة جزاء ونعود الى مجموع مربعات البواقي لينتج منحنى استكمال للبيانات

وعندما $I \in \mathbb{R}^+$ فان $\int_0^1 [g(x)]^2 dx \in \mathbb{R}^+$ ويكون المقدر عبارة عن خط مستقيم وهو تقدير المربعات الصغرى ($I = \infty$).

وبكتابة المعيار (6) بصيغة المصفوفات

$$PRSS = (y - g)^T (y - g) + I g^T K g \quad (8)$$

وبيجاد المشتقة ومساواة الناتج للصفر فان المقدر الناتج يكون بالشكل

$$\hat{g} = (I - I K)^{-1} y \quad (9)$$

$$\hat{g} = A_I y \quad (10)$$

الصيغة (9) تسمى صيغة Reinsch والمصفوفة A_I تسمى (Hat Matrix)

4-2 اختيار المعلمة التمهيدية [16],[11],[2]

لاختيار المعلمة التمهيدية نستخدم معيار العبور الشرعي المعمم GCV (Generalized Cross Validation)

$$GCV = n^{-1} \sum_{i=1}^n (y_i - \hat{g}(x_i; I))^2 / (1 - n^{-1} tr(A_I))^2 \quad (11)$$

حيث المقدار $tr(A_I)$ يساوي مجموع عناصر القطر الرئيسي للمصفوفة Hat ويسمى هذا المجموع Effective degree of freedom (df) أي إن

$$df = trace(A_I) = tr(A_I) \quad (12)$$

ولإيجاد المعلمة التمهيدية نختار من بين مجموعة كبيرة من القيم المرشحة من قيم l

تلك القيمة التي تجعل قيمة المعيار في (11) أصغر ما يمكن.

5-2 انحدار الشرائح الجزائية (B) [6],[7] Penalized Spline Regression

قدم هذا النموذج من قبل Eilers&marks عام 1986 وسمياه P-Spline وتعني التمهيد باستخدام الشرائح (B) مع عقد بمسافات متساوية ولكن عددها دائما اقل من عدد البيانات وللسيطرة على مشكلة نقص التمهيد (Under Smoothing) استخدم الباحثان جزءا متقطع (discret penalty) من فروقات من رتب عليا لمعاملات الشرائح (B) المتجاورة وبعد حساب الشرائح (B) فان معيار المربعات الصغرى الجزائية يكتب بالصيغة الآتية:

$$PRSS = \frac{\sum_{i=1}^n y_i^2 - \sum_{j=1}^m a_j B_j(x_i; q)}{\sum_{j=k+1}^m (D^k a_j)^2} + l \quad (13)$$

حيث m عدد الشرائح (B) ، $B_j(x; q)$ يمثل قيمة الشريحة B_j من الدرجة q عند x ، a_j معامل الشريحة B_j ، $D^k a_j$ يمثل الفروق من الرتبة k للمعامل a_j ويحسب كما يلي: عندما $k=1$ فان $Da_j = a_j - a_{j-1}$ وعندما $k=2$ فان

$$D^2 a_j = D(Da_j) = a_j - 2a_{j-1} + a_{j-2}$$

l هي معلمة التمهيد (Smoothing Parameter) .

ويمكن كتاب الشرائح حيث (13) على شكل مصفوفات كالآتي:

$$PRSS = (y - Ba)^T (y - Ba) + l D_k^T D_k a \quad (14)$$

حيث $y_{n \times 1}$: متجه المشاهدات

$a_{m \times 1}$: متجه المعاملات

$B_{n \times m}$: مصفوفة الشرائح حيث $B_{i,j} = B_j(x_i)$

D_k : مصفوفة الفروقات من الرتبة k .

وبتصغير المعيار (14) للحصول على متجه المعامل المقدر نحصل على نظام المعادلات الطبيعية التالي :

$$(B^T B + l D_k^T D_k) \hat{a} = B^T y \quad (15)$$

ونلاحظ عندما $l = 0$ فان النظام السابق يقلل الى نظام المعادلات الطبيعية للانحدار الخطي الى y على B وعند استخدام عدد كبير من العقد في النظام فان المنحنى الناتج يكون أكثر تذبذبا وبزيادة قيمة l تزداد كمية التمهيد ولكن الزيادة الكبيرة ($l \in \mathbb{R}^+$) فان المقدر الناتج يكون خط مستقيم عندما ($k = 2$) و يكون منحنى تربيعي عندما ($k = 3$) .

إذا كانت $g(x) = \sum_{j=1}^m a_j B_j(x; q)$ فإن $g = Ba$ وبالتالي

$$\hat{g} = B\hat{a} \quad (16)$$

$$\hat{g} = B(B^T B + I D_k^T D_k)^{-1} B^T y \quad (17)$$

$$\hat{g} = H_I y \quad (18)$$

حيث المصفوفة H_I تسمى (Hat Matrix)

ولحساب معلمة التمهيد I نستخدم معيار GCV مع استبدال مصفوفة التمهيد H_I بدلا من A_I .

