

## ”المقدر الحصين لنموذج تجميعي معمّم شبه معلمي“

أ. د. ظافر حسين رشيد  
م. د. عمر عبد المحسن علي  
جامعة بغداد/ كلية الادارة والاقتصاد – قسم الاحصاء

### الخلاصة

النموذج التجميعي المعمّم GAM من الممهدات متعددة المتغيرات حديثة الاستعمال في تحليل الانحدار اللامعلمي، ولذا تم تكريس هذا البحث لدراسته ولكن بالصيغة الهجينة، أي عند للظاهرة التي تغير سلوكها من خطي ذات شكل دالي مسبق (معلمي)، ولاخطي بشكل غير معلوم (لامعلمي) والذي سيكون هنا الشريحة التمهيدية. هذا، بالإضافة الى تقديم مقترح لأيجاد مقدر GAM شبه معلمي حصين ومقارنته بأسلوب تكراري وآخر غير تكراري.

### Abstract

Generalized Additive Model has been considered as a multivariate smoother that appeared recently in Nonparametric Regression Analysis. Thus, this research is devoted to study the mixed situation, i.e. for the phenomena that changes its behaviour from linear (with known functional form) represented in parametric part, to nonlinear (with unknown functional form: here, smoothing spline) represented in nonparametric part of the model. Furthermore, we propose robust semiparametric GAM estimator, which compared with two other existed techniques.

### 1.1 المقدمة وهدف البحث

يعدّ النموذج التجميعي المعمّم Generalized Additive Model: GAM ممهداً معقولاً بل وجيداً عند حالة وجود أكثر من متغير توضيحي واحد في تحليل الانحدار اللامعلمي. إذ سيمثل ممهداً شاملاً Global Smoother يجمع ضمناً التأثيرات الرئيسية Main effects للمتغيرات التوضيحية والممثلة بممهدات موضعية Local Smoothers أحادية المتغير وبصورة تجميعية Additively.

فلحالة الانحدار متعدد المتغيرات، ويكون هناك تغير في سلوك الظاهرة (الدالة) المدروسة بأسلوب معين، أو تكون هناك معلمات مجتزأة عن خطية البيانات، أو معرفة غير أكيدة عن الصيغة الدالية المسبقة لجزء من نموذج (سطح) الانحدار المزمع استخدامه لتحليل الظاهرة، فسيكون استعمال نموذجاً هجيناً من معلمي ولا معلمي، أمراً ممكناً بل وضروري كأحد الحلول العملية الملائمة لهكذا حالات، وهو ما يدعى بنموذج الانحدار شبه المعلمي.

وبالإستناد الى ماتقدم، يمكن استعمال ما يدعى بـ GAM شبه المعلمي، إذ سيتكون من مجموع جزء دالي وبمعلمات معروفة، وجزء حرّ يتمثل بممهد معين، وسيكون لهذا البحث عبارة عن الشريحة التمهيدية. وبهذا نجمع بين القوة الإستدلالية للجزء المعلمي من حيث توفر صفة الخطية المهمة، و بين المرونة العالية في تقدير الجزء اللامعلمي من حيث أفضل توفيق للبيانات عن طريق التمهيد.

ويهدف البحث الى مقارنة الطرائق التكرارية وغير التكرارية لإيجاد مقدر GAM شبه معلمي، فضلاً عن إقتراح طريقة حصينة في حالة الأخطاء العشوائية بتوزيع ملوث.

## 1.2 النموذج التجميعي المعمم شبه المعلمي Semiparametric GAM

قد تظهر في بعض الأحيان، شكوك حول تجانس التباينات في حالات تتوزع زمنياً أو مكانياً. أو أن الظاهرة تغير من سلوكها بشكل يدعو الى التساؤل عن جدوى إستعمال نموذجاً معلمياً أو لامعلمي بعينه. ولذا يكون إقتراح نموذجاً هجيناً يجمع مابين النموذجين أمراً لا بد منه. فالنموذج [8], [13]:

$$Y_i = X_i' \beta + g(t_i) + \text{error} \quad \dots \quad (1)$$

إذ سيبدو الأمر كأن الجزء المعلمي  $X_i' \beta$  يسلك سلوك "مقطع" Intercept أو عتبة Threshold تتباين وتتغير بمقدار  $g(t_i)$  اللامعلمي [10], [9], [3]. ويعتبر النموذج (1) نقطة إلتقاء GLIM المعلمي مع GAM اللامعلمي، ولذا يمكن تسميته بـ "GAM شبه معلمي" كذلك. وعند إستعمال الطرائق الشرائحية في تقدير الجزء اللامعلمي منه، عندئذ سيدعى النموذج (1) بنموذج "الشرائح الجزئية" Partial Splines Model. مع العلم إن التعميم لهذا النموذج يدعى نموذج "الشرائح الهجينة" Hybrid Splines Model.

## 2.2 طرائق تقدير GAM شبه المعلمي

ومن أبرز هذه الطرائق لتقدير (1)، منها ما يأتي في الفقرات اللاحقة.

### 1.2.2 أسلوب Backfitting التكراري

عندما يتم إفتراض أن جميع المركبات الداخلة في GAM هي خطية ماعدا واحدة فقط، فعندئذ يتم إعتبار أن الـ Backfitting عملية تكرارية لممهدين فقط هما، الأول: بشكل مسقطي هو:  $S_1 = X(X'X)^{-1}X'$  والذي ينتج عنه تقدير OLS أي  $(X\hat{\beta})$  لواحد أو أكثر من المتغيرات التنبؤية  $X's$  لتمثيل مصفوفة التصميم  $X$  ذات الرتبة الكاملة. وأما الآخر: فهو الممهد  $S_2$  والذي ينتج عنه تقدير الشريحة التمهيدية  $\hat{g}$ . فيمكن مطابقة النموذج (1)، وذلك بتقدير  $\beta$  و  $g$  من خلال تصغير المربعات الصغرى الجزائية الموزونة WPLS، وكالاتي:

$$S(\beta, g) = \sum_{i=1}^n \{Y_i - X_i' \beta - g(t_i)\}^2 + \lambda \int g''(t) dt \quad \dots \quad (2)$$

وعندما تكون  $t_i$  جميعها مختلفة Distinct - أي لا يوجد "ties" - فسيكون حل المعادلات الآتية، كالاتي:

$$(3) \begin{bmatrix} X'X & X' \\ X & I + \alpha K \end{bmatrix} \begin{bmatrix} \beta \\ g \end{bmatrix} = \begin{bmatrix} X' \\ I \end{bmatrix} Y$$

وعند حذف الجزء المعلمي  $X$  و  $\beta$  من المعادلة (3)، نحصل على:

$$(I + \lambda K)g = Y$$

وسيطهر الممهده للبيانات Y للقيم المطابقة g، بالصيغة:

$$g = S(Y - X\beta) \quad \dots (4)$$

$$S = (I + \lambda K)^{-1}$$

ويمكن كتابة (3) بشكل معادلتين آتيتين<sup>[5]</sup>:

$$X'X\beta = X'(Y - g) \quad \dots (5)$$

$$(I + \lambda K)g = (Y - X\beta) \quad \dots (6)$$

ويكون:

$$\hat{g} = (I + \lambda K)^{-1} (Y - X\hat{\beta}) \quad \dots (7)$$

$$\hat{\beta} = (X'X)^{-1} X'(Y - \hat{g}) \quad \dots (8)$$

إذ أثبت [Heckman , 1986]<sup>[5]</sup>، أن تقدير  $\hat{\beta}$  بهذا الأسلوب يكون تقديراً متسقاً Consistent. فلو كانت g معلومة، فإن المعادلة (5) يمكنها تقدير  $\beta$  بالـ WLS للفروق  $[y_i - (Ng)_i]$ . أما لو كانت  $\beta$  معلومة، فإن المعادلة (6) يمكنها تقدير g بمطابقة للشريحة التمهيدية التكعيبية للفروق  $[y_i - x_i'\beta]$ . وهنا يأتي دور الـ Backfitting بعمل تقدير تناوبي مابين (5) و (6). وهذا يمكن عمله دائماً طالما كانت القيم الذاتية للمصفوفة (S) ذات قيم ذاتية أقل من الواحد، أي أن  $|S| < 1$ . ومن المفضل أن تبدأ عملية التقدير بقيم  $\beta$  صفرية. أما إختيار المعلمة التمهيدية الداخلة في تقدير دالة g فيمكن إيجادها بطرق عديدة. فالحالة الاعتيادية هي استعمال معيار GCV الاعتيادي وبالصيغة:

$$\min OGCV(\lambda) = \frac{\frac{1}{n} \sum_{i=1}^n \{y_i - X\beta - g_\lambda(t_i)\}^2}{\left[1 - \frac{\text{trace}(S(\lambda))}{n}\right]^2} \quad \dots (9)$$

فقيمة  $\lambda$  المثلى هي التي تحقق (9).

