المجلة العراقية للعلوم الإحصائية

www.stats.mosuljournals.com

توظيف تقنية التقليص المويجي في تقدير نموذج الإنحدار الجمعي اللامعلمي المعمم (WGAM)، دراسة مقارنة مع المحاكاة والتطبيق

الاء عبد الستار داوٗد حمودات و بشار عبد العزيز الطالب

قسم الاحصاء والمعلوماتية ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل ، الموصل ، العراق

معلومات النشر

تاريخ المقالة:
تم استلامه في 13 كانون الثاني 2020
تم القبول في 18 اذار 2020
متاح على الإنترنت في 1 كانون الاول 2020

الكلمات الدالة:
طريقة تقدير النموذج الجمعي المعمم اللامعلمية ،التمهيدية
استخدام تقنية التقليص المويجي

المراسلة:
الاء عبد الستار داوٗد حمودات
allahamoodat@uomosul.edu.iq

الخلاصة

تم في هذا البحث تناول مشكلة عدم معلومية التوزيع الاحتمالي للبيانات وتوظيف طريقة تقدير النموذج الجمعي المعمم Generalized Additive Models (GAM) اللامعلمية المستندة إلى الشرائح التمهيدية Smoothing Splines كممهدات والتعامل مع هذه الحالة بالأسلوب التكراري . وتم استخدام تقنية التقليص المويجي Wavelet Shrinkage والاعتماد عليها في تقدير نموذج الإنحدار Wavelet Generalized Additive Models (WGAM) والتي تم إقتراح توظيفها كممهد للبيانات وذلك من خلال استخدام بعض المويجات كمرشحات في حساب التحويل المتقطع للمويجة، وقد تم الاعتماد على بعض المعايير الاحصائية لمقارنة طرق التقدير، وذلك من خلال توظيف اسلوب المحاكاة وتحليل بيانات حقيقية، وقد تم إختبار كفاءة الطريقة المقترحة على بيانات تم جمعها من مستشفى ابن سينا التعليمي، على حالات مصابة بقصر القامة، وقد اعطت مرشحات تقليص المويجة افضل النتائج مقارنة بطريقة GAM الاعتيادية وساعدت المويجة على تمهيد البيانات وذلك من خلال الحصول على أكفأ النتائج.

DOI: 10.33899/IQJOSS.2020.167385 , ©Authors, 2020, College of Computer Science and Mathematics, University of Mosul.
This is an open access article under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

1.المقدمة Introduction

يهدف البحث إلى توظيف تقنية المويجة الصغيرة (Wavelet) لفلترة وتمهيد البيانات بإستخدام بعض أنواع المويجات كمرشحات في حساب التحويل المتقطع للمويجة وإستخدام البيانات الممهدة لتقدير النموذج الجمعي المعمم (WGAM) والحكم على كفاءة النتائج بإستخدام بعض معايير المقارنة الإحصائية لتحديد أسلوب التقدير الأفضل من بين نموذجي GAM وأسلوب WGAM المقترح بأسلوب المحاكاة ولبيانات الواقعية.

2.النماذج الجمعية المعممة Generalized Additive Models (GAM)

تعد النماذج الجمعية المعممة (GAM) احد مداخل الانحدار اللامعلمي وهي نماذج اكثر تقييداً والفرض الوحيد الذي تقوم عليه عملية التقدير هذه هو أن دوال التقدير تكون جمعية ومتغيراتها ممهدة، ويعتبر الانموذج الجمعي المعمم GAM اداة فعالة لتحليل البيانات وتمتلك القدرة على التعامل مع العلاقات غير الخطية الشديدة بين المتغير المعتمد والمتغيرات المستقلة وهو اسلوب مقترح من قبل (Friedman & Jorome, 1999) و (Werner Stuetzle) عام 1981 . وهذا النموذج يخلط بين خصائص النماذج الخطية المعممة (GLM) Generalized Linear Models والنماذج الجمعية (Additive Models AM) كما بين (Friedman & Jorome, 1999) وإن دراسة الانحدار اللامعلمي المتعدد بشكل عام ستفرز عنه مشكلة البعدية (زيادة ابعاد النموذج) والتي يعاني منها الباحثين حيث تقيدهم نحو تعميم حالة احادي المتغيرات الى حالة متعدد المتغيرات والتي ستتعامل مع الحالة الاعم للشرائح التمهيدية وهي Thin Plate Spline (TPS). فعند ايجاد مقدار تقاربي لبعدين او اكثر يصعب عمل مصفوفة المتغيرات المقاسة بوحدات مختلفة وغيرها من المشاكل، لذلك ظهرت الحاجة

لظهور النموذج الجمعي كحل عملي ، حيث يعمل بصفة تجميعية تساعد في سهولة تفسير الظواهر المختلفة، وبصورة عامة فإن النموذج الجمعي المعمم يعد إمتداداً شبه معلمياً للنموذج الخطي المعمم. وعليه فان النموذج الجمعي اللامعلمي Non-Parametric Additive Model يكون كالاتي (He, 2004) :-

$$Y_i = \alpha + \sum_{i=1}^{n} m_i(x_i) + \varepsilon_i \qquad (1)$$

m_i : متجه دوال مقدرة من البيانات والتي تعبر عن المتغير التوضيحي i والتي يمكن تقديرها عن طريق الشرائح التمهيدية (Smoothing Spline) وهنالك عدة طرق للحصول على تقدير للنموذج الجمعي المعمم GAM ومن أهمها خوارزمية التوفيق التراجعي.

3.خوارزمية التوفيق التراجعي Backfitting Algorithm

ان اول من قدم هذه الطريقة هما الباحثان (Friedman & Stuetzle) عام 1981 ، حيث اعتمدت هذه الطريقة على الاسلوب التكراري ، وان النموذج الجمعي يفترض دالة التوقع الشرطي لمتغير الاستجابة y والتي يمكن ان تكتب كمجموع الحدود الممهدة للمتغيرات $x_1, x_2, \ldots, x_1$ ، ولمطابقة المعادلة (1) سيكون التوقع الشرطي لكل K من الـ $x's$ كالاتي (Friedman & Jorome, 1999) :

$$E(Y/X = x_1, x_2, \ldots, x_k) = m(x_1, x_2, \ldots, x_k) = m_1(x_1) + m_2(x_2) + \cdots + m_k(x_k)$$

الخطوة الاولى لتقدير هذه المعادلة هو البدء مع قيم اولية كخطوة ابتدائية (°) في عملية تقدير تكرارية:

$$m_i = m_i^{\circ} \quad , \quad i = 1,2, \ldots, p$$

الخطوة الثانية هو القيام بالدورة :

$$\hat{m}_i = S_i\left(y - \alpha - \sum_{k=i}^{p} m_k/x_i\right) \quad , \quad k, i = 1,2, \ldots, p \qquad (2)$$

اذ ان S_i : تمثل مصفوفة تحويل ممهدة $(n \times n)$.

$$S_i = (I + \lambda_i K_i)^{-1}$$
$$K_i = Q_i R_i^{-1} Q_i'$$

يعاد إيجاد K_i لكل i من المتغيرات التوضيحية. ويتم تكرار هذه العمليات لإيجاد التقديرات الجديدة للدوال بواسطة تمهيد الاخطاء حتى تتقارب الدوال الجزئية ، ويتم التوقف عندما تستقر الدوال الممهدة $m_i(x_i)$ بدون تغيير . وحسب معيار التقارب (Buja et. al., 1989) :

$$\Delta\left(m_i^d, m_i^{d-1}\right) = \sum_{i=1}^{p}\left(m_i^d, m_i^{d-1}\right)^2$$

