مجلة القادسية لعلوم الحاسوب والرياضيات المجلد(٧) العدد(٢) السنة(٢٠١٥)



صفحة ٢٩ ـ ٤٢

**مقارنة بعض المقدرات شبه المعلمية للأنموذج الانحدار الخطي الجزئي باستخدام المحاكاة**

**محمد طالب خنجر**
**جامعة القادسية**
**كلية علوم الحاسوب والرياضيات**

**محمد حبيب الشاروط**
**جامعة القادسية**
**كلية الادارة والاقتصاد**

**استلام البحث ٢٠ / ٤ / ٢٠١٥** **إرسال التعديلات ٢٤ / ٥ / ٢٠١٥** **قبول النشر ٧ / ٦ / ٢٠١٥**

**المستخلص:**

يهدف هذا البحث الى إجراء دراسة مقارنة بين بعض طرائق تقدير نموذج الانحدار الخطي الجزئي شبه المعلمي باستخدام المحاكاة وايجاد افضل مقدر لتمثيل البيانات المدروسة إعتماداً على نتائج تجارب المحاكاة ، حيث تم استخدام طريقة المربعات الصغرى (OLS) لتقدير الجزء المعلمي اما الجزء اللامعلمي فقد تم تقديره بطريقتين هما ممهد متعدد الحدود الموضعي ( Local Polynomial Smoother) وطريقة تمهيد الشريحة التكعيبية (Cubic Smoothing Spline) . وأظهرت نتائج تجارب المحاكاة ان نموذج الانحدار الخطي الجزئي شبه المعلمي المقدر بطريقة تمهيد الشريحة التكعيبية هو النموذج الافضل من حيث النماذج المطبقة.

## 1- المقدمة

لقد تناولت العديد من الدراسات والبحوث السابقة مواضيع متعددة شملت الطرائق المعلمية التي تكون فيها توزيعات المتغيرات المدروسة معلومة التوزيع ، واستخدمت الطرق المعروفة في التقدير كطريقة المربعات الصغرى (LS) وطريقة الامكان الاعظم(ML) واتجهت بحوث اخرى الى التقديرات اللامعلمية عندما تكون التوزيعات لتلك المتغيرات المدروسة غير معلومة ، اضافةً الى ذلك فقد دمجت بحوث اخرى بين الاسلوبين السابقين لمعالجة مشكلة ظهور جزء من المتغيرات بشكل معلمي وجزء اخر بشكل لامعلمي وقد تم تقدير النموذج المدروس باساليب مختلفة لتلك العلاقات ، حيث تم عمل تقديرات للمعلمات وعمل تقديرات للمتغيرات ذات التوزيع غير المعروف .

وفي هذا البحث تم الاتجاه الى الاسلوب الاخير في التقدير عند التعامل مع البيانات ، ولغرض تحليل تلك البيانات احصائياً قام الباحث باجراء عملية محاكاة لمجموعة دوال وباستخدام طريقتين تتعامل مع النماذج اللامعلمية وهي ممهد متعدد الحدود الموضعي و تمهيد الشريحة التكعيبية . اما الجزء المعلمي فقد تم تقديره بطريقة المربعات الصغرى (OLS)

في عام 1988 قدم الباحث (Speckman)[4] دراسة عرض فيها استعمال ممهد اللب (Kernel Smoothing)في تقدير النماذج الخطية الجزئية ، وفي عام 2003 قدم الباحثون (List,Millimet&Stengos)[5] دراسة تضمنت اجراء مقارنة بين نموذج شبه معلمي ونموذج معلمي عند دراستهم مشكلة تلوث الهواء في الولايات المتحدة .

## 2- الجانب النظري

### 2-1 نموذج الانحدار شبه المعلمي Partial Linear Regression Model

يعد نموذج الانحدار الخطي الجزئي (PLM) هو احد نماذج الانحدار شبه المعلميه[4] وهو من النماذج التي تعتمد على متغيرات خطية (Linear) واخرى غير خطية لامعلميه (Nonparametric) وهي عادةً ما تكون متغيرات مستمرة ، كما وان كافة هذه المتغيرات تؤثر في المتغير المعتمد (متغير الاستجابة) .