ولكن مع هذا قد استعمل الباحثان أسلوباً إضافياً لإيجاد قيمة المعلمة التمهيدية بافتراض قيمتها مسبقاً، أسوة بالحالة اللامعلمية للـ backfitting، أي الأسلوب غير الآلي (الشخصي) كأسلوب آخر للتعامل مع تقدير المعلمة التمهيدية.

## 2.2.2 طريقة غير تكرارية

هنا لا تكون الحاجة الى تكرارات لعملية التقدير يتم تحديثها في إطار تحسين المطابقة، وإنما يتم الحل بشكل مباشر عن طريق ماتدعي بالشريحة التمهيدية الجزئية Partial Smoothing Spline لتمثيل حل عام لنظام معادلات خطي. فعندما تكون القيمة الذاتية الكبرى من القيم الذاتية للـ "S" قريبة من الواحد، فسيكون تقارب backfitting بطيء جداً. ولذا كان لابد من إيجاد أسلوب بديل لحل معادلات (3) وبـ O(n) من العمليات الحسابية أيضاً، وبدون الحاجة لإعادة التكرار. فعند تعويض (4) في (5)، يتم الحصول على نظام خطي P×P بالنسبة للـ  $\beta$ ، ويكون:

$$X'X\beta = X'(Y - S(Y - X\beta)) \quad \dots (10)$$

$$X'(I - S)X\beta = X'(I - S)Y \quad \dots (11)$$

...(12)

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

$$\hat{g} = S(Y - X\hat{\beta})$$

...(13)

وبذا يمكن إستعمال (GLS) لحل المعادلات الطبيعية، بحيث أن:

$$\Omega = (I - S)$$

$\Omega$  : مصفوفة أوزان غير قطرية.

فعلى الرغم من أن  $S$  مصفوفة تمهيد من درجة  $n \times n$ ، فكل مايلزم فعله هو القيام بتمهيد كل عمود من أعمدة  $X$  البالغة  $p$ . وسيتم إنجاز هذه العملية بـ  $O(np)$  من العمليات الحسابية. وهذا ما يمكن عمله مع  $X'(I-S)Y$ . وهذا ماسيدو أسلوباً بسيطاً لتحليل التباين المشترك اللامعلمي **Nonparametric Covariance Analysis**. ومن اللافت للنظر، أن المصفوفة  $S$  بهيئة مصفوفة وزن للإحدار على  $X$ . وسيكون الحل وحيداً **Unique** ومتسقاً **Consistent** طالما وجدت معكوسة المصفوفة  $X'(I-S)X$ . في حين أن إستعمال أسلوب آخر، مثل تجزئة جولسكي **Cheolsky Decomposition** لنظام  $p \times p$  كهذا قد يحتاج الى  $O(n^2p)$  من العمليات. ومع هذا فيجب الحذر عند التعامل مع هذه الطريقة لأنه لا يمكن تضمينها أساليب تجزئة متعامدة لتلافي الحصول على مصفوفة  $p \times p$  وهي:  $X'(I-S)X$ ، وبالتالي قد تكون هناك فرصة لظهور ما يدعى بخطأ التدوير **Rounding Error**.