وتعرف البواقي الجزئية بأنها القيم المتطابقة لكل دالة مضافاً اليها البواقي الاجمالية من النموذج التجميعي بشرط أن: $E\left(m_i(x_i)\right) = 0$ وان مصفوفة التمهيد S لا تعتمد على قيم متغير الاستجابة y بل تعتمد على قيم المتغيرات التوضيحية $x_i's$ وهذا يشابه الحالة المعلمية حيث ان :

$$\hat{\beta} = Hy \quad , \quad H = X(X'X)^{-1}X'$$

حيث يتم استبدال مصفوفة التمهيد S_i بمصفوفة تمهيد اخرى S_i^* حيث تقوم بعملية التمهيد اولاً ثم تقوم بحذف متوسط التمهيد للحصول على تقديرات يمكن التعرف عليها Identifiable للتقديرات وهذا ما يسمى بالممهد الممركز Centered Smoother بحيث أن:

$$S_i^* = (I - 11')S_i$$

حيث ان 1 عبارة عن متجه (nx1) كل عناصره واحد . ويتم الحصول على تقدير للنموذج الجمعي المعمم GAM، والذي يوظف لتصغير المقدار التالي:

$$\sum_{i=1}^{n}(y_i - a - m(x_i))^2 + \sum_{i=1}^{n}\lambda \int_a^b m''(x_i)^2 \, dx \qquad (3)$$

ويلاحظ من المعادلة (3) بأنها تكون جزائية بثابت منفصل λ_i لكل حد ، ويمكن كتابة المعادلة (3) بصيغة المصفوفات وكالاتي :

$$(y_i - a - \sum_{i=1}^{n} m_i)'(y_i - a - \sum_{i=1}^{n} m_i) + \sum_{i=1}^{n} \lambda_i \, m_i' \, k_i m_i \qquad (4)$$

k_i : تمثل مصفوفة جزاء لكل . وبأخذ التفاضل للمعادلة (4) بالنسبة m_i ومساواتها بالصفر نحصل على:

$$\hat{m}_i = S_k(y - a - \sum_{i \neq k}^{n} \hat{m}_i) \qquad (5)$$

ويمثل المقدار $(y - a - \sum_{i\neq k}^{n} \widehat{m}_i)$ البواقي الجزئية من التمهيد ، وان $S_k = (I + \lambda_k \mathrm{K}_k)^{-1}$ ، ومن ثم يتم تكرار او اعادة عملية التمهيد للبواقي الجزئية .

ويتم اختيار المعلمة التمهيدية (Smoothing Parametric) (λ) بطريقة الاسلوب المجرد (الالي) Objective (Automatic) Method، اذ تترك البيانات لتحدد بذاتها قيمة المعلمة التمهيدية (المثلى) عن طريق تصغير معيار معين يعتمد بدوره على البواقي ، ومن اشهر المعايير لاختيار المعلمة التمهيدية هو (Generalized Cross Validation GCV(λ)) والذي تم إعتماده كمعيار رئيسي للمقارنة في هذا البحث (Green and Siverman,1994).

المويجة wavelet

المويجة (Wavelet) او ما تسمى بالموجة الصغيرة (Small Wave) هي احد انواع الدوال الرياضية تتيح إمكانية تجزئة الدالة المعطاة الى مركبات تردد مختلفة ودراسة كل مركب مع اعادة التصميم Resolution الملائم عند كل قياس ، وتعرف المويجة رياضياً بانها دالة قيمة حقيقية معرفة على محور حقيقي كامل وتتذبذب صعوداً ونزولاً بشكل منتظم حول الصفر (Crowley et. al. ,2006) . وسميت بالمويجة لصغرها وتميزها عن اشارة الموجة الكبيرة (Big wave) مثل موجة دالة الجيب (Sine) وموجة الجيب تمام (cosine) . وتتكون المويجات عادة من جزئيين رئيسين هما (Kumar & Lavanya, 2011):

1- **دالة القياس (Scaling Function)**

$\emptyset(.)$ التي تعرف ايضا بدالة الاب (Father Function) وتعتبر الجزء التقريبي للبيانات ، ويتم الحصول عليها من خلال الصيغة الاتية :-

$$\mathrm{f}(x) = \sum_{\mathrm{k}=0}^{\mathrm{N}} C(\mathrm{k})\mathrm{f}(2x - \mathrm{k}) \qquad (6)$$

حيث ان :-

$C(\mathrm{k})$: تمثل معاملات مرشح التمرير الواطئ Low-Pass Filter .

2- **دالة المويجة Wavelet Function** $\Psi(.)$ والتي تعرف ايضاً بدالة الام Mother function والتي تمثل معادلة المويجة Wavelet Equation ، صيغتها كالاتي :-

$$\mathrm{w}(x) = \sum_{\mathrm{k}=0}^{\mathrm{N}} \mathrm{d}(\mathrm{k})\mathrm{f}(2x - \mathrm{k}) \qquad (7)$$

حيث ان $\mathrm{d}(\mathrm{k})$ يمثل مرشح التمرير العالي High-Pass Filter .

4. بعض دوال المويجة Some Wavelet Functions فيما يلي بعض الدوال المويجية شائعة الاستخدام :

اولاً – مويجة هار Haar wavelet

درست المويجة Haar من قبل العالم الرياضي (Alfred Haar) في الفترة (1909–1910)، وتعرف المويجة الام لـ هار (Haar Mother Wavelet) كالآتي:-

$$\Psi(\mathrm{X}) = \begin{cases} 1 & \text{if} \quad 0 < \mathrm{X} \leq \frac{1}{2} \\ -1 & \text{if} \quad \frac{1}{2} < \mathrm{X} \leq 1 \\ 0 & \text{if} \quad \text{o.w} \end{cases}$$

اما دالة القياس للمويجة الاب لها (Haar Father Wavelet) فهي تعرف كالاتي :- (Green , 1994)

$$\emptyset(\mathrm{X}) = \begin{cases} 1 & \text{if} \quad 0 \leq \mathrm{X} \leq 1 \\ 0 & \text{if} \quad \text{o.w} \end{cases}$$

ثانيا– المويجة دوبشيز Daubechies wavelet

سميت هذه المويجة نسبة الى الباحثة Ingrid Daubechies ، وتكتب مويجات هذه العائلة اختصاراً (dbLi) او (DN) حيث ان N تمثل رتبة المرشح في حين ان Li هو عدد العزوم المتلاشية والذي يعني تقارب Convergence مفكوك المويجة بحيث ان :

$$\sum_{k=0}^{n-1}(-1)^k k^v C(k) = 0 \qquad , \qquad v = 0,1,2, \dots, \frac{n}{2} - 1$$

$C(k)$: تمثل معاملات مرشح التمرير الواطئ (Low-Pass Filter) . (Crowley et. al. ,2006)

ثالثاً–المويجة (كويفليتز) Coiflets Wavelet

بناءاً على طلب قدمه الباحث Coifman في 1989 اوجدت الباحثة Daubechies المويجة كويفليتز ونسب اليه ، حيث قام الباحث بطرح فكرة الحصول على العزوم المتلاشية او الزائلة لمرشحات التمرير الواطئ ومرشحات التمرير العالي معاً لكلتا الدالتين الام والاب . (Rajeev et. al., 2011)