والصيغة العامة لهذا النوع من النماذج هي:[6]

$$Y_i = X_i'\beta + g(t_i) + \varepsilon_i \qquad ,1,2,\dots\dots,n \qquad \dots\dots\dots\dots\dots\dots\dots\dots..(1)$$

حيث الاخطاء العشوائية $\varepsilon_i$ تتوزع توزيع طبيعي بمتوسط صفر وتباين ثابت $\sigma^2$ أي ان :

$\varepsilon_i \sim N(0, \sigma^2)$

ويعد نموذج الانحدار الخطي الجزئي حالة خاصة من النماذج التجميعية Additive models ولهذا فهو أفضل من النماذج اللامعلميه بسبب تجنبه مشكلة البعدية (The curse of dimensionality) التي تحدث في النماذج اللامعلميه عند زيادة عدد المتغيرات ومن جهة أخرى ايضاً يعد أكثر مرونة من النماذج الخطية القياسية وذلك بسبب التقليل من الافتراضات الخطية المفروضة على هذه النماذج .[7]

ويمكن التعبير عن النموذج الموصوف بالمعادلة (1) بشكل مصفوفات كما يلي :

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1K} \\ 1 & X_{21} & X_{22} & \cdots & X_{2K} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_n) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad \dots\dots..(2)$$

أي ان :

$$Y = X\beta + g + \varepsilon \qquad \dots\dots\dots\dots\dots\dots\dots.(3)$$

حيث ان

Y : متجة متغير المعتمد أو متغير الاستجابة من الدرجة( $n \times 1$)

X : مصفوفة المتغيرات التوضيحية (المعلمية) من الدرجة ($n \times (k+1)$

□ : متجة المعلمات المجهولة من درجة($(k+1) \times 1$)

$g$ : متجه دالة تمهيدية غير معرفة من الدرجة( $n \times 1$).

□ : متجة الاخطاء العشوائية من درجة (nx1) وتكون مستقلة بمتوسط صفر وتباين $\sigma^2$.

وللانموذج أعلاه تسميات اخرى ، اذ يسمى بنموذج الانحدار شبه المعلمي (Semiparametric regression model)[8,4] أو النموذج الخطي الجزئي (partially linear model)[9] وسبب تسمية النموذج بالخطي الجزئي لكونه يتضمن جزء معلمي خطي وجزء لامعلمي وترتبط هذه الاجزاء مع بعضها بعلاقة تجميعية [6] .

## 2-2 طرائق تقدير نموذج الانحدار الخطي الجزئي (PLM)

من الواضح ان النماذج الخطية الجزئية شبه المعلمية تتضمن جزئين معلمي ولامعلمي لذلك استعملنا في هذا البحث طريقة المربعات الصغرى لتقدير الجزء المعلمي ، اما الجزء اللامعلمي فتم تقديره وفقاً للطرائق الاتية .

- Local Polynomial Smoother
- Cubic smoothing spline Estimator

### 1-2-2 ممهد متعدد الحدود الموضعي[10] Local Polynomial Smoother

يعد هذا الممهد من الممهدات الجيدة المستخدمة في التمهيد اللامعلمي ويفضل على ممهدات Kernel فضلاً عن ذلك يتمتع هذا الممهد بقابليته على التكييف مع التصميم ، سواء كان التصميم ثابت او عشوائي . وتمتلك ممهدات الانحدار الخطي الموضعي ( Local Linear Regression Smoother) كفاءة عالية عند مقارنتها مع الممهدات الاخرى المختلفة ، أي من الممكن ان تكون ملائمتها بنسبة عالية جداً مع خيار دالة اللب وكذلك عرض الحزمة من بين كل الممهدات الممكنة .[11]

فاذا افترضنا ان الدالة $g$ تمتلك مشتقات من الرتبة p عند النقطة $t$ ، والنقاط $T_i$ تقع في جوار النقطة $t$ فان توسيع تايلر (TaylorExpansion)[1] للدالة g يعطى بالشكل الآتي :

$$\mathrm{g}(T_i) = a_0 + a_1(T_i - t) + a_2(T_i - t)^2 + \cdots + a_P(T_i - t)^P$$

وللحصول على قيم $a_j$ يتم اجراء التفاضل المتسلسل للدالة اعلاه :

$$\mathrm{g}'(T_i) = a_1 + 2a_2(T_i - t) + \cdots + pa_p(T_i - t)^{p-1}$$

$$\mathrm{g}''(\mathrm{T_i}) = 2a_2 + 6a_3(\mathrm{T_i} - \mathrm{t}) + \cdots + \mathrm{p}(\mathrm{p}-1)a_{\mathrm{p}}(\mathrm{T_i} - \mathrm{t})^{\mathrm{p}-2}$$

$$\vdots$$

$$\mathrm{g}^p(T_i) = p!\, a_p$$

وبافتراض ان :

$$a_j = \frac{\mathrm{g}^j(t)}{j!} = \alpha_j \qquad , \quad j = 0,1,2, \ldots \ldots . . p$$