Nadaraya-Watson Estimator

6-2 مقدر N-W [9],[8],[12]

يعتبر من ابسط أنواع الممهيات وأكثرها شيوعا كما يظهر مفهوم المعدل المحلي الموزون (locally Weighted Averaging) بصورة مباشرة من خلال صيغة الأنموذج الذي وضع من قبل كل من Nadaraya(1964) و Watson(1964) لإيجاد تقدير الدالة g عند النقطة x في مجالها نحدد حولها جوارا بعرض مقداره h يسمى عرض الحزمة (Bandwidth) يتحكم بعرض الجوار حول x ولضمان حصول التمهيد تعطى المشاهدات القريبة من x وزنا اكبر من المشاهدات الأبعد وذلك الوزن يحدد من قبل دوال Kenel . ويكتب تقدير الدالة g بالشكل الآتي:

$$\hat{g}_{NW}(x) = \frac{\sum_{i=1}^n k\left(\frac{x_i - x}{h}\right) y_i}{\sum_{j=1}^n k\left(\frac{x_j - x}{h}\right)} \quad (19)$$

حيث $k(u)$ دالة Kernel ، h عرض الحزمة ويتحكم بحجم الجوار في دالة Kernel وبالتالي يتحكم بكمية التمهيد للمقدر الناتج حيث

عندما $h \rightarrow 0$ أي عرض الحزمة صغير جدا فان

$$\hat{g}_{NW}(x_i) \rightarrow \frac{k(0)y_i}{k(0)} = y_i \quad (20)$$

أي إن المنحنى المقدر يمر بجميع نقاط البيانات ويكون منحنى متذبذب وله تباين عالي وتحيز اقل ويسمى منحنى تحت التمهيد (Under Smooth).

وعندما $h \rightarrow \infty$ أي إن عرض الحزمة كبير جدا فان

$$\hat{g}_{NW}(x) = \frac{1}{n} \sum_{i=1}^n \frac{k\left(\frac{x-x_i}{h}\right) y_i}{\sum_{i=1}^n k\left(\frac{x-x_i}{h}\right)} \quad (21)$$

والمنحنى الناتج قريب من الخط المستقيم أو ما يسمى منحنى فوق التمهيد (Over smooth) ذو تباين قليل ولكن تحيز عالي.

ولذلك يتم اختبار عرض الحزمة ليوازن بين هاتين الغايتين المتعارضتين . ونلاحظ إن مقدر NW يعاني من التحيز عند حدود الفترة (Boundary Biase) لان صيغة التحيز تعتمد على

المقدار $\frac{f_x'(x)}{f_x(x)}$ حيث f_x دالة الكثافة الاحتمالية بمعنى إن التحيز يعتمد على مواقع نقاط البيانات x_i

وبكتابة الأنموذج بصيغة المصفوفات فان

$$\hat{g}_{NW} = WY \quad (22)$$

حيث $W = (w_1^T, \dots, w_n^T)$ ، $Y = (y_1, y_2, \dots, y_n)^T$ إذن

$$\hat{g}_{NW} = WY = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ \vdots & \vdots & \dots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad (23)$$

$$w_{ij} = \frac{k\left(\frac{x_i - x_j}{h}\right)}{\sum_{j=1}^n k\left(\frac{x_i - x_j}{h}\right)} \quad (24)$$

7-2 اختيار دالة Kernel :- [2],[4],[12]

ذكر (Hurdel) أن اختيار دالة Kernel ليس بالأمر المؤثر جدا في كمية التمهيد لذلك يتم اختيار دالة Kernel بحيث تحقق ما يلي :

دالة Kernel $k : R \otimes R$ وتحقق $\int_{-\infty}^{\infty} k(u) du = 1$

1- أنها دالة غير سالبة أي أن $k(u) \geq 0$

2- العزوم تحسب كما يلي $k_j(t) = \int_{-\infty}^{\infty} u^j k(u) du$

3- دالة متناظرة (symmetric) أي أن $k(-u) = k(u)$

والجدول التالي يوضح بعض دوال kernel من الرتبة الثانية

Kernel	Equation
Uniform	$k_0(u) = \frac{1}{2} \quad (u \leq 1)$
Epanechnikov	$k_1(u) = \frac{3}{4}(1-u^2) \quad (u \leq 1)$
Biweight	$k_2(u) = \frac{15}{16}(1-u^2)^2 \quad (u \leq 1)$
Triweight	$k_3(u) = \frac{35}{32}(1-u^2)^3 \quad (u \leq 1)$
Gaussian	$k_r(u) = \frac{1}{\sqrt{2p}} \exp(-\frac{u^2}{2}) \quad (-\infty < u < \infty)$

جدول (1) بعض دوال Kernel من الرتبة الثانية

8-2 اختيار عرض الحزمة [2],[12]

لاختيار عرض الحزمة نستخدم معيار العبور الشرعي CV (Cross Validation)

$$cv_h = \min_n n^{-1} \sum_{i=1}^n (y_i - \hat{g}_{i,h}(x_i))^2 w(x_i) \quad (25)$$

حيث $w(x_i)$ دالة وزن و $\hat{g}_{i,h}(x_i)$ تقدير NW للدالة g عند النقطة x_i بعد حذف المشاهدة y_i .