رابعاً المويجة ساملت The Lest Asymmetric Daubechies Symmetric Wavelet (Symmlets)

المويجة ساملت هي مويجة متعامدة قريبة من التماثل اقترحتها الباحثة Daubechies حيث اجرت تعديلات على عائلة (db) بزيادة التماثل مع بقاء بساطة المويجة وهي متماثلة ولها نفس خصائص (دويجيز) (Zhang et. al. , 2016)

5.قطع العتبة الناعمة Soft Thresholding

ان الخطوة الثانية من خطوات تقدير دالة المويجة هي ازالة التشويش الموجودة في الاشارة عن طريق حد العتبة ، وباستخدام التحويل المويجي يتم وضع عتبة ترددية بحيث تلغي هذه العتبة معاملات التشويش وتحافظ على معاملات الاشارة الاصلية K واحد انواع قطع العتبة هو قطع العتبة الناعمة ويستخدم لإزالة التشويش في الاشارة والتي قدمها كل من الباحثين (Donoho & Johnstone) ، حيث ان قطع العتبة الناعمة تدفع كل المعاملات باتجاه الصفر أي يتم فيها انهاء القيم ما دون العتبة الى الصفر والمحافظة على القيم الاعلى من العتبة ، وتعرف قطع العتبة الناعمة بالشكل الاتي (Rajeev et. al., 2011) :

$$W_n^{st} = \text{sign}\,(W_n)\,(|Wn| - \delta)_+ \quad (8)$$

$$\text{sign}\,(W_n) = \begin{cases} +1 & \text{if} \quad Wn > 0 \\ 0 & \text{if} \quad Wn = 0 \\ -1 & \text{if} \quad Wn < 0 \end{cases}$$

$$(|W_n| - \delta)_+ = \begin{cases} 0 & \text{if} \quad (|W_n| - \delta) < 0 \\ (|W_n| - \delta) & \text{if} \quad \text{o.w} \end{cases}$$

6.معايير المقارنة Comparison Criteria's

هنالك العديد من المعايير تقيس مقدار الجودة في تقدير دوال الانحدار اللامعلميه، ومنها:

1- معيار العبور الشرعي المعمم Generalized Cross Validation (GCV)

وضع معيار Generalized Cross Validation (GCV) من قبل الباحث Wahba Graven عام 1979 ، و يستخدم معيار العبور الشرعي GCV في اتجاهين الاول ضمن سياق اختيار انسب معلمة تمهيد لمقدر الشريحة الممهدة ، اما الاتجاه الثاني فإستخدامه للمقارنة بين طرائق تقديرات الانحدار اللامعلمي، وتأخذ صيغته الرياضية الشكل التالي (Roppert , 2003) :

$$\text{GCV} = \frac{n^{-1}\sum_{i=1}^{n}\left\{y_j - a - \sum_{j=1}^{p}\widehat{m}_j(x_i,\lambda)\right\}^2}{(1-\text{tr}(S_\lambda)/n)^2} = \frac{n^{-1}\sum_{i=1}^{n}\{y_j - \hat{y}_i\}^2}{(1-\text{tr}(S_\lambda)/n)^2} \quad (9)$$

2- معيار Akaike Information Criterion (AIC) وضع من قبل Akaike عام 1973 ،

والصيغة العامة لمعيار (AIC) تكون كما يلي:

$$AIC_\lambda = -2Ln\,(Lik) + 2df \quad (10)$$

df: درجة حرية التمهيد ويعبر عنها بأثر مصفوفة التمهيد S_λ ، $df = \text{tr}(S_\lambda)$. (Ulrich,2001)

3- معيار Bayesian Information Criterion (BIC)

الصيغة العامة لمعيار (BIC) تكون كالاتي :- (Li & Ruppert, 2008)

$$BIC = -2Ln(Lik) + \text{Ln}(n)df \quad (11)$$

4- معيار الإعتمادية غير الخطية Concurvity

وهو معيار يبين تعميم لمشكلة تعدد العلاقة الخطية (التي من الممكن أن تحدث في النموذج الخطي العام GLM) بين المتغيرات المستقلة ولكن هذه المرة في النموذج الجمعي العام والذي قد يؤدي بالنتيجة إلى الحصول على مقدرات غير مستقرة وهذا المعيار يعبر عن الإعتمادية غير الخطية، وكما أن تعدد العلاقة

الخطية ينتج عنها تضخم في التباين فإن مشكلة الإعتمادية غير الخطية (Concurvity) ينتج عنها معاملات مقدرة غير مستقرة (Amodio et. al., 2015). ويمكن التعبير رياضياً عن هذه المشكلة كما هو موضح في الصيغة الاتية : (He , 2004)

$$concurvity = E\big(Corr.(\hat{y}.\hat{y}_{new})\big) \quad (12)$$
$$\hat{y}_{new} = \hat{y} + N(0,\sigma^2)$$

7.تجارب المحاكاة Simulation trials

دوال التوليد المستخدمة Used Generation Functions

تتنوع الدوال بتنوع الظاهر التي تمثلها ، حيث تتميز هذه الدوال كونها صممت لتعرض مجموعة من الظواهر غالباً ما تحدث في واقع الحياة، وقد تم توظيف دالتين معتمدتين في أغلب البحوث العالمية وهما:-

1- دالية خطية من الدرجات العليا :(Augustin et. al., 2012)

$f_1(x) = 0.2*x\text{^}11 * (10 * (1 - x))\text{^}6 + 10*(10*x)\text{^}3*(1-x)\text{^}10$

2- دالة Doppler (Donoho & Johnstone, 1994)

$$f_2(x) = \{x(1-x)\}^{1/2} sin\{2\pi(1+\varepsilon)/(x+\varepsilon)\} \quad , \quad \varepsilon = 0.05$$

خوارزمية تجارب المحاكاة Simulation Trials Algorithm

تم تطبيق العديد من السيناريوهات لتجارب المحاكاة ، اذ تم تلويث المتغيرات التوضيحية تارة وتلويث y تارة اخرة بمختلف التوزيعات ((t). , Laplace , Exp.) وبمعلمات مختلفة ، ولأحجام عينة مختلفة (300 , 150 , 50) ولنموذج بأربعة متغيرات وبنسب تلويث (35%, 15% , 5%) وذلك لتوليد البيانات الخاصة بالمتغيرات العشوائية ، ثم تكرار كل تجربة (Replicates=100) مرة لغرض الحصول على نتائج متسقة ، ولإعطاء صورة شاملة عن كفاءة الطرق تم اختيار معلمات مختلفة للتوزيعات الاحتمالية وكما يلي :

1- توليد اربع متغيرات توضيحية تتوزع توزيع منتظم قياسي مستقل .
2- توليد الاخطاء العشوائية من توزيع طبيعي بوسط حسابي يساوي صفر وتباين معين .
3- توليد المتغير العشوائي y مباشرة من خلال النموذج المستخدم في تجارب المحاكاة ، وذلك باستخدام دوال الانحدار بدلالة المتغيرات التوضيحية التي تم توليدها اعلاه ، مضافاً اليهما الخطأ العشوائي .
4- تقدير النموذج الجمعي المعمم GAM الاعتيادي ، ومن ثم تمهيد البيانات بدوال المويجة (Db, Haar, Least A. , Coiflets) ومن ثم تقدير النموذج الجمعي المعمم (WGAM) وذلك للحصول على المقدرات الممهدة (DWGAM, HWGAM, LAWGAM, CWGAM).
5- المقارنة بين GAM و WGAM للمقدرات الممهدة الأربعة في النقطة (4) وذلك من خلال معايير المقارنة (GCV ,Con. , BIC , AIC) .