وبالتعويض عن $a_j$ بـ $\alpha_j$ في المعادلة نحصل على :

$$\mathrm{g}(T_i) = \alpha_0 + \alpha_1(T_i - t) + \alpha_2(T_i - t)^2 + \cdots + \alpha_P(T_i - t)^P$$

$$g(T_i) = \sum_{j=0}^{p} \alpha_j (T_i - t)^j \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(4)$$

وبالتعويض عن $g(Ti)$ بما يساويها في المعادلة رقم (1) نحصل على :

$$Y_i = X_i'\beta + \sum_{j=o}^{p} \alpha_j \left(T_i - t\right)^j + \varepsilon_i \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(5)$$

$$\therefore\ \varepsilon_i = Y_i - X_i'\beta - \sum_{j=o}^{p} \alpha_j (T_i - t)^j \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(6)$$

بتربيع طرفي المعادلة (6) وضربها بـ $k_h(T_i - t)$ ثم ادخال المجموع على الطرفين ينتج:

$$\sum_{i=1}^{n} \varepsilon_i{}^2 * k_h(Ti- t) = \sum_{i=1}^{n} \left(Y_i - X_i'\beta - \sum_{j=0}^{p} \alpha_j\, (T_i - t)^j\right)^2 * K_h(T_i - t)$$

$$i = 1,2, \ldots\ldots, n \quad ; \quad j = 0,1, \ldots\ldots, p$$

حيث ان $kh(.)$ هي دالة Kernel ومن خصائص هذه الدالة ان تكون دالة حقيقية متماثلة محددة ومستمرة وتكاملها مساوي الى الواحد أي ان :

$$\int k(u)du = 1$$

وان $h$ تمثل معلمة عرض الحزمة .

وبتحويل المعادلة (5) الى شكل مصفوفات فتكون كما ياتي :

$$\alpha = [\alpha 0, \alpha 1, \ldots\ldots, \alpha p]$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad T = \begin{bmatrix} 1 & (T_1 - t) \ldots\ldots (T_1 - t)^p \\ \vdots & \vdots \qquad\qquad \vdots \qquad \vdots \\ 1 & (T_n - t) \ldots\ldots (T_n - t)^p \end{bmatrix}$$

$$W = \begin{bmatrix} k_h(t - T_1) \ldots\ldots & 0 \\ \vdots \qquad\qquad \ddots & \vdots \\ 0 \cdots \qquad k_h(t - T_1) \end{bmatrix} = \text{diag}k_h(t - T_i)$$

اذ ان المعادلة (5) تصبح بالشكل الاتي :

$$Y = X\beta + T\alpha + \varepsilon \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7)$$

وباستعمال طريقة المربعات الصغرى الموزونة (WLS) للتقدير وكما يلي :[10]

$$\frac{\partial \varepsilon' w \varepsilon}{\partial \alpha'} = -T'W(Y - X\beta - T\alpha)$$

وبجعل $\frac{\partial \varepsilon' w \varepsilon}{\partial \alpha'}|_{\alpha=\hat{\alpha}} = 0$ فان الحل $\hat{\alpha} = (\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_p)'$ هو تقدير المربعات الصغرى الموزونة ويكون كما يلي :[4]

$$\hat{\alpha} = (T'WT)^{-1}T'W\big(Y - X\hat{\beta}_{LP}\big) \qquad \dots\dots\dots\dots\dots\dots\dots\dots. (8)$$

اما بالنسبة لتقدير معالم الجزء المعلميβ سيكون من خلال استعمال طريقة المربعات الصغرى وكما يلي :[10]

$$\varepsilon'\varepsilon = (Y - X\beta - T\alpha)'(Y - X\beta - T\alpha)$$

$$\frac{\partial \varepsilon'\varepsilon}{\partial \beta'} = -X'(Y - X\beta - T\alpha)$$

وبجعل $\frac{\partial \varepsilon'\varepsilon}{\partial \beta'}|_{\beta=\hat{\beta},\alpha=\hat{\alpha}} = 0$ ينتج ان :

$$X'X\hat{\beta} = X'Y - X'T\hat{\alpha} \qquad \dots\dots\dots\dots\dots\dots\dots\dots. (9)$$

وبتعوض المعادلة (8) في المعادلة (9) ينتج :

$$X'(I - T(T'WT)^{-1}T'W)X\beta = X'(I - T(T'WT)^{-1}T'W)Y$$

وبافتراض ان

$$H = I - T(T'WT)^{-1}T'W$$

$$\therefore\ \hat{\beta}_{LP} = (X'HX)^{-1}X'HY \qquad \dots\dots\dots\dots\dots\dots\dots\dots.. (10)$$