9-2 مقدر الانحدار الخطي الموضوعي [9],[10],[14] Local Linear Regression Estimator

إذا كانت الدالة g تمتلك مشتقة ثانية فان توسيع تايلر للدالة يكون بالشكل

$$g(u) \approx a + b(u - x) \quad (26)$$

أن الانحدار بمتعددات الحدود الموضوعية يفترض أن الدالة تنتمي الى عائلة معلمية (Parametric Family) معينة وهي هنا متعددة حدود من الدرجة الأولى ولتقدير المعالم نستخدم معيار المربعات الصغرى الموزون التالي:-

$$\min_{a,b} \sum_{i=1}^n (y_i - a - b(x_i - x))^2 k\left(\frac{x_i - x}{h}\right) \quad (27)$$

حيث k دالة Kernel و h عرض الحزمة (Bandwidth)

ويمكن كتابة المعيار السابق بصيغة المصفوفات

$$\min_b (Y - Xb)^T W (Y - Xb) \quad (28)$$

$$b = \begin{pmatrix} a \\ b \end{pmatrix}, \quad X = \begin{pmatrix} (x_1 - x) \\ (x_2 - x) \\ \dots \\ (x_n - x) \end{pmatrix}, \quad Y = (y_1, y_2, \dots, y_n)^T \quad \text{حيث}$$

$$W = \text{diag} \left\{ k \left(\frac{x_i - x}{h} \right) \right\}_{i=1}^n$$

وبتصغير المعيار (28) نحصل على متجه المعلمات المقدرة كما يلي

$$\hat{b} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = (X^T W X)^{-1} X^T W Y \quad (29)$$

من (26) نلاحظ إن تقدير الدالة g عند النقطة x يكون بالشكل :-

$$\hat{g}(x) = \hat{a} \quad (30)$$

$$\hat{g}(x) = \frac{\sum_j w_j y_j}{\sum_j w_j} \quad (31)$$

حيث

$$w_j = k \left(\frac{x_j - x}{h} \right) [s_{n,2} - (x_j - x) s_{n,1}] \quad (32)$$

$$s_{n,L} = \sum_{i=1}^n k \left(\frac{x_j - x}{h} \right) (x - x_j)^L, \quad L = 1, 2 \quad (33)$$

إن مقدر الانحدار الخطي الموضوعي لا يحتاج الى معالجة للتحيز قرب حدود الفترة (Boundary Bias) لأنه يتكيف بصورة مباشرة عكس مقدر NW حيث إن معدل التقارب لا يتأثر بمواقع نقاط البيانات. أما اختيار دوال Kernel و عرض الحزمة فهي كما سبق.

3- الجانب التجريبي (مقدمة)

لغرض تطبيق النماذج اللامعلمية الذي تم التطرق اليها في الجانب النظري- لا بد من وضع بعض الافتراضات المهمة للحصول على تحليل أكثر شمولية من خلال استخدام عدد كبير من العينات وبأحجام مختلفة أو اختيار 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) .

1-3 أسلوب المحاكاة

تعرف المحاكاة بأنها تقليد للواقع العملي بحيث تقوم بتوظيف نماذج تظهر فيها عدد كبير من الحالات الافتراضية لتكون نتائج التحليل أكثر شمولية وتعميم، وقد ظهرت الاستعانة بالمحاكاة أساسا كأحدى افرازات التقدم الحاصل في مجال الحاسبات الالكترونية من جانب ، ومن جانب آخر بسبب التطور الحاصل في مجال التحليل العددي. أما مسوغات العمل بالمحاكاة تكون 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.[2]

1-1-3 النماذج المستخدمة في تجارب المحاكاة:

تناولت الدراسة ثلاثة نماذج من دوال الاختبار موضحة كما يلي :-

$$f(x) = \sin(2px^3) + 0.5(x - 0.75)^2 \quad \text{النموذج الأول:}$$

$$f(x) = \sin(12(x + 0.2)) / (x + 0.2) + \exp(-12x) \quad \text{النموذج الثاني:}$$

$$f(x) = 1 - 48x + 18x^2 - 15x^3 + 45x^4 - x^5 \quad \text{النموذج الثالث:}$$

٢-١-٣ الأخطاء العشوائية وحجوم العينات:

1- الأخطاء العشوائية: تم توليدها لتتوزع توزيعا طبيعيا بمتوسط صفر وتباين ثابت σ^2 وتم تناول ثلاث مستويات من الانحراف المعياري :-