نتائج تجارب المحاكاة Simulation Results

تم وضع اربع نماذج من الجداول وهي (4, 5 , 6 , 7 , 8) وذلك لمحدودية المساحة في البحث أما بقية الجداول فهي متوفرة لدى الباحثين ولنسب تلويث واحجام عينات وتوزيعات احتمالية مختلفة، وعليه وللاختصار تم انتخاب اربعة جداول بشكل عشوائي لعرضها ومقارنة أداء أساليب التقدير وستتم المناقشة داخل الجداول.

جدول (1) : المقارنة مع اسلوب GAM الاعتيادي للنموذج الاول وعند تلويث y بتوزيع t وبنسب التلويث واحجام عينات مختلفة

تلويث متغير الاستجابة y بتوزيع (t)	معايير المقارنة	GCV	Con.	AIC	BIC	GCV	Con.	AIC	BIC	GCV	Con.	AIC	BIC
	نسب التلويث n	5%				15%				35%			
50	GAM	2666.2	2.78E-28	533.07	540.3	2290	1.03E-28	529.4	536.5	1667	1.74E-27	512.8	521.1
	DWGAM	2516.7	2.44E-29	266.75	271.8	2240	3.06E-29	264.0	268.8	1686	3.32E-29	256.9	261.4
	HWGAM	2569.9	7.40E-29	267.44	272.0	2283	9.75E-29	264.7	269.5	1740	1.00E-28	257.4	262.0
	LAWGAM	2527.1	1.22E-28	266.93	271.4	2242	2.82E-28	264.2	268.9	1696	1.01E-27	257.0	261.5
	CWGAM	2544.2	2.39E-28	266.99	271.4	2284	1.19E-28	265.6	268.9	1688	8.09E-28	258.2	261.6
		نلاحظ عند نسبة تلويث 5% وحجم عينة 50 ان افضل طريقة كانت أسلوب التمهيد بدالة بويسجز (DWGAM) رغم ان معيار BIC كانت قيمته اصغر لمقدري LAWGAM ومقدر CWGAM لكن قيمة بقية المعايير هي الأصغر لمقدر DWGAM مقارنة ببقية تقديرات التقليص المويجي				نلاحظ عند نسبة تلويث 15% وحجم عينة 50 ان افضل طريقة كانت أسلوب التمهيد بدالة بويسجز (DWGAM)				نلاحظ عند نسبة تلويث 35% وحجم عينة 50 ان افضل طريقة كانت أسلوب التمهيد بدالة بويسجز (DWGAM) رغم ان معيار GCV كانت قيمته اصغر لمقدر GAM ولكن قيمة هذا المعيار هو الأصغر لمقدر DWGAM مقارنة ببقية تقديرات التقليص المويجي			
150	GAM	2378.5	4.48E-28	1592.8	1605.	2498	5.19E-28	1600.	1612.	1651	1.26E-27	1537.	1550.
	DWGAM	2456.8	9.77E-29	799.17	807.8	2531	8.09E-29	801.4	810.0	1619	1.04E-29	771.0	780.2
	HWGAM	2411.5	1.51E-28	797.79	806.7	2530	8.57E-29	801.7	810.3	1729	6.32E-28	771.2	780.2
	LAWGAM	2442.8	4.97E-28	798.76	807.9	2530	8.17E-29	801.5	810.2	1721	2.33E-28	771.6	780.3
	CWGAM	2435.7	1.83E-28	798.56	807.5	2530	1.35E-28	801.5	810.1	1716	2.85E-29	771.3	780.9
		نلاحظ عند نسبة تلويث 5% وحجم عينة 150 ان افضل طريقة كانت الأفضلية فيها لأسلوب التمهيد بدالة هار (HWGAM) ولكن النتيجة بالعموم غير حاسمة بشكل قطعي				نلاحظ عند نسبة تلويث 15% وحجم عينة 150 ان افضل طريقة كانت أسلوب التمهيد بدالة بويسجز (DWGAM) رغم ان معيار GCV كانت قيمته اصغر لمقدر GAM ولكن قيمة هذا المعيار هو الأصغر لمقدر DWGAM مقارنة ببقية تقديرات التقليص المويجي				نلاحظ عند نسبة تلويث 35% وحجم عينة 150 ان افضل طريقة كانت أسلوب التمهيد بدالة بويسجز (DWGAM)			
300	GAM	2481.0	2.51E-27	3187.1	3203.1	2138	2.50E-26	3152.	3169.	1692	1.02E-25	3078.1	3095
	DWGAM	2424.5	6.27E-28	1595.4	1606.8	2101	3.86E-28	1577	1587.2	1676	8.75E-28	1539.5	1550.6
	HWGAM	2428.8	1.86E-27	1595.1	1607.2	2106	4.95E-28	1578	1588.3	1683	6.49E-27	1540.3	1552.4
	LAWGAM	2428.5	4.65E-28	1595.6	1607.2	2115	1.58E-26	1596.9	1588.7	1677	1.60E-26	1539.9	1551.6
	CWGAM	2426.3	8.43E-28	1595.5	1607.1	2143	9.31E-28	1596.8	1588.6	1678	1.60E-26	1539.6	1551.6
		نلاحظ عند نسبة تلويث 5% وحجم عينة 300 عدم وضوح الأفضلية بين مقدرات التقليص المويجي ولكنها جميعاً افضل من تقدير GAM الاعتيادي				نلاحظ عند نسبة تلويث 15% وحجم عينة 300 ان افضل طريقة كانت أسلوب التمهيد بدالة بويسجز (DWGAM)				نلاحظ عند نسبة تلويث 35% وحجم عينة 300 ان افضل طريقة كانت أسلوب التمهيد بدالة بويسجز (DWGAM)			

وكخلاصة للجدول (1) فان اسلوب التمهيد بدوال المويجة الصغيرة قد ساهمت في تحسين كفاءة مقدرات النموذج الجمعي المعمم الاعتيادي GAM

جدول (2) : يمثل تقدير GAM وتقدير WGAM للنموذج الاول وعند حالة تلويث X بتوزيع Laplace وبنسب التلويث واحجام عينات مختلفة