ويمكن الحصول على خصائص المقدر المعلمي وذلك من خلال أيجاد المتوسط والتباين له وكما ياتي :

بتعويض المعادلة (7) في المعادلة (10) ينتج ان :

$$\hat{\beta}_{LP} = (X'HX)^{-1}X'H(X\beta + T\alpha + \varepsilon)$$

$$\hat{\beta}_{LP} = (X'HX)^{-1}X'HX\beta + (X'HX)^{-1}X'HT\alpha + (X'HX)^{-1}X'H\varepsilon$$

وعند ملاحظة المعادلة اعلاه نجد ان الحد الوسط فيها مساوي الى الصفر :

$$\therefore\ \hat{\beta}_{LP} = \beta + (X'HX)^{-1}X'H\varepsilon \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots. (11)$$

وباخذ التوقع لطرفي المعادلة اعلاه ينتج ان[3] :

$$E(\hat{\beta}_{LP}) = \beta$$

اما تباين المقدرات $\hat{\boldsymbol{\beta}}_{LP}$ كما يلي :

$$\hat{\beta}_{LP} - \beta = (X'HX)^{-1}X'H\varepsilon$$

$$V - COV(\hat{\beta}_{LP}) = E(\hat{\beta}_{LP} - \beta)(\hat{\beta}_{LP} - \beta)'$$

$$= (X'HX)^{-1}X'HH'X(X'HX)^{-1}\sigma_{\varepsilon}^{2}$$

### 1-1-2-2 طريقة العبور الشرعي لتقدير معلمة عرض الحزمة[13,12]

تعد هذه الطريقة احد افضل طرائق إختيار معلمة عرض الحزمة واكثرها استعمالاً وتسمى ايضا بطريقة **leave-one-out)** وان عمل هذه الطريقة انه في كل مرة تستبعد قيمة واحدة من قيم المتغيرات $\{\mathbf{x_i}, \mathbf{t_i}, \mathbf{y_i}\}_{\mathbf{i=1}}^{\mathbf{n}}$ بشكل تدريجي لتحديد المعلمة $(\boldsymbol{h})$ التي تجعل مجموع مربعات البواقي للدالة اقل مايمكن بالاعتماد على $(\boldsymbol{n-1})$ من المتغيرات .

ولحساب قيمة معلمة التمهيد وذلك من خلال الصيغة التالية:

$$\boldsymbol{CV(h) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{g}_h^{(-1)}t_i)^2}$$

وبالتالي يتم اختيار قيمة المعلمة التمهيدية $\boldsymbol{h}$ التي تعطي اصغر قيمة للمعيار $(\boldsymbol{CV})$أي ان

$$\boldsymbol{h_{cv}} = \mathbf{arg\,min}\ \boldsymbol{cv_{(h)}}$$

### 2-2-2 تمهيد الشريحة التكعيبية [6,2]

تعد تمهيد الشريحة التكعيبية من الممهدات المستخدمة في موائمة المنحنى اللامعلمي لمجموعة من البيانات والذي تعتمد على احتساب مجموع مربعات البواقي **(RSS)** مضافا اليها حد الجزاء **(roughness penalty)** وكما ياتي :

$$\sum_{\mathbf{i=1}}^{\mathbf{n}}(\mathbf{y_i} - \mathbf{x'_i\beta} - \mathbf{g(t_i)})^2 + \boldsymbol{\lambda}\int_{\mathbf{a}}^{\mathbf{b}}(\mathbf{g^m(t)})^2\mathbf{dt} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{12})$$

$\boldsymbol{m}$: تمثل المشتقة من الدرجة $m^{th}$ للدالة g

ولايجاد المقدر يتم عن طريق تصغير الصيغة رقم (12) .

حيث يمثل الحد الاول من الصيغة رقم (12) مجموع مربعات البواقي و λ : تمثل معلمة التمهيد . أما الحد الثاني من الصيغة رقم (12) فيمثل حد الجزاء والذي يكون موزوناً بـ λ ، وان هذا الحد يكون كبير عندما يكون تكامل مربع المشتقة الثانية لدالة الانحدار g(t) كبير ، كما إن معلمة التمهيد λ تؤدي دوراً مهماً في تحديد قيمة الجزاء غير الممهد (roughness penalty) ،أي عندما معلمة التمهيد λ تقترب من الصفر هذا يعني ان الحد الثاني (حد الجزاء) سيختفي من المعادلة وان مقدر مجموع مربعات البواقي سيوضح البيانات أما اذا كانت قيمة معلمة التمهيد $\lambda$ كبيره جداً أي$\lambda \to \infty$ فأن المقدار يسيطر على مجموع مربعات البواقي وينتج منحني ثابتاً للانحدار الخطي اي يطابق الأنموذج المفسر التوضيحي مع توضيح كامل للمنحنى وتوضح البيانات بمجموعة غير منتهية من عرض الحزمة .