$$.s = 0.125 \quad \text{(أ)}$$

$$.s = 0.25 \quad \text{(ب)}$$

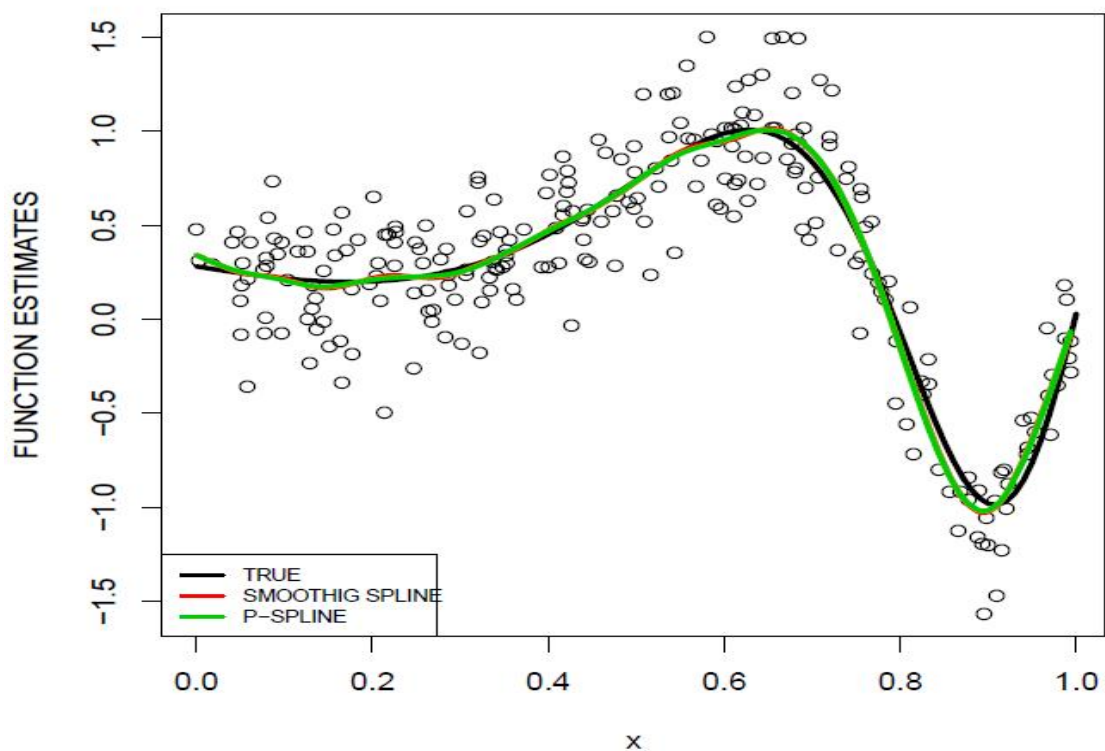
$$.s = 0.5 \quad \text{(ج)}$$

2- حجوم العينات :

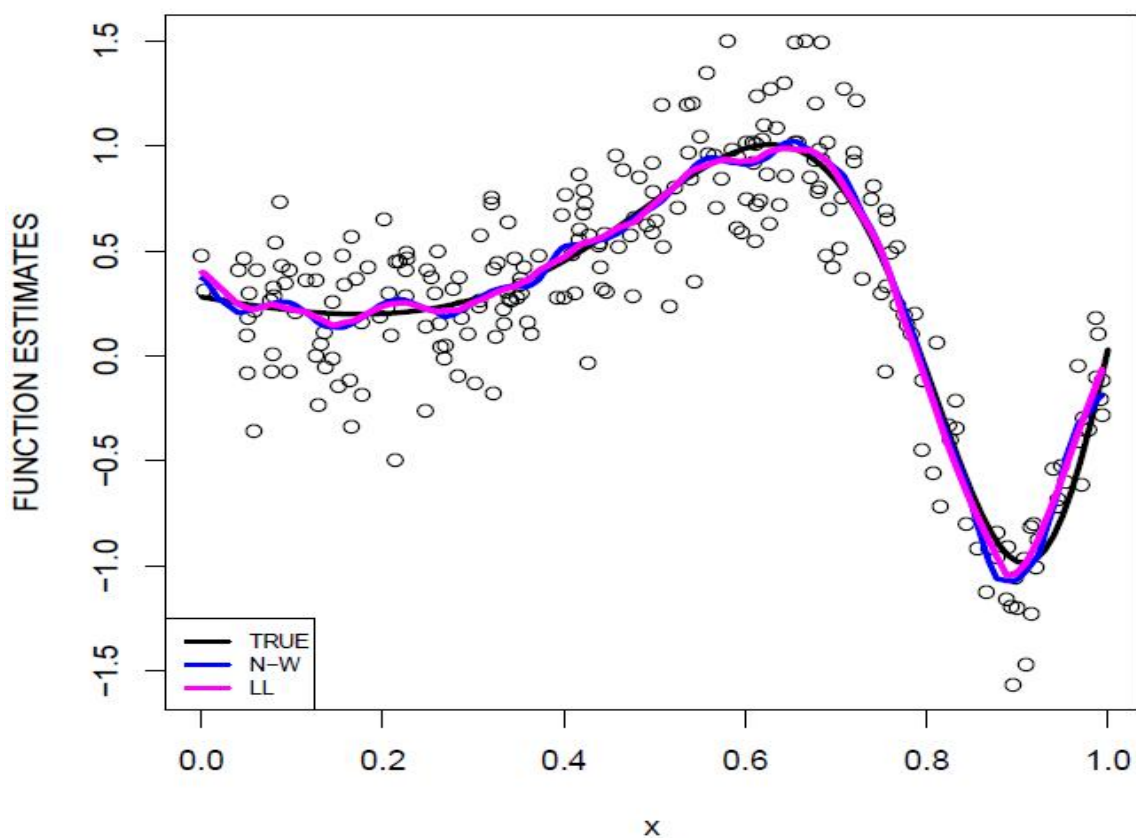
تم تناول ثلاث حجوم للعينات حجم عينة (٥٠) وحجم عينة (١٥٠) وحجم عينة (٢٥٠).