تلويث متغير الاستجابة X بتوزيع (Laplace)	n	معايير المقارنة	GCV	Con.	AIC	BIC	GCV	Con.	AIC	BIC	GCV	Con.	AIC	BIC
		نسب التلويث	5%				15%				35%			
	50	GAM	811.3	3.73E-20	7641.1	7648.4	809.5	9.09E-19	8088	8094	775.2	7.77E-25	8440.8	8447
		DWGAM	767.1	1.64E-23	3638.3	3651.0	702.3	2.02E-22	3812	3825	733.8	3.13E-26	4165.2	4176
		HWGAM	562.1	3.30E-23	3544.0	3057.0	742.2	1.68E-25	3232	3245	631.0	4.10E-25	4005.3	4018
		LWGAM	704.9	9.18E-20	3507.9	3520.6	783.3	2.25E-21	3674	3687	711.1	1.89E-25	4121.6	4133
		CWGAM	551.6	6.19E-24	3402.9	3015.5	792.1	1.34E-20	3534	3547	788.1	2.17E-25	4111.8	4123
			نلاحظ عند نسبة تلويث 5% وحجم عينة 50 ان افضل طريقة كانت أسلوب التمهيد بدالة كويفلتز (CWGAM)				نلاحظ عند نسبة تلويث 15% وحجم عينة 50 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM) رغم ان معيار .Con لمقدر (DWGAM) كان اصغر ولكن بقية النتائج هي الأفضل مقارنة ببقية تقديرات التقليص المويجي				نلاحظ عند نسبة تلويث 35% وحجم عينة 50 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM) رغم ان معيار .Con لمقدر DWGAM كان اصغر ولكن بقية النتائج لمقدر HWGAM هي الأفضل مقارنة ببقية تقديرات التقليص المويجي		
	150	GAM	763.1	6.18E-20	23483	23495	771.0	5.22E-24	24921	24932	792.5	4.10E-28	25328	25340
		DWGAM	719.2	4.60E-22	11124	11150	762.7	4.04E-25	12297	12321	786.2	7.01E-26	12579	12601
		HWGAM	614.4	4.70E-24	9408.4	9433.4	755.9	4.53E-25	11982	12007	741.1	6.00E-26	12129	12453
		LWGAM	738.1	1.91E-21	10702	10727	766.2	1.82E-25	12220	12244	772.3	5.14E-26	12545	12567
		CWGAM	747.6	4.31E-22	10324	10349	766.2	4.13E-25	12196	12221	761.4	2.18E-26	12533	12554
			نلاحظ عند نسبة تلويث 5% وحجم عينة 150 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM)				نلاحظ عند نسبة تلويث 15% وحجم عينة 150 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM) رغم ان معيار .Con لمقدر (LWGAM) كان اصغر ولكن بقية النتائج هي الأفضل مقارنة ببقية تقديرات التقليص المويجي				نلاحظ عند نسبة تلويث 35% وحجم عينة 150 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM) رغم ان معيار .Con لمقدر كويفلتز (CWGAM) كان اصغر ولكن بقية النتائج هي الأفضل مقارنة ببقية تقديرات التقليص المويجي		
	300	GAM	774.3	4.39E-21	49416	49430	814.9	8.51E-28	51848	51862	802.4	7.58E-28	53253	53267
		DWGAM	762.1	2.76E-24	24128	24161	812.9	1.83E-25	25744	25775	796.9	4.64E-26	26518	26547
		HWGAM	732.4	4.43E-24	22965	22998	812.3	7.48E-26	25512	25544	781.5	7.09E-26	26380	26411
		LWGAM	751.4	1.03E-23	23946	23979	812.6	1.11E-25	25636	25668	795.4	3.42E-26	26461	26490
		CWGAM	737.0	3.34E-24	23801	23834	812.4	8.30E-26	25611	25642	794.9	4.25E-26	26432	26461
			نلاحظ عند نسبة تلويث 5% وحجم عينة 300 أن أفضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM) رغم أن معيار .Con لمقدر (DWGAM) كان اصغر ولكن بقية النتائج هي الأفضل مقارنة ببقية تقديرات التقليص المويجي				نلاحظ عند نسبة تلويث 15% وحجم عينة 300 أن أفضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM)				نلاحظ عند نسبة تلويث 35% وحجم عينة 300 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM) رغم ان معيار .Con لمقدر كويفلتز (LWGAM) كان اصغر ولكن بقية النتائج هي الأفضل مقارنة ببقية تقديرات التقليص المويجي		

جدول (3) : يمثل تقدير GAM وتقدير WGAM للنموذج الثاني وعند حالة تلويث y بتوزيع .Exp وبنسب تلويث واحجام عينات مختلفة

تلويث متغير الاستجابة y بتوزيع (Exp.)	n	معايير المقارنة	GCV	Con.	AIC	BIC	GCV	Con.	AIC	BIC	GCV	Con.	AIC	BIC
		نسب التلويث	5%				15%				35%			
	50	GAM	2689.9	2.62E-27	531.7	538.7	2439	3.50E-27	527.9	535.2	1859	8.43E-27	511.9	519.3
		DWGAM	2440.7	1.83E-28	266.1	270.8	2334	2.89E-28	265.1	269.6	1766	1.71E-29	257.0	261.7
		HWGAM	2518.3	2.94E-28	267.2	271.7	2317	1.55E-28	264.8	269.4	1777	3.26E-28	256.9	261.8
		LWGAM	2447.4	1.49E-28	266.4	271.0	2342	1.46E-28	265.2	269.7	1783	6.49E-29	257.3	261.9
		CWGAM	2457.4	5.63E-28	266.3	270.9	2335	3.18E-28	265.1	269.6	1767	9.79E-28	256.9	261.9
			نلاحظ عند نسبة تلويث 5% وحجم عينة 50 ان افضل طريقة كانت أسلوب التمهيد بدالة بويبجيز (DWGAM) رغم ان معيار .Con كانت قيمته اصغر لمقدر (LWGAM) ولكن قيمة بقية المعايير هي الأصغر لمقدر (DWGAM) مقارنة ببقية تقديرات التقليص المويجي				نلاحظ عند نسبة تلويث 15% وحجم عينة 50 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM) رغم ان معياري .Con و AIC لمقدري (DWGAM) وكان أصغر ولكن بقية النتائج هي الأفضل مقارنة ببقية تقديرات التقليص المويجي				نلاحظ عند نسبة تلويث 35% وحجم عينة 50 ان افضل طريقة كانت أسلوب التمهيد بدالة بويبجيز (DWGAM) رغم ان معيار AIC لمقدر (HWGAM) كان أصغر بفارق قليل ولكن بقية النتائج لمقدر (DWGAM) هي الأفضل مقارنة ببقية تقديرات التقليص المويجي		
	150	GAM	2408.9	4.50E-27	1594.	1606.	2286	2.69E-27	1582.	1593.	1875	2.33E-27	1539	1551
		DWGAM	2402.1	4.70E-29	797.6	805.9	2269	1.12E-28	792.9	801.1	1703	5.78E-28	771.0	779.9
		HWGAM	2414.1	1.54E-28	798.0	806.3	2262	8.74E-29	792.8	801.7	1702	1.35E-28	771.2	780.0
		LWGAM	2397.0	3.04E-28	797.7	806.3	2273	2.24E-28	793.0	802.5	1705	2.65E-28	771.2	780.5
		CWGAM	2397.6	1.76E-28	797.8	806.9	2278	3.28E-28	793.1	802.2	1704	2.74E-28	771.2	780.2
			نلاحظ عند نسبة تلويث 5% وحجم عينة 150 ان افضل طريقة كانت غير واضحة بشكل حاسم حيث أن أسلوب التمهيد بدالة بويبجيز (DWGAM) واسلوب دالة (LWGAM) كان لهما الأداء الأفضل مقارنة ببقية الطرق.				نلاحظ عند نسبة تلويث 15% وحجم عينة 150 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM) رغم ان معيار BIC لمقدر (DWGAM) كان اصغر ولكن بقية النتائج هي الأفضل مقارنة ببقية تقديرات التقليص المويجي				نلاحظ عند نسبة تلويث 35% وحجم عينة 150 ان أفضل طريقة كانت غير واضحة بشكل حاسم حيث أن اسلوب التمهيد بدالة بويبجيز (DWGAM) واسلوب دالة هار (HWGAM) كان لهما الأداء الأفضل مقارنة ببقية الطرق.		
	300	GAM	2381.4	2.18E-26	3184	3199	2390	4.49E-26	3159.	3172	1696	2.74E-25	3078.	3096
		DWGAM	2419.2	1.42E-27	1595	1606	2201	2.59E-27	1580.1	1591	1699	9.55E-28	1541	1552
		HWGAM	2400.1	7.99E-28	1594	1605	2201	2.41E-27	1580.2	1592	1683	4.55E-27	1539	1552
		LWGAM	2418.2	4.88E-27	1595	1607	2297	3.20E-28	1580.5	1593	1691	8.68E-28	1540	1553
		CWGAM	2425.8	8.60E-28	1595	1606	2291	5.95E-28	1580.4	1592	1688	3.84E-27	1540	1553
			نلاحظ عند نسبة تلويث 5% وحجم عينة 300 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM)				نلاحظ عند نسبة تلويث 15% وحجم عينة 300 ان افضل طريقة كانت أسلوب التمهيد بدالة بويبجيز (DWGAM) رغم ان معيار .Con لمقدر (LWGAM) كان اصغر ولكن بقية النتائج لمقدر (DWGAM) هي الأفضل مقارنة ببقية تقديرات التقليص المويجي				نلاحظ عند نسبة تلويث 35% وحجم عينة 300 ان أفضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM) رغم ان معيار .Con لمقدر (LWGAM) كان أصغر ولكن بقية النتائج لمقدر (HWGAM) هي الأفضل مقارنة ببقية تقديرات		