ولذلك فان معلمة التمهيد λ تعمل كمفتاح للتحكم بين حسن المطابقة المتمثلة بـ $\sum_{i=1}^{n}(y_i - x_i{}'\beta - g(t))^2$ وممهد التقدير مقاساً بـ $\int[g^m(t)^2]$ .

ويكون هناك شرطاً اساسياً على الدالة g هو أن تكون قابلة للاشتقاق مرتين ومع امكانية لاجراء تكامل مربع المشتقة الثانية لها .[3,2]

وان المقدر بهذه الطريقة يدعى بمقدر الشريحة التمهيدية من الدرجة $2m - 1$ ، وبصورة خاصة عند افتراض m=2 نحصل على شريحة ممهدة تكعيبية (Cubic smoothing spline) .[6]

إن دالة الجزاء من الصعب جدا تقديرها برمجياً وتحتاج الى جهد كبير وامكانيات رياضية متطورة ،ولكن في عام 1994 ناقش الباحثان (Green&Silverman)[6] طريقة لحل الجزاء غير الممهد وكما يأتي :

بفرض ان هناك n من المشاهدات $(t_1, t_2, \dots, t_n)$ في الفترة $[a, b]$ ،وان الدالة $g$ تمثل شريحة تكعيبية (cubic spline) اذا تحقق الشرطين الاتيين :[6]

- الدالة $g$ متعددة حدود تكعيبية في كل فترة $(t_i\ , t_{i+1})$ .
- الدالة $g$ ومشتقاتها الاولى والثانية هي دوال مستمرة عند نقاط البيانات $t_i$ .

واذا كانت المشتقة الاولى والثانية للدالة $g$ هي صفر عند نقاط الحد a و b ، لذلك فان الدالة $g$ تكون دالة خطية (linear function) في النقطتين $(a, t_1)$ و $(t_n, b)$ وبفرض ان $a = 0 , b = 1$ وموافقة الدالة $g$ لشروط الحد الطبيعية فان المقدر يسمى شريحة تكعيبية طبيعية (natural cubic spline).

وبفرض أن $g_i = g(t_i)$ و $\upsilon_i = g\,''(t_i)$ $\quad i = 1,2, \dots , n$

وان من خواص الشريحة التكعيبية الطبيعية فان $g\,''(a) = 0\ ,\ g\,''(b) = 0$

نفرض أن $g$ هي متجة $(n \times 1)$ حيث ان $g = (g_1, g_2, \dots, g_n)'$ وان $\upsilon$ تمثل متجة المشتقة الثانية $\upsilon = (\upsilon_2, \upsilon_3, \dots \dots \upsilon_{n-1})'$ من الدرجة $((n-2) \times 1)$ ، وان المتجهين $g$ و $\upsilon$ هما اللذان يحددان شكل المنحنى $g$ .

ويمكن تعريف هذه المتجهات عن طريق أثنين من المصفوفات $R$ و $Q$ و كما يأتي :

a: يمثل متجه الفروق بين المشاهدتين $t_{i-1}, t_i$ و عناصرها $(i^{th})$ كالاتي

$$a_i = t_{i+1} - t_i \qquad i = 1,2, \dots \dots, n-1$$

$Q$ مصفوفة بدرجة $(n \times (n-2))$ عناصر $q_{\text{ij}}$ ويتم حسابها كما يأتي :

$$q_{(j-1,j)=a_{(j-1)}^{-1}}$$

$$q_{jj=-a_{(j-1)}^{-1}-a_j^{-1}}$$

$$q_{(j+1),j=}a_j^{-1}$$

$$q_{(i,j)=0\,,}\forall |i-j| \geq 2$$

$$i = 1,2, \dots . n \;\; ; j = 2,3, \dots . n-1$$

$R$ مصفوفة متماثلة (symmetric) من الدرجة $(n-2) \times (n-2)$ ويتم حساب عناصرها وفق الاتي:

$$r_{ii} = (a_{i-1} + a_i)/3 \qquad ; i = 2,3 \dots \dots, n-1$$

$$r_{i,i+1} = r_{i+1,i} = \frac{a_i}{6}; i = 2,3, \dots \dots ., n-2$$

$$r_{ij} = 0 \quad , \forall \; |i-j| \geq 2$$

$$i = 2, \dots . n-1 \,; j = 2, \dots . n-1$$

وبما أن $R^{-1}$ موجود لذلك يمكن تعريف المصفوفة $K$ كما ياتي :