الجدول (2) نتائج تجارب المحاكاة للأنموذج الأول

تسلسل	الطريقة	حجم العينة	$\sigma = 0.125$		$\sigma = 0.25$		$\sigma = 0.5$	
			MSE	MAE	MSE	MAE	MSE	MAE
1	SMOOTHING SPLINE (GCV)	50	0.005479912	0.05525699	0.02204265	0.1080041	0.07483071	0.1970757
		150	0.001544138	0.03025819	0.005223824	0.05539687	0.01788708	0.1019358
		250	0.000950368	0.02382111	0.003377054	0.04481693	0.01125022	0.08119141
2	CUBIC P-SPLINE (GCV) B-BASIS	50	0.003497538	0.04563198	0.01281836	0.08652352	0.04428603	0.1615319
		150	0.001426871	0.02928732	0.004976031	0.05435558	0.01738032	0.1012834
		250	0.000875240	0.02298567	0.00324816	0.04402314	0.01115892	0.08102562
3	N-W EPAK	50	0.006794419	0.05974781	0.01658158	0.09656149	0.04342099	0.1598896
		150	0.002758852	0.03939301	0.007318801	0.06473335	0.02006932	0.1081003
		250	0.001897864	0.0326529	0.00496465	0.05325545	0.01435199	0.09021332
4	LL EPAK	50	0.005172739	0.05316037	0.01497336	0.09282982	0.04268578	0.1578225
		150	0.001912644	0.03381333	0.005843462	0.05922971	0.01836237	0.1043433
		250	0.001292871	0.02787973	0.003864246	0.04826274	0.01233156	0.08493575



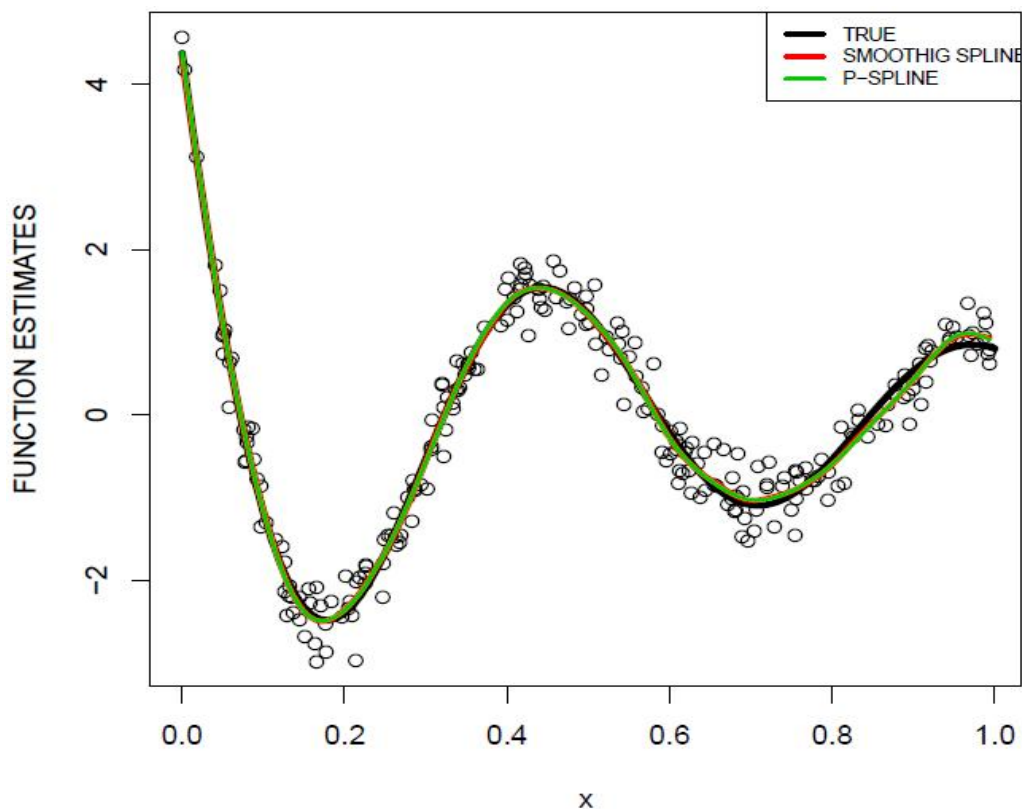
شكل(1) تقدير الشريحة الممهدة (Smoothing Spline) والشرائح الجزائية (P-Spline) للأنموذج الاول.



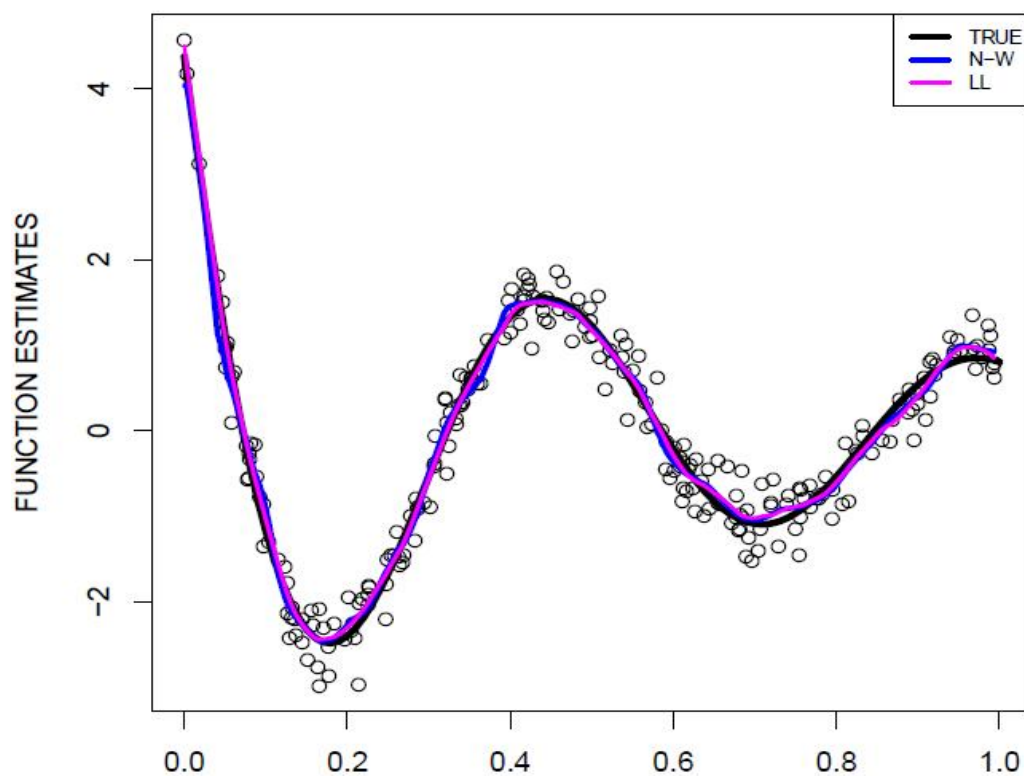
شكل(2) تقدير كل من NW ، الانحدار الخطي الموضعي(LL) للأنموذج الاول