### 3.2.2 مقترح الشرائح الجزئية الحصينة **Robust Partial Spline**

إذ يتم تمثيل البواقي التي سيحتسب الوزن فيما بعد طبقاً لها، كالآتي:

$$r_i = \frac{y_i - x\beta - g}{\hat{\sigma}} \quad ; \quad i=1,2, \dots, n \quad \dots (14)$$

لتصغير المقدار:

$$\sum_{i=1}^n \hat{\sigma}^2 \rho \left\{ \frac{y_i - x\beta - g}{\hat{\sigma}} \right\} + \frac{1}{2} \lambda \int_{[0,1]} g''(t)^2 dt$$

ليصبح المقدر:

$$\hat{g} = (W + \lambda K)^{-1} W(Y - X\hat{\beta}) \quad \dots (15)$$

$$W = \text{diag}(w_i)$$

بأوزان قطرية:

$$w_i = \frac{\psi(r_i)}{r_i} \quad \dots (16)$$

K : مصفوفة جزاء تربيعية متماثلة ومن درجة  $n \times n$  :

$$\begin{aligned} K &= QR^{-1}Q' \\ \int g''(t)^2 dt &= \gamma' Q' g \\ &= \gamma' R \gamma \end{aligned} \quad \dots (17)$$

إذ أن:

باستعمال الشرط  $Q'g = R\gamma$ ، يتم الحصول على [1], [5], [3]:

$$\begin{aligned} \int g''(t)^2 dt &= g' QR^{-1} Q' g \\ &= g' K g \end{aligned}$$

h: يمثل متجه المسافات (الفروق) بين أي المشاهدين  $t_i$  و  $t_{i-1}$ ، وعناصره  $(i^{th})$ ، كالآتي:

$$h_i = t_i - t_{i-1} \quad ; \quad i=2,3, \dots, n$$

γ: متجه المشتقات الثانية للدوال  $(g_i)$ ، ومن درجة  $1 \times (n-2)$ .Q: مصفوفة tridiagonal بدرجة  $(n-2) \times n$ . ويتم احتساب عناصرها كالآتي [3], [5].

$$\text{where: } \left. \begin{aligned} q(j-1, j) &= h^{-1}(j-1) \\ q(j, j) &= -h^{-1}(j-1) - h^{-1}(j) \\ q(j+1, j) &= h^{-1}(j) \\ q(i, j) &= 0, \forall |i-j| \geq 2 \end{aligned} \right\}$$

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

R: مصفوفة tridiagonal متماثلة وبدرجة  $(n-2) \times (n-2)$ . ويتم احتساب عناصرها كالآتي [3], [5].

$$\text{where: } \left. \begin{aligned} r(i, i) &= (1/3)[h(i-1) + h(i)]; i = 2,3, \dots, n-1 \\ r(i, i+1) &= r(i+1, i) = (1/6)h(i); i = 2,3, \dots, n-2 \\ r(i, j) &= 0; \forall |i-j| \geq 2 \end{aligned} \right\}$$

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

### 3.2 مقترح استعمال الأسلوب الشخصي

يتم استعمال الأسلوب الشخصي إختيار المعلمة التمهيدية في حالة ممد واحد، بدلاً من الأسلوب الآلي المؤلف في مثل هذه الحالات<sup>[1]</sup>. إذ يمكن عمل ذلك عن طريق تثبيت قيمتها مسبقاً، قبل إجراء عملية تقدير GAM شبه المعلمي، في النموذج (1). فللمعادلة (7)، يمكن اعتبار  $\lambda$  بقيمة ثابتة، أي أن:

$$\lambda = \frac{\rho}{1-\rho} \quad \dots (18)$$

where:  $\rho \in [0,1]$

### 1.3 الجانب التجريبي

تم تنفيذ تجارب المحاكاة بلغة MATLAB 7.5 البرمجية وباستخدام ثلاثة أحجام للعينات وكما يأتي: (n=50)، (n=100) و (n=200). وب تكرارات (Replicates=300) لكل تجربة. فالمتغيرات التوضيحية هي الـ  $t_j$  يتوزعان توزيع منتظم قياسي مستقل لكل منهما، أي:  $t_j \sim U(0,1)$

والأخطاء العشوائية تتوزع توزيعاً طبيعياً، أي:  $e_i \sim N(0, \sigma^2)$  ;  $j=1,2$  ;  $i=1,2, \dots, n$

يتم توليد الأخطاء العشوائية باستعمال طريقة Box-Muller. وقد تم تناول ثلاث مستويات من التباين على أساس S.N.R.<sup>[14]</sup>:

تباين عالي High Noise :  $\sigma=(1/2)*\text{Function Range}$   
 تباين متوسط Medium Noise :  $\sigma=(1/4)*\text{Function Range}$   
 تباين واطيء Low Noise :  $\sigma=(1/8)*\text{Function Range}$

إذ أن  $\sigma$  يمثل الإنحراف المعياري للخطأ e. أما التوزيع الآخر للخطأ العشوائي فهو التوزيع الملوث، ويستعمل في حالة تلوث البيانات بقيم شاذة. إذ تم توليد بيانات متغيرات تتبع التوزيع الملوث عن طريق توليد متغيرات تتبع التوزيع الطبيعي بالأسلوب السابق ذاته بنسبة  $(1-\zeta)$ ، وتوليد متغيرات التوزيع الطبيعي ذاته لكن بتباين مختلف وبنسبة  $\zeta$ ، إذ أن:  $0 \leq \zeta \leq 1$  ، أي أن:

$$e \sim [(1-\zeta) * N(0, \sigma^2), \zeta * N(0, \gamma^2 \sigma^2)] ; \gamma^2 > 1 \quad \dots (19)$$

وعلى هذا الأساس سيتم توليد المتغير  $Y_i$  مباشرة من خلال النماذج المستخدمة في تجارب المحاكاة، وذلك باستخدام دوال الإنحدار بدلالة المتغيرين التوضيحين اللذان تم توليدهما في الفقرة (I) أعلاه، مضافاً إليهما الأخطاء العشوائية بأحد أشكالها سواء الملوث منها أم غير الملوث والتي تم توليدها في الفقرة (II)، ولكل نموذج من النماذج قيد البحث.

### 2.3 النماذج المستعملة في المحاكاة

تم إستعمال صيغة شبه المعلمية مشتقة من النماذج أدناه بحيث تكون :

$$Y_i = X\beta + g_2(t_{2i}) + e_i \quad \dots (20)$$

$$X = \begin{bmatrix} 1 & t_1 \end{bmatrix} \quad \beta = \begin{bmatrix} b0 \\ b1 \end{bmatrix} \quad ;$$

وبذا كانت المركبة الثانية (اللامعلمية) من كل نموذج كما يأتي:

#### 1. للنموذج الأول [11]:

$$g_2(t_{2i}) = t_{2i}^3$$

إذ أن هذه الدالة من النوع **Spatially homogenous**.

$$g_2(t_{2i}) = \begin{cases} -2t_{2i} & \text{if } t_{2i} < 0.6 \\ -1.2 & \text{if } t_{2i} \geq 0.6 \end{cases} \quad \dots [6] \quad \text{2. للنموذج الثاني}$$

صيغة خطية لدالة تغير سلوكها الى صيغة ثابتة عندما  $(t_{2i} \geq 0.6)$ .

### 3.3 معايير مقارنة النماذج

هناك عدد من المعايير لقياس مقدار الجودة في تقدير النماذج، مع الأخذ بنظر الإعتبار تنوع النماذج (الظواهر) التي يراد توظيف الشرائح التمهيدية لتمثيلها. ومن هذه المعايير (والتي يتم إختيار الأصغر منها عندما يراد قياس الأفضلية):

#### 1. معيار AMSE [12]

ويمثل معدل متوسط مربعات الخطأ [Average of Mean Squared Error]، وبالصيغة:

$$AMSE = n^{-1} \sum_{i=1}^n E\{\eta_i - \hat{\eta}_i\}^2 \quad \dots (21)$$

#### 2. معيار GCV [4]

من المعروف عن هذا المعيار إنه غالباً ما يستعمل في تحديد قيمة (أو قيم) المعلمة (أو المعلمات) التمهيدية، ضمن ما يعرف بالأساليب الآلية [6] المعدة لهذا الغرض، والمبينة في المعادلة (9). وهو يعتمد بصورة أساسية على درجات الحرية المكافئة **EDF**.