$$K = QR^{-1}Q\,' \qquad \dots \dots \dots \dots \dots \dots \dots . . (13)$$

واثبت الباحثان (Green and Silverman )[8] ان المتجهين $g$ و $\upsilon$ هما شرائح تكعيب طبيعية إذا تحقق الشرط التالي :

$$Q\,'g = R\upsilon$$

اذ يمكن حساب تكامل مربع المشتقة الثانية للدالة $g$ عند توفر الشرط اعلاه وكما يلي

$$\int_a^{\mathrm{b}} (g''(\mathrm{t}))^2 \mathrm{dt} = \upsilon' \mathrm{R} \upsilon = g' \mathrm{k} g$$

وفي عام 1988 ناقش الباحث( Speckman )[15,14,6] تقدير نموذج الانحدار شبه المعلمي بالاعتماد على شريحه التمهيد التكعيبية حيث إفترض ان متجه القيم للدالة $g$ عند نقاط العقد التي تمثل المشاهدات $(\mathrm{t}_1, \mathrm{t}_2, \dots \dots \mathrm{t}_\mathrm{n})$ هو $g = (g(t_1), \dots \dots, g(t_n))$ . أما $\hat{g}_\lambda$ فتمثل مقدر شريحة تكعيبية (cubic spline estimator) للمتجه . وأن

$$Y = (y_1 \dots \dots . y_n)\,'$$

$$\hat{g} = [\hat{\mathrm{g}}_\lambda(\mathrm{t}_1) \ldots \ldots \ldots . \hat{\mathrm{g}}_\lambda(\mathrm{t}_\mathrm{n})]'$$

$$= (\mathrm{S}_\lambda)(\mathrm{y}_1 \ldots \ldots . \mathrm{y}_\mathrm{n})'$$

ويمكن تمثيل المقدر باستعمال صيغة المصفوفات وكما ياتي:

$$\hat{g}_\lambda = \mathrm{S}_\lambda Y$$

$\lambda$ تمثل معلمة التمهيد

$\mathrm{S}_\lambda$ تمثل مصفوفة التمهيد من الدرجة $(n \times n)$ وتحسب وفق الصيغة الاتية

$$S_\lambda = (I + \lambda K)^{-1} \qquad \ldots \ldots \ldots \ldots \ldots \ldots . \ldots (14)$$

حيث يتم حساب هذه المصفوفة $S_\lambda$ وفق الصيغة اعلاه اعتماداً على معلمة التمهيد $\lambda$ وعلى نقاط العقد $(t_1, t_2 \ldots \ldots t_n)$ .

وباستعمال طريقة المربعات الصغرى يتم تقدير متجه المعلمات $\beta$ للجزء المعلمي في نموذج الانحدار شبه المعلمي ووفق الصيغة الآتية:

$$\varepsilon'\varepsilon = (Y - X\beta - g)'(Y - X\beta - g)$$

$$\frac{\partial \varepsilon' \varepsilon}{\partial \beta'} = -X'(Y - X\beta - g)$$

وبجعل $\frac{\partial \varepsilon'\varepsilon}{\partial \beta'}|_{\beta=\hat{\beta}} = 0$ فان الحل $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)'$ هو تقدير المربعات الصغرى ويكون كما يلي :

$$\hat{\beta} = (X'X)^{-1}X'Y^*$$

$$Y^* = Y - g$$

وبالتعويض عن $\beta$ بـ $\hat{\beta}$ في المعادلة رقم (1) لنموذج (PLM) فنحصل على:

$$Y_i^{**} = Y_i - X\,'\hat{\beta}$$

$$Y_i^{**} = \mathrm{g}(t_i) + \varepsilon_i \qquad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . \ldots (15)$$

وباستعمال طريقة (Cubic Smoothing Spline) نحصل على تقدير للجزء اللامعلمي (الدالة $\mathrm{g}(t)$ ) في نموذج (PLM) ويكون كالاتي :

$$\hat{g} = \mathrm{S}_\lambda\big(Y - X\hat{\beta}\big) \qquad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . . \ldots (16)$$

### 3-2 اختيار المعلمة التمهيدية[13,12]

استخدمنا في هذا البحث معيار العبور الشرعي المعمم لاختيار المعلمة التمهيدية $GCV$ (Generalized Cross Validation) ووفق الصيغة الاتية :

$$GCV_{\lambda} = \sum_{i=1}^{n} (y_i - \hat{g}(t_i;\ \lambda)^2 / (1 - n^{-1} tr(S_{\lambda})^2$$

$tr(S_{\lambda})$ : تمثل مجموع عناصر القطر الرئيسي لمصفوفة التمهيد .