جدول (3) نتائج تجارب المحاكاة للأنموذج الثاني

تسلسل	الطريقة	حجم العينة	$\sigma = 0.125$		$\sigma = 0.25$		$\sigma = 0.5$	
			MSE	MAE	MSE	MAE	MSE	MAE
1	SMOOTHING SPLINE (GCV)	50	0.005534268	0.0562402	0.02035086	0.1061829	0.08101789	0.2090057
		150	0.001733316	0.03205713	0.005845146	0.05870988	0.01939803	0.1079478
		250	0.001097556	0.02538902	0.003649871	0.04654172	0.01247312	0.08531095
2	CUBIC P-SPLINE (GCV) B-BASIS	50	0.003451452	0.04544333	0.01321104	0.08846263	0.04705696	0.1658207
		150	0.001537470	0.03036274	0.005441852	0.05682596	0.01888439	0.1064380
		250	0.000953082	0.02396824	0.003460585	0.04526596	0.01190135	0.08399692
3	N-W EPAK	50	0.03969211	0.1246173	0.05400925	0.1579092	0.0976164	0.2308780
		150	0.008161169	0.06225233	0.01662504	0.0941459	0.04239526	0.1533778
		250	0.004803725	0.04954502	0.01109984	0.07760494	0.02841703	0.1253187
4	LL EPAK	50	0.01492486	0.08721263	0.02869382	0.1277828	0.07241153	0.2092728
		150	0.002892833	0.04201458	0.008618604	0.07271241	0.02632823	0.1278209
		250	0.001785538	0.03304284	0.005535537	0.05849917	0.01730099	0.1028079



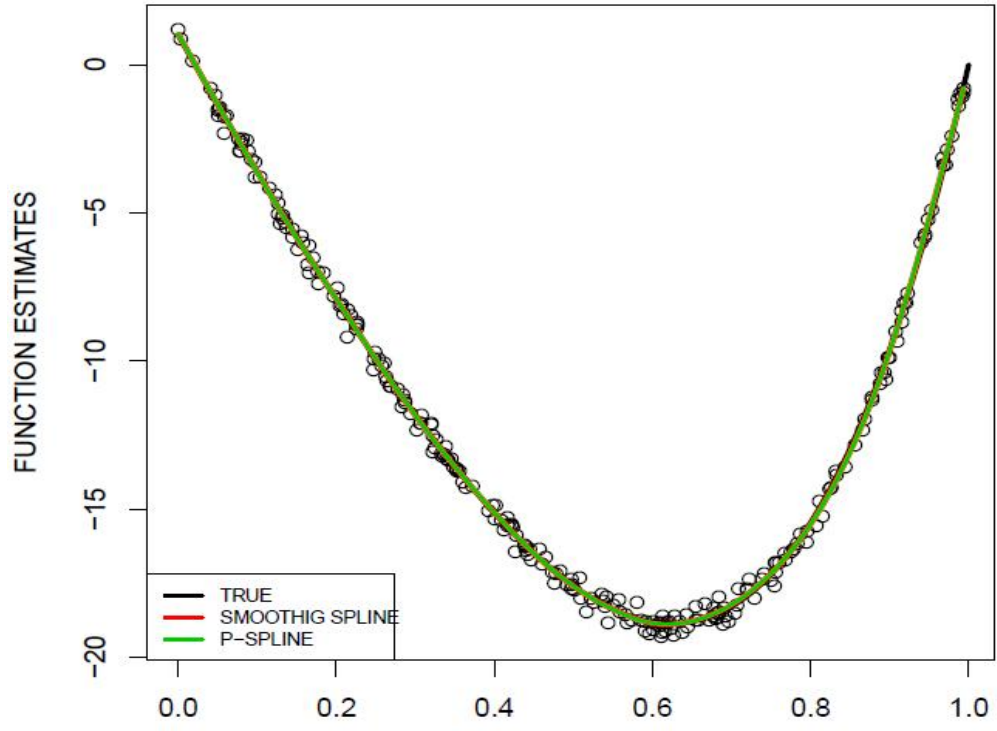
شكل (3) تقدير الشريحة الممهدة (Smoothing Spline) والشرائح الجزائية (P-Spline) للأنموذج الثاني