فللمهد  $t_j^{\text{th}}$  فإن درجة حريته تكون:  $df_j = \text{trace}(S_j) - 1$  ;  $j = 1, 2, \dots, p$  وبالتالي يحقق نتائجاً أقل تحيزاً من **MSE** [2]، وكما يأتي:

... (22)

$$OGCV = \frac{\frac{1}{n} \sum_{i=1}^n \{y_i - X\beta - g(t_i)\}^2}{\left[1 - \frac{\text{trace}(S)}{n}\right]^2}$$

#### 3. معيار AMDSE [11]

ويمثل معدل وسيط مربعات الخطأ [Average of Median Squared Error] وبالصيغة:

$$AMDSE = E\left(\text{Median}\{\eta_i - \hat{\eta}_i\}^2\right) \quad (23)$$

#### 4. معيار Concurvity [7]

وهو معيار يبين فيهما معرفة قوة مشكلة الـ Concurvity، وبالصيغة:

$$\text{Concurvity} = E(\text{Corr}(\hat{Y}, \hat{Y}_{new})) \quad \dots (24)$$

$$\hat{Y}_{new} = \hat{Y} + N(0, \sigma^2)$$

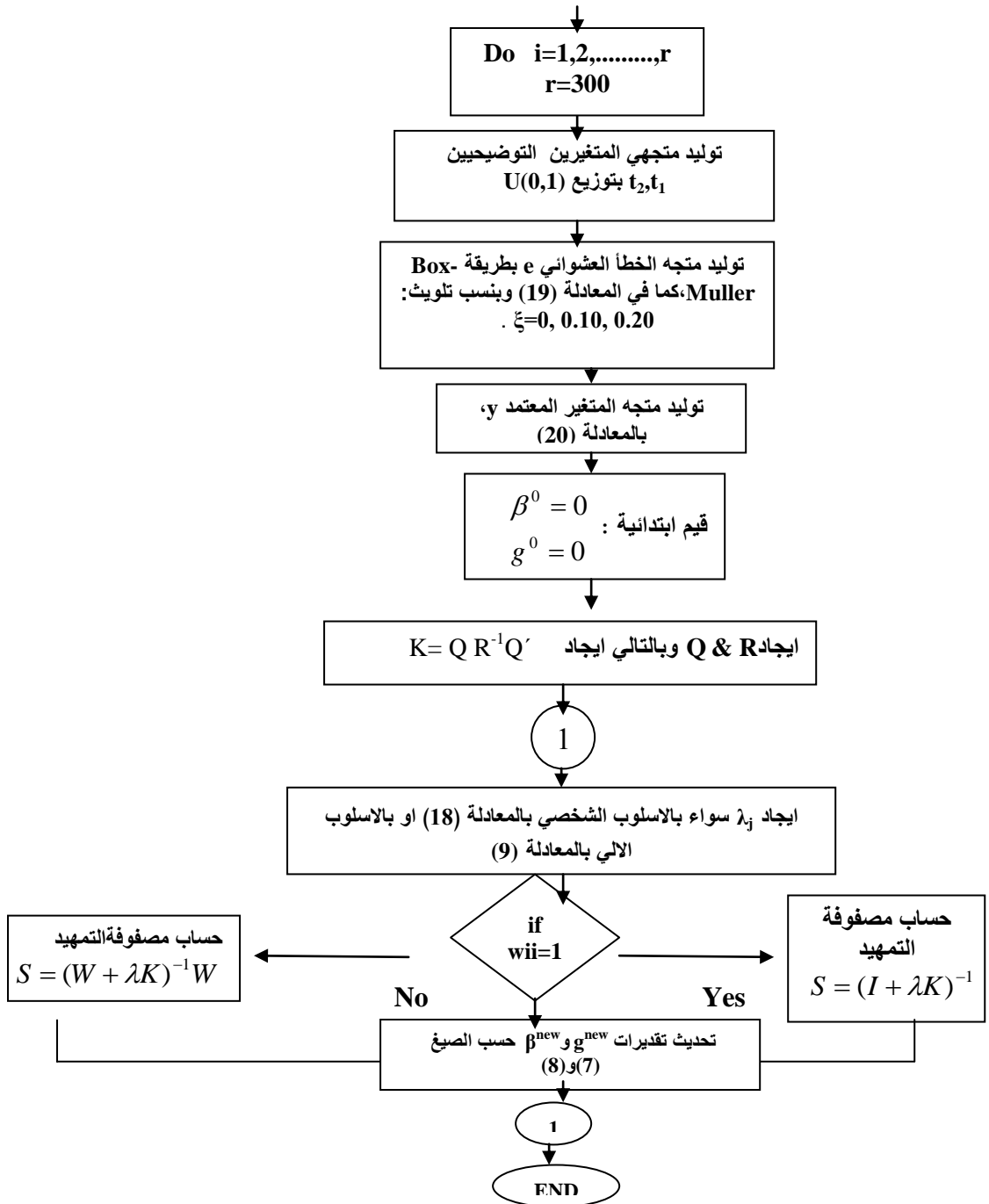
### 4.3 تنفيذ تجارب المحاكاة

- لكلا النموذجين الواردين في الفقرة (2.3) أعلاه، تم القيام بما يأتي :
1. توليد المتغيرات: توليد المتغير  $t_2$  التوضيحي ليتوزع توزيعاً منتظماً قياسيً  $U(0,1)$ . مع توليد أخطاء عشوائية تتبع توزيع Gaussian، ناهيك عن توزيع ملوث للخطأ العشوائي وهو التوزيع الملوث  $N(0,4)$ ، وبنسب  $\xi=0, 0.10, 0.20$ ، وبحسب المعادلة (19).
  2. طريقة Backfitting: تستعمل الصيغة (9) متمثلة بالمعيار الاعتيادي OGCV لإختيار معلمة واحدة، ومن ثم تطبيق المعادلتين (7) و(8) معاً، لإيجاد تقدير GAM شبه المعلمي.
  3. الطريقة المباشرة: تستعمل الصيغة (9) متمثلة بالمعيار الاعتيادي OGCV عند إختيار معلمة تمهيدية واحدة، ومن ثم تطبيق المعادلتين (12) و(13) معاً، لإيجاد تقدير GAM شبه المعلمي.
  4. إقتراح: تم إقتراح إستعمال الأسلوب الشخصي لإختيار معلمة تمهيدية واحدة كما في الصيغة (18)، وإيجاد تقدير GAM شبه المعلمي من المعادلتين (7) و (8) معاً، وبأسلوب Backfitting كمحاولة لإختزال وقت تنفيذ المحاكاة مما يؤدي الى فائدة عملية.