وبالتالي يتم اختيار قيمة المعلمة التمهيدية $\lambda$ التي تعطي اصغر قيمة للمعيار $(GCV)$أي ان

$$h_{GCV} = \arg\min GCV_{(\lambda)}$$

### 3- الجانب التجريبي

### 1-3 المقدمة

لغرض تطبيق طرائق تقدير النماذج شبه المعلمية التي تم التطرق اليها في الجانب النظري تم اللجوء الى اسلوب المحاكاة (Simulation) والذي يعد تقليداً للواقع من خلال استعمال نموذج رياضي محدد للنظام الحقيقي الذي قد يصعب وضعه ، وايضاً لابد من وضع بعض الفروض الضرورية للحصول على تحليل اكثر شمولية من خلال استعمال عدد كبير من العينات وباحجام مختلفة او اختيار قيم مختلفة لتباين الخطأ .

### 2-3 توليد المتغيرا ت العشوائية

تمت عملية تنفيذ تجارب المحاكاة باستعمال اربعة احجام للعينة $(n = 10,50,100,150)$ وبتكرار (replicate=500) لكل تجربة محاكاة . وتوليد متغيرات توضيحية تتبع التوزيع الطبيعي القياسي .

3-3 الاخطاء العشوائية وحجوم العينات

تم توليد الاخطاء العشوائية تتبع التوزيع الطبيعي بمتوسط صفر وتباين ثابت $\sigma^2$ ، وقد تم افتراض ثلاث مستويات لقيم تباين الخطأ وهي (0.25,0.5,0.75) وتم تناول اربعة احجام للعينة$(n = 10,50,100,150)$ .

### 4-3 النماذج المستعملة في تجارب المحاكاة

تم تناول نموذجين من دوال الاختبار وكما يلي

$$1 -\ g(t) = 0.5\sin(2\pi t)$$

$$2 -\ g(t) = 1 - t + exp(-200(t - 0.5)^2$$

جدول (1) يبين معدل قيم متوسط مربعات الخطأ لطرق التقدير شبه المعلمية للنموذج الاول

| Model | N | $\sigma^2$ | L.P | C.S.S |
|---|---|---|---|---|
| 1 | 10 | 0.25 | 0.06300023 | 0.011006304 |
| | | 0.5 | 0.06440579 | 0.01383334 |
| | | 0.75 | 0.06937782 | 0.01870536 |
| | 50 | 0.25 | 0.06311191 | 0.010994810 |
| | | 0.5 | 0.06489809 | 0.01354913 |
| | | 0.75 | 0.06777605 | 0.01697365 |
| | 100 | 0.25 | 0.06317212 | 0.009805388 |
| | | 0.5 | 0.06486348 | 0.01167300 |
| | | 0.75 | 0.06797348 | 0.01602960 |
| | 150 | 0.25 | 0.06311607 | 0.009832413 |
| | | 0.5 | 0.06492750 | 0.01162979 |
| | | 0.75 | 0.06799796 | 0.01490421 |

جدول (2) يبين معدل قيم متوسط مربعات الخطأ لطرق التقدير شبه المعلمية للنموذج الثاني

| Model | N | $\sigma^2$ | L.P | C.S.S |
|---|---|---|---|---|
| 2 | 10 | 0.25 | 0.06317590 | 0.026804699 |
| | | 0.5 | 0.06479370 | 0.02833062 |
| | | 0.75 | 0.06793182 | 0.03208815 |
| | 50 | 0.25 | 0.063088548 | 0.027348564 |
| | | 0.5 | 0.06475945 | 0.028876555 |
| | | 0.75 | 0.06799441 | 0.03365421 |
| | 100 | 0.25 | 0.06306852 | 0.02682400 |
| | | 0.5 | 0.06486816 | 0.028520846 |
| | | 0.75 | 0.06801144 | 0.03205541 |
| | 150 | 0.25 | 0.06306598 | 0.02714706 |
| | | 0.5 | 0.06495793 | 0.02871869 |
| | | 0.75 | 0.06809486 | 0.03192303 |

## 4- الاستنتاجات

### 1-4 النموذج الاول

- عند تقدير النموذج شبه المعلمي وباستخدام الطريقتان أظهرت النتائج لجميع قيم احجام العينات والتباينات المستعملة ان مقدر (C.S.S) هو أفضل من مقدر (L.P) .
- كما بينت النتائج ان هناك تذبذب قليل في (Average Mean Square Error) لجميع المقدرات وعند كافة احجام العينات وباختلاف التباينات.
- ان بعض النتائج لمعيار الاختبار تزداد بزيادة حجم العينة وبزيادة التباين والبعض الآخر يقل والعكس ايضاً .