شكل (4) تقدير NW والاتحادار الخطي الموضعي (LL) للأنموذج الثاني

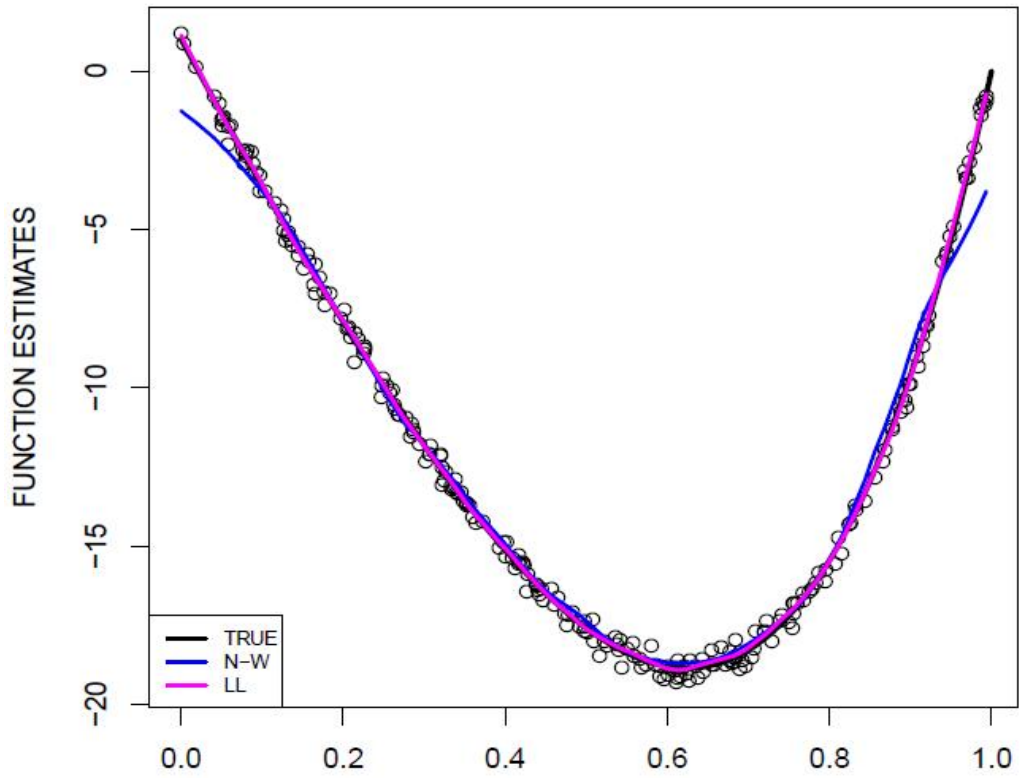
جدول(4)نتائج تجارب المحاكاة للأنموذج الثالث

تسلسل	الطريقة	حجم العينة	$\sigma = 0.125$		$\sigma = 0.25$		$\sigma = 0.5$	
			MSE	MAE	MSE	MAE	MSE	MAE
1	SMOOTHING SPLINE (GCV)	50	0.00551977	0.05562645	0.02010282	0.1045088	0.0765481	0.1996439
		150	0.001758145	0.03199512	0.005641662	0.05729555	0.01885706	0.1048534
		250	0.001163704	0.02582681	0.003712082	0.04634281	0.01253821	0.0842926
2	CUBIC P-SPLINE (GCV) B-BASIS	50	0.003702819	0.04674886	0.01330864	0.08793449	0.04500055	0.1619541
		150	0.00160635	0.03078419	0.005459213	0.05657466	0.01863855	0.1040385
		250	0.001000640	0.02435113	0.003575588	0.04580297	0.01199707	0.08332756
3	N-W EPAK	50	0.1511819	0.2566094	0.1695506	0.2809081	0.2173623	0.3360412
		150	0.02348108	0.1060473	0.0347509	0.1354244	0.07000816	0.1985256
		250	0.01140826	0.07603184	0.01980168	0.1048284	0.0450142	0.1593378
4	LL EPAK	50	0.01415359	0.0844769	0.02707644	0.1249777	0.06730523	0.2019558
		150	0.002731360	0.04080923	0.008317931	0.07206578	0.02555664	0.1256302
		250	0.001760647	0.03298168	0.0053379	0.05737761	0.01689004	0.1018573



شكل (٥)

تقدير الشريحة الممهدة (Smoothing Spline) والشرائح الجزائية (P-Spline) للأنموذج الثالث.



شكل (٦) تقدير NW والانحدار الخطي الموضعي (LL) للأنموذج الثالث.