جدول (4) : يمثل تقدير GAM وتقدير WGAM للنموذج الثاني وعند حالة تلويث X بتوزيع t(7) وبنسب التلويث واحجام عينات مختلفة

تلويث المتغير المستقل X بتوزيع (t)	n	معايير المقارنة / نسب التلويث	GCV	Con.	AIC	BIC	GCV	Con.	AIC	BIC	GCV	Con.	AIC	BIC
			5%				15%				35%			
	50	GAM	2681.0	3.19E-26	535.50	542.7	2771.7	2.96E-27	535.0	542.2	2577	4.76E-27	1598	1610.
		DWGAM	2598.7	3.04E-28	267.58	272.3	2707.4	8.09E-28	268.5	273.3	2537	1.22E-28	801	810.4
		HWGAM	2690.4	3.96E-28	268.89	273.4	2736.3	3.72E-29	268.9	273.3	2518	4.02E-28	800	809.2
		LWGAM	2632.8	7.56E-28	268.22	272.7	2796.1	1.60E-27	268.6	273.0	2539	4.95E-28	801.8	810.0
		CWGAM	2635.4	4.35E-28	268.31	272.8	2765.8	1.13E-27	268.2	272.7	2538	3.35E-28	801.9	810.2
			نلاحظ عند نسبة تلويث 5% وحجم عينة 50 ان افضل طريقة كانت أسلوب التمهيد بدالة دوبيجيز (DWGAM)				نلاحظ عند نسبة تلويث 15% وحجم عينة 50 ان افضل طريقة لم يكن حاسماً تجاه أي من المقدرات ولكن مقدر كوبفيلتز (CWGAM) نتج عنه اصغر قيمتين لمعياري AIC و BIC مقارنة ببقية تقديرات التقليص الموجي				نلاحظ عند نسبة تلويث 35% وحجم عينة 50 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (DWGAM) رغم ان معيار Con. لمقدر (CWGAM) كان اصغر بفارق قليل ولكن بقية النتائج لمقدر (HWGAM) هي الأفضل مقارنة ببقية تقديرات التقليص الموجي		
	150	GAM	2762.2	4.46E-27	1603.6	1615.	2670.1	5.64E-28	1598.	1610	2677	4.76E-28	1598	1610.
		DWGAM	2611.3	7.95E-27	803.65	812.6	2506.5	3.89E-28	800.8	809.9	2537	1.22E-28	801	810.4
		HWGAM	2628.4	5.27E-28	804.05	812.6	2504.0	7.45E-28	800.8	809.6	2518	4.02E-28	800	809.2
		LWGAM	2613.3	1.11E-27	803.72	812.7	2510.2	2.95E-28	800.9	809.9	2539	4.95E-28	801	810.0
		CWGAM	2634.0	1.35E-27	803.79	812.7	2511.6	9.12E-28	800.9	810.1	2538	3.35E-28	801	810.2
	300		نلاحظ عند نسبة تلويث 5% وحجم عينة 150 ان افضل طريقة كانت أسلوب التمهيد بدالة دوبيجيز (DWGAM) رغم ان معيار Con. لمقدر التمهيد بدالة هار (HWGAM) كان أصغر بفارق قليل ولكن بقية النتائج لمقدر (DWGAM) هي الأفضل مقارنة ببقية تقديرات التقليص الموجي				نلاحظ عند نسبة تلويث 15% وحجم عينة 150 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM) رغم ان معيار Con. لمقدر التمهيد بدالة (LWGAM) كان أصغر بفارق قليل ولكن بقية النتائج لمقدر (HWGAM) هي الأفضل مقارنة ببقية تقديرات التقليص الموجي				نلاحظ عند نسبة تلويث 35% وحجم عينة 150 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM) رغم ان معيار Con. لمقدر التمهيد بدالة دوبيجيز (DWGAM) كان أصغر بفارق قليل ولكن بقية النتائج لمقدر (HWGAM) هي الأفضل مقارنة ببقية تقديرات التقليص الموجي		
		GAM	2687.7	4.74E-23	31980	3212.	2498.8	3.33E-28	3194.	3207.	2511	4.13E-27	3201	3215.
		DWGAM	2552.4	1.69E-24	1603.2	1614	2474.6	1.89E-28	1598	1609	2507	1.85E-28	1603	1613.4
		HWGAM	2526.5	1.50E-26	1601.7	1613	2460.1	3.48E-28	1597	1608	2500	5.56E-29	1602	1613
		LWGAM	2551.4	3.32E-24	1603.1	1614	2473.3	2.34E-28	1598	1609	2502	1.74E-28	1603	1613.1
		CWGAM	2552.7	6.45E-24	1603.2	1614	2470.2	3.60E-28	1598	1609	2507	1.63E-28	1603	1613.7
			نلاحظ عند نسبة تلويث 5% وحجم عينة 300 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM)				نلاحظ عند نسبة تلويث 15% وحجم عينة 300 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM) رغم ان معيار Con. لمقدر التمهيد بدالة دوبيجيز (DWGAM) كان أصغر بفارق قليل ولكن بقية النتائج لمقدر (HWGAM) هي الأفضل مقارنة ببقية تقديرات التقليص الموجي				نلاحظ عند نسبة تلويث 35% وحجم عينة 300 ان افضل طريقة كانت أسلوب التمهيد بدالة هار (HWGAM)			