  6. إقتراح الشريحة الجزئية الحصينة: إختيار المعلمة التمهدية بالأسلوب الآلي الاعتيادي وبمعيار OGCV كما في المعادلة (9)، وإيجاد مقدر GAM بتوظيف طريقة M الحصينة كما في المعادلة (15).

مخطط (1): أسلوب Backfitting لتقدير GAM شبه المعلمي الحصين وغير الحصين

Start





## 4.3 النتائج

1. للمودج الأول: فمن نتائج الجدول (1) يلاحظ تفوق أسلوب **backfitting** عند أحجام العينات  $(n=100,200)$  وللمعيارين **AMSE** و **AMDSE** ولكل أحجام العينات عند المعيارين **GCV** و **Concurvity**. ويتضح أن أسلوب **PLS with OGCV** كان له الأفضلية بالمعيارين **AMSE** و **AMDSE** عند حجم العينة  $(n=50)$ . وأن المعيار **Concurvity** يسجل إنخفاضاً واضحاً مع زيادة مقابلة في حجم العينة. أما بقية المعايير فهي تنخفض مع زيادة حجم العينة أيضاً ولكن مع  $(\sigma=1/2)$ ، وتزداد فيما عداها. ويلاحظ كذلك التقارب والتشابه بين نتائج طريقتي **backfitting** و **Direct** عموماً. ولحالة التلوث، لم تتقدم الطريقة الحصينة المقترحة على مثيلاتها غير الحصينة، كما هو مبين في الجدول (2).

2. للمودج الثاني: من نتائج الجدول (3) يلاحظ تفوق أسلوب **backfitting** وللمعيارين **GCV** و **Concurvity** ولكل أحجام العينات عند المعيارين. ويتضح أن أسلوب **PLS with OGCV** كان له الأفضلية بالمعيارين **AMSE** و **AMDSE** عند كل أحجام العينات. وأن المعيار **Concurvity** يسجل إنخفاضاً واضحاً مع زيادة مقابلة في حجم العينة. أما معيار **AMSE** فينخفض مع زيادة حجم العينة أيضاً ولكن مع  $(\sigma=1/2)$ ، وبقية المعايير تزداد فيما عداها. ويلاحظ كذلك التقارب والتشابه بين نتائج طريقتي **backfitting** و **Direct** عموماً. أما حالة التلوث فقد أبدت الطريقة الحصينة المقترحة أفضلية، ولجميع أحجام العينات، كما هو مبين في الجدول (4).