4- الاستنتاجات:

مما تقدم يمكن استنتاج النتائج الآتية:

- 1- بالنسبة للأنموذج الأول أظهرت النتائج تفوق الطرق الشرائحية وهي الشريحة الممهدة والشرائح الجزائية على طرق NW ، LL في حجوم العينات المتوسطة والكبيرة ولجميع قيم الخطأ ، بينما حققت طرق NW ، LL تفوقا في حجم العينة الصغير . وكانت النتائج متقاربة بالنسبة للشرائح الجزائية والشريحة الممهدة مع تفوق للشرائح الجزائية، أما طرائق Kernel فكان التفوق لطريقة LL . ونلاحظ من خلال الشكل (1) أن مقدر الشريحة الممهدة يعاني من التذبذب في الجهة اليسرى وكذلك في الشكل (2) يبدو التذبذب واضحا لكلا الطريقتين NW ، LL ويمثل هذا التذبذب إخفاقا من طريقة العبور الشرعي في اختيار قيمة مثلى للمعلمة التمهيدية.
- 2- بالنسبة للأنموذج الثاني أظهرت النتائج تفوقا واضحا للطرائق الشرائحية على طرائق Kernel ويلاحظ كذلك أن نتائج NW ، LL كانت بعيدة الى حد ما عن نتائج الطرائق الشرائحية ولجميع الحالات وتظهر النتائج تقاربا في النتائج لطريقتي الشريحة الممهدة والشرائح الجزائية مع تفوق الأخيرة ، إما طرائق Kernel فيبقى التفوق لطريقة LL على NW . ويلاحظ من شكل (3) بعض التذبذب لمقدر الشريحة الممهدة بين القيم 0.6 ، 0.8 أما الشكل (4) فيلاحظ إخفاقا واضحا من كلا الطريقتين NW ، LL مع تمهيد أفضل لمقدر LL.
- 3- بالنسبة للأنموذج الثالث كانت النتائج متقاربة مع الأنموذج الثاني وكان ترتيب النتائج الشرائح الجزائية ، الشريحة الممهدة ، LL ، NW . ونلاحظ من شكل (6) التحيز الواضح عند أطراف الفترة لمقدر NW والأمر لا يظهر بالنسبة لمقدر LL والذي يعالجه بصورة مباشرة كما ذكرنا في الجانب النظري .
1. نلاحظ من خلال الجداول (1) ، (2) ، (3) أن قيم المعيارين MSE ، MAE تقل كلما زاد حجم العينة.

5- التوصيات:

- 1- بالنسبة للطرائق الشرائحية نوصي باستخدام معيار اخلاختيار المعلمة التمهيدية أكثر ملائمة من معيار GCV ، مثل معيار AIC حيث يعاني معيار GCV من التذبذب أو استخدام نماذج أخرى مثل مقدر الشريحة الموضوعية (Local Penalty Spline) حيث تتغير المعلمة التمهيدية / عبر العقد عكس النماذج التقليدية التي تستخدم معلمة تمهيدية ثابتة عبر العقد (Global Penalty).
- 2- استخدام درجة متعددة حدود اكبر من الأولى مثل متعددة حدود تربيعية أو تكعيبية في الانحدار بمتعددات الحدود الموضوعية لكونها أكثر تكيفا مع النماذج غير البسيطة.

المصادر العربية:

١. علي ، عمر عبد المحسن،(2007)"مقارنة مقدرات النماذج التجميعية المعممة باستخدام الشرائح التمهيدية عند تحليل الانحدار اللامعلمي والشبه معلمي"، أطروحة دكتوراه فلسفة في الإحصاء ،كلية الإدارة والاقتصاد،جامعة بغداد.
٢. يوسف،خلود يوسف خمو،(2004)"مقارنة أساليب بيز مع طرائق أخرى لتقدير منحني الانحدار اللامعلمي". أطروحة دكتوراه فلسفة في الإحصاء ،كلية الإدارة والاقتصاد،جامعة بغداد.

المصادر الأجنبية:

3. Azzalini, A. and Bowman, A. W. (1997). Applied Smoothing Techniques for Data Analysis. Oxford University Press, Oxford.
4. CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. J. Amer. Statist. Assn. 74, 829-836.
5. de Boor, C. (2001). A Practical Guide to Splines. Revised edition. Applied Mathematical Sciences 27. Springer-Verlag, New York
6. Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing using B-splines and penalized likelihood (with Comments and Rejoinder). Statistical Science 11(2): 89-121.
7. Eilers, P.H.C. and Marx, B.D.(2004)."Splines,Knots,and Penalties".Research supported in part for Brian Marx by NSF Grant DMS-0102131.
8. Eubank, R. L. (1999).Nonparametric Regression and Spline Smoothing. Marcel Dekker, NewYork, NY.
9. Fan, J. (1992a). Design-adaptive nonparametric regression, J. Am. Statist. Assoc.87, 998{1004.
10. FAN, J. and GIJBELS, I. (1992b). Variable bandwidth and local linear regression smoothers. Ann. Statist. 20, 2008-2036
11. Hastie.T,Tibsherani.R and Friedman.J(2009)."The Element of Statistical Learning".Springer.ISBIN:978-387-84857-0.
12. HURDLE, W. (1990). Applied Nonparametric Regression. Oxford University Press, Oxford.
13. Ibrahim.N.A and Suliadi.(2009)"Nonparametric Regression for Correlated Data".WSEAS TRANSACTION ON MATHEMATICS ,ISSN : 1109-2769,Issue7,Volume8
14. LOADER, C. R. (1996). Local Regression and Likelihood. Electronic book, <http://cm.bell-labs.com/stat/project/locfit/>.
15. Ruppert, D., Wand, M.P., and Carroll, R.J. (2003). Semiparametric Regression. CambridgeUniversity Press, New York.
16. 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.
17. WAHBA, G. (1990). Spline Functions for Observational Data. SIAM, Philadelphia.
18. Wand,M. P. (2000), "A Comparison of Regression Spline Smoothing Procedures," Computational Statistics, 15,443-462.