جدول (5) : خلاصة نتائج المحاكاة للمقارنة بين طرق التقدير

معايير المقارنة / نسب التلويث	n	GCV	Con.	AIC	BIC	GCV	Con.	AIC	BIC	GCV	Con.	AIC	BIC	الخلاصة (الغالبية)
		5%				15%				35%				
y توزيع t	50	DWGAM				DWGAM				DWGAM				**DWGAM**
	150	HWGAM				DWGAM				DWGAM				**DWGAM**
	300	غير محسوم				DWGAM				DWGAM				**DWGAM**
X توزيع t	50	CWGAM				HWGAM				HWGAM				**HWGAM**
	150	HWGAM				HWGAM				HWGAM				**HWGAM**
	300	HWGAM				HWGAM				HWGAM				**HWGAM**
y توزيع Exp.	50	DWGAM				HWGAM				DWGAM				**DWGAM**
	150	DWGAM				HWGAM				DWGAM & HWGAM				**DWGAM HWGAM**
	300	HWGAM				DWGAM				HWGAM				**HWGAM**
X توزيع Exp.	50	DWGAM				CWGAM				DWGAM				**DWGAM**
	150	DWGAM				HWGAM				HWGAM				**HWGAM**
	300	HWGAM				HWGAM				HWGAM				**HWGAM**

وكحصيلة نهائية لتجارب المحاكاة ومن خلال الجدول (5) أعلاه نلاحظ بأن ال 36 تجربة محاكاة لتوزيعات t والتوزيع الأسي .Exp وتوزيع لابلاس ولنسب التلويث المختلفة فإن أسلوب التمهيد بدالة دوبيجيز (DWGAM) وأسلوب التمهيد بدالة هار (HWGAM) كان لهما الأداء الأفضل مقارنة ببقية الطرق لسيناريوهات المحاكاة التي تم تناولها.

الجانب التطبيقي

تناول الجانب التطبيقي مقارنة الطريقة المقترحة التي تقوم على توظيف ممهدات تقنية الموجة الصغيرة المستندة على الشرائح التمهيدية في تقدير وذلك للحصول على نموذج جمعي معمم مستند إلى تقنية التقليص الموجي والذي أشير له في هذا البحث إختصاراً (WGAM) ومقارنتها مع النموذج الجمعي المعمم الاعتيادي GAM بتطبيقهما على البيانات الأصلية والمرشحة والمؤلفة من تسعة من المتغيرات التوضيحية ومتغير استجابة واحد، وقد تم الاعتماد على البرنامج R لتطبيق الطريقة المقترحة الواردة في الجانب النظري.

جمع البيانات

طبقت هذه الدراسة على بيانات حقيقية تم جمعها من مستشفى ابن سينا التعليمي (مركز الوفاء التخصصي لأمراض السكري والغدد الصم استشارية قصار القامة) لمحافظة نينوى-2019 ، على حالات مصابة بقصر القامة ، وقد تم في هذا البحث جمع بيانات لـ 150 من الاشخاص المصابين بهذا المرض وهي عينة بحجم مناسب جداً لنموذج بتسعة متغيرات فقط (تجمع أغلب البحوث أن حجم العينة الملائم لتقدير نماذج الإنحدار هو أن يكون عشرة أضعاف عدد المتغيرات المستقلة على الأقل) حيث أن هذا العدد هو حوالي 17 ضعف عدد المتغيرات، وبعد مراجعة مجموعة من الاطباء الاختصاصيين تم اختيار تسعة متغيرات توضيحية ومتغير استجابة واحد (الطول) حيث بين الاطباء الاختصاصيون في مستشفى أبن سينا التعليمي والذين تمت استشارتهم أنها العوامل الرئيسية التي تؤثر على الاصابة بهذا المرض وكما هي موضحة بالجدول رقم (6) :-

جدول رقم (6) : يوضح متغيرات الدراسة

y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
الطول	العمر	الوزن	حساسية الحنطة	هرمون الغدة الدرقية	فحص هرمون النمو قبل التحفيز	فحص هرمون النمو بعد التحفيز	فحص لسكري	فحص اليوريا	فحص الكرياتين

اختبار التوزيع الطبيعي (Normality Test)

تم اختبار التوزيع الطبيعي باستخدام إختبار كولموكوروف سميرنوف وكذلك من رسم الاحتمال الطبيعي:

1- **اختبار Kolmogorov-Smirnov** ويعد من اهم اختبارات التوزيع الطبيعي ويسمى أحياناً بإختبار حسن المطابقة، كما موضح بالجدول (7)، حيث أن فرضيتي الاختبار هي:

H_0: عدم وجود مشكلة للتوزيع الطبيعي (البيانات تتبع التوزيع الطبيعي)

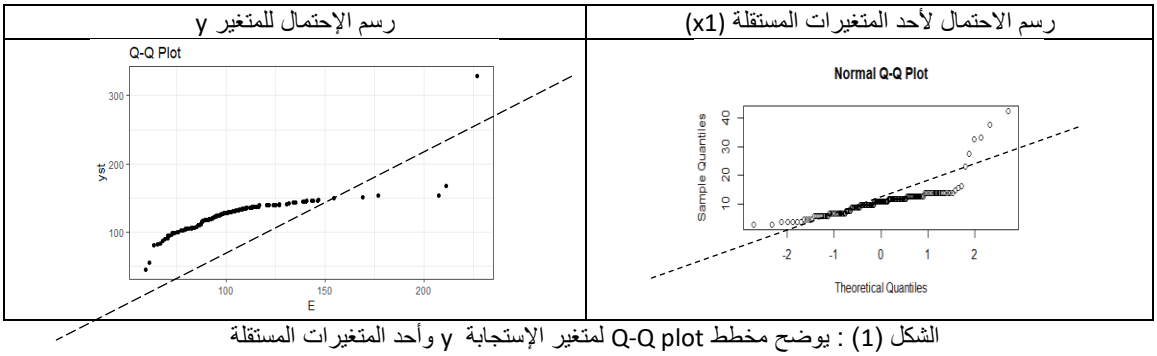
H_1: وجود مشكلة للتوزيع الطبيعي (البيانات لا تتبع التوزيع الطبيعي)

الجدول (7): جدول اختبار Kolmogorov-Smirnov

Statistic (D)	N	p-value
0.23111	150	1.091e-06

وبما ان قيمة p-value تساوي (1.091e-06) هي اقل من مستوى المعنوية 0.05 فان البيانات لا تتبع التوزيع الطبيعي .

2- **رسم الاحتمال الطبيعي Q-Q plot** لمتغير الإستجابة وأحد المتغيرات المستقلة (البقية كانوا على نفس الشاكلة)، ويوضح الشكل (1) يوضح ان البيانات لدينا لا تتوزع وفق التوزيع الطبيعي.



الشكل (1) : يوضح مخطط Q-Q plot لمتغير الإستجابة y وأحد المتغيرات المستقلة

تقدير النموذج الجمعي المعمم المقترح Wavelet Generalized Additive Model (WGAM)

تم تطبيق الخوارزمية المقترحة لتقدير النموذج الجمعي المعمم الممهد بالتقليص المويجي (المويجة الصغيرة) WGAM وحسب الخطوات الاتية:

1- البدء بإعطاء قيم اولية للدالة $m_i = m_i^{\circ} = 0 \quad , \quad i = 1,2,\dots,p$

2- ايجاد المصفوفتين Q و R كما في المعادلتين (1) و (2) على التوالي لكل j^{th} من المتغيرات التوضيحية ، ومن ثم ايجاد مصفوفة الجزاء التربيعية $K = QR^{-1}Q'$.