## 1.4 الإستنتاجات

- يمكن إدراج الإستنتاجات الخاصة بالبحث على ضوء نتائج التحليل للتجارب المقامة بالإستناد الى النموذجين قيد البحث، وكما يأتي:
1. إستطاعت الطريقة الحصينة أن تتبنى النموذج الثاني عند جميع أحجام العينات، في حين أخفقت الطريقة مع النموذج الأول ولجميع أحجام العينات.
2. تبدأ قيم المعايير - ماعدا **Concurvity** - بالتقارب عند مستويات التباين المتوسطة والواطنة عند تثبيت حجم العينة.
3. تبينت أمثلية إستعمال الأسلوب المباشر لتحليل **GAM** شبه المعلمي، عند حجم العينة  $(n=50)$  وذلك لسرعة الإنجاز وإختزال الوقت اللازم للتحليل، مقابل بعض التضحية الطفيفة بأفضلية هذه النتائج مقارنة مع الطريقتين الأخرين. وتكون أفضل طريقة للإستعمال في الحالة شبه المعلمية هي طريقة **Backfitting** المقترحة لحجم العينة  $(n=100)$  في حين يفضل إستعمال أسلوب **OGCV** عند حجم العينة  $(n=200)$ .

## المصادر العربية:

1. علي، عمر عبد المحسن؛ (2007)؛ "مقارنة مقدرات النماذج التجميعية المعممة باستخدام الشرائح التمهيدية عند تحليل الإنحدار اللامعلمي وشبه المعلمي"؛ أطروحة دكتوراه في الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد.

## المصادر الأجنبية:

2. Buja, A.; Hastie, T.J. and Tibshirani, R.J.; (1989); "Linear smoothing and additive models" ( with discussion ); Annals of Statistics; Vol.17; pp. 453-555.
3. Canotoni, E.; and Ronchetti, E.; (2001); "Robust inference for Generalized Linear Models"; JASA, Vol.96, No. 455, pp. 1022-1030.
4. Eubank, R.L.; (1988); "Spline smoothing and nonparametric regression "; Marcel Dekker, New-York.
5. Green; P.J. and Silverman ; B.W.; (1994); "Nonparametric regression and generalized linear models: A roughness penalty approach"; Chapman and Hall, London.
6. Hastie, T.J and Tibshirani, R.J.; (1990); "Generalized additive models"; Chapman and Hall, London .
7. He, Shui; (2004); "Generalized additive models for data with concavity: statistical issues and a novel model fitting approach"; Ph.D. Dissertation, School of public health, University of Pittsburgh.
8. Kauermann, G.; (2005); "Non- and Semiparametric Models and their estimation"; Physica-Verlag 0, ISSN 0002-6018. <http://www.uni-bielefeld.de/~kauermann>.
9. Liu, A.; Meiring, W.; and Wang, Y.; (2005); "Testing generalized linear models using smoothing spline methods "; Statistica Sinica; Vol. 15, pp. 235-256.
10. Osullivan, F; Yandell, B.S; and Raynor, W.J.; (1986); "Automatic smoothing for regression functions in generalized linear models"; JASA; Vol.81; No393, pp.96 -103.
11. Ruppert, D.; and Carroll, R.J.; (2000); "Spatially-Adaptive Penalties for Spline Fitting "; Australian & New-Zealand J.Statist., Vol.42, No.2, pp. 205-223.
12. Wahba, G; ( 1990); "Spline models for observational Data"; SIAM: Society for Industrial and Applied Mathematics Philadelphia, PA.
13. Wang, M.; (1989); "Semiparametric Model for Randomly Truncated Data"; JASA, Vol.84, No.407, pp. 742-748.
14. Wood, S.; Kohn, R.; Shively, T.; Jiang, W.; (2002); "Model Selection in Spline Nonparametric Regression"; J.R.Statist.Soc.B., Vol. 64, Part1, pp. 119-139.