### 2-4 النموذج الثاني

- أفرزت النتائج لكافة قيم احجام العينات ولجميع مستويات التباين ان مقدر (C.S.S) كذلك هو أفضل من مقدر (L.P) .
- اما مقدر (C.S.S) فان قيمة AMSE تكون متذبذبة .
- ولمقدر (L.P) نلاحظ ان قيمة AMSE ترتفع عند حجم العينة الواحدة بزيادة مقدار التباين . ولكن تكون قيمة المعيار متذبذبة عند زيادة حجم العينة .

## 5- التوصيات

- يفضل استعمال مقدر (C.S.S) عند تقدير نموذج الانحدار الخطي الجزئي شبه المعلمي عند تطبيق دالة الانحدار اللامعلمية الاولى والثانية .
- ضرورة تقدير دالة الانحدار شبه المعلمي في حالة عدم تجانس تباين الخطأ مع التطبيق .
- ينبغي استعمال طرائق تقدير اخرى لنموذج الانحدار شبه المعلمي غير المذكورة في هذا البحث .
- يفضل تقدير نموذج الانحدار شبه المعلمي عندما تكون هناك قيم مفقودة في المتغيرات التوضيحية ومتغير الاستجابة يكون تام المشاهدة .

## المصادر

1- متي،نور صباح و الصفاوي ، صفاء يونس .تقدير دوال الانحدار اللامعلمي باستخدام بعض اساليب التمهيد . المجلة العراقية للعلوم الاحصائية(20)،عدد خاص بوقائع المؤتمر العلمي الرابع لكلية علوم الحاسبات والرياضيات.ص373-392,2011.

2- عبد الحسين علي حسون،علي عزيز علي .الرياضيات العالية .الطبعة الاولى – دار الكتب للطباعة والنشر، 1981.

3- حسين،شيرين علي. مقدرات الامكان الاعظم الموزونة الحصينة ومقارنتها مع طرائق اخرى لانموذج اللوجستك مع تطبيق عملي. رسالة ماجستير في الاحصاء،كلية الادارة والاقتصاد،جامعة بغداد،2009 .

4.P.Speckman, Kernel Smoothing in partially Linear Models.Journal of Royal Statistical Soc. 50, No.3,pp. 413-436,1988.

5.D.Millimet J.List, and T.Stengos. The environmental Kuznets curve. Real progress or misspecified models, Rev Econ Stat (85)4 .pp1038–1047,2003.

6-P.J.Green and B.W.Silverman . Nonparametric regression and generalized linear models : A roughness penalty approach.Chapman and Hall, London,1994.

7-W.Hardle and M.Muller. Multivariate and Semiparametric Kernel Regression. in M. G. Schimek (ed),Smoothing and regression.Approaches,Computation and Application,Wiley,2000.

8-D.Aydin.Partially Linear Models Based on Smoothing Spline Estimated by Different Selection Methods: A Simulation Study. Department of Statistics, Faculty of Arts and Sciences, Mugla University,2011.

9.A.Yatchew.semiparametric regression for the applied econometrician.Cambridge university press,2003.

10.D.Ruppert, M.P.Wand and R.J.Carroll.Semiparametric Regression.Cambridgeuniversity Press,New York,2003.

11-D.Aydin.Acomparison of the nonparametric regrestion model using smoothing spline and kernel regression .international journal of mathematical.physical and engineering scinces –vol .2 ,N0.2, pp.75-79,1999.

12-D.Aydin and M.Tuzemen. Estimation in Semi-parametric and Aditive Regression using Smoothing and Regression Spline. Second International Conference on Computer Research and Development, computer society ,2010.

13-D.Ruppert. Selecting the Number of Knots for Penalized Splines. Journal of Computational and Graphical Statistics, Vol.11, N0. 4, PP735–757,2002.

14-R.L.Eubank,E.Kambour,K.Klippe,C.Reese and M.schimek

.Estimation in Partially Linear Models. computational Statistics & data analysis 29.pp27-34,1998.

15- W.Hurdle. Applied Nonparametric Regression. Oxford University press,oxford,1990.

محمد طالب / محمد الشاروط

# A comparison of some semi-parametric Estimators For Partial Linear Regression Model by using simulation

**Abstract:**

the purpose of this paper is to compare between some of methods of semiparametric model by using simulation to estimate the best model for simulation data . the method which is used to estimate the parametric part is (OLS) while Nonparametric part was estimated by using local polynomial smoother and Cubic smoothing Spline .The simulation results shows that the semi-parametric model estimated by using the Cubic smoothing spline is the best one .