3- ايجاد المعلمة التمهيدية λ بالاسلوب الالي بالاعتماد على المعيار الشرعي المعمم GCV والمشار اليه في المعادلة (9) ، ومن ثم حساب مصفوفة التمهيد S وكما تم توضيحه تحت المعادلة (2) .

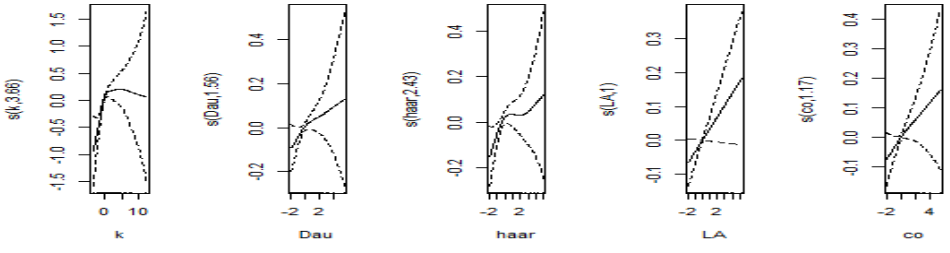
4- تحديث تقديرات m^{new} وذلك بإعادة او تكرار عملية التمهيد للبواقي الجزئية.

5- تقدير النموذج التجميعي المعمم المقترح المستند إلى التقليص المويجي WGAM باستخدام أربع انواع من دوال المويجة الاكثر شيوعاً (Daubechies,Haar, Least Asymmetric, Coiflets) وباستخدام قطع العتبة الناعم ومقارنتها مع GAM الاعتيادية ، والنتائج موضحة في الجدول (8) لأربعة مقدرات مقترحة (DWGAM, HWGAM, LAWGAM, CWGAM):-

جدول رقم (8) : يوضح نتائج تقدير GAM الاعتيادي و WGAM

معايير المقارنة		GCV	Concurvity	AIC	BIC
GAM		0.8318	0.026306	399.9243	416.9655
Wavelet Estimators (WGAM)	Daubechies (DWGAM)	0.9076	2.33E-05	207.4834	215.7273
	Haar (HWGAM)	1.0710	0.00119	219.8305	230.0873
	Least Asymmetric (LAWGAM)	0.975	1.29E-26	212.9603	219.9128
	Coiflets (CWGAM)	0.7772	1.18E-08	204.9526	212.2963

من ملاحظة الجدول (8) وباستخدام البيانات الحقيقية نجد ان هناك انخفاضاً في قيم معايير الكفاءة (GCV, BIC, AIC, Con.,) عند تقدير النموذج الجمعي المعمم وفق تقنية التقليص المويجي المقترحة WGAM مقارنة بالنموذج الجمعي المعمم الاعتيادي GAM. كما اظهر المرشح Coiflets (CWGAM) تفوقاً واضحاً على باقي المرشحات وذلك من خلال انخفاض قيم المعايير (GCV , Con. , AIC, BIC) مقارنة بباقي المقدرات الممهدة، وعليه وبالمجمل فإن المقدرات المقترحة تمتلك كفاءة عالية مقارنة بمقدر GAM (كفاءتها حوالي ضعف كفاءة GAM) وخاصة مقدر CWGAM والذي سجل أصغر قيمة مقارنة بالقيم المقترحة الثلاثة الأخرى فضلاً عن المقدر GAM قيمة منخفضة لمؤشر الإعتماد غير الخطي (Concurvity) للسب المذكور آنفاً. والشكل (2) يوضح رسم النموذج الجمعي المعمم GAM الاعتيادي ونماذج WGAM المقترحة.



شكل (2): يوضح رسم الـ GAM الاعتيادي و WGAM

ويلاحظ من الشكل (2) الذي يسمى رسم دالة الصفيحة Spline Function Plot المحور X يحوي على قيم المتغيرات التوضيحية، بينما يحتوي المحور y على قيم متغير الاستجابة ، ويشير الرقم في المحور y الى الدرجات الحرية الفعالة للمتغيرات التوضيحية ، وان الشكل K يمثل رسم النموذج GAM الاعتيادي ، اما باقي الاشكال فتشير الى رسم النموذج GAM بدوال التقليص المويجي الأربعة. ويرجع السبب في الاشكال المتعرجة (المكسورة) للمتغيرات التوضيحية الى الخطوط التمهيدية التي تصوغ النماذج اللاخطية في البيانات ، اي رسم كل دالة انحدار جزئي بشكل منفصل وهذه احد مزايا نموذج الانحدار الجمعي ، حيث يقلل من مشكلة الانحدارات متعددة الابعاد ، وتوفر هذه الخطوط نقطة سماح بنسبة 95% حول النموذج الموفق ومن ملاحظتنا للرسم نجد أن طريقة CWGAM كانت الأكثر تمهيداًحيث نلاحظ ان الحدود في دوال المويجة اقتربت من الصفر ونلاحظ ان رسم k ابتدا تقريبا من $0.5_$ اما رسم دوال المويجة فقد اقترب من الصفر ابتدا من $0.2-$ و 0.1 تقريبًا.

الاستنتاجات :

يمكن ادراج الاستنتاجات الخاصة بالبحث على ضوء نتائج التحليل :

1- ان استخدام نموذج GAM الاعتيادي يمثل طريقة مرنة للغاية لمشكلة البيانات ، بالإضافة الى انها لا تحتاج الى تحديد اولي لشكل العلاقة بين المتغيرات التوضيحية ومتغير الاستجابة ولكنها تنهار عندما تكون البيانات ملوثة.

2-عند تقدير النموذج الجمعي المعمم وفق اسلوب تقليص المويجة المقترحة (WGAM) باستخدام البيانات الحقيقية فقد اعطت مرشحات تقليص المويجة افضل النتائج مقارنة بطريقة GAM الاعتيادية حيث ساعدت ممهدات التقليص المويجي على تمهيد البيانات وذلك من خلال الحصول على اقل القيم لمعايير الكفاءة (أي كفاءة أعلى)، واظهرت المويجة Coiflets تفوقاً على باقي المويجات الاخرى .

في حالة استخدام اسلوب المحاكاة وعند استخدام النموذج الاول والثاني وتلويث البيانات بالتوزيعات (Laplace dist. , Exp. dist. ، Dist. (t)) بنسب (35%, 15% , 5%) وبأحجام عينة (300 , 150 , 50) لوحظ ايضاً ان أسلوب التمهيد بدالة دوبيجيز (DWGAM) وأسلوب التمهيد بدالة هار (HWGAM) كان لهما الأداء الأفضل مقارنة ببقية الطرق لسيناريوهات المحاكاة التي تم تناولها.

Refference

1- Amodio, S. ; Aria, M ; and D'Ambrosio, A, (2015), "**On Concurvity in nonlinear and nonparametric models**", STATISTICA, vol. 74, no.1 , pp. 85-98.

2- Augustin , N. H. H. , Sauleau , E. j A. , & Wood S. N. ,(2012) , **" On Quantile Quantile Plots for** Gen**eralized Linear Models"** , Computational Statisstics & Data Analysis 56 , 2404-2409 .

3- B. Mphan Kumar , R. Vidhya Lavanya , (2011**) , " Signal Denoising with Soft Threshold By Using Chui-Lian (CL) Multiwavelet " ,** International Journal of Electronics & Communication Technology IJECT Vol . 2 , Issue 1 , pp 38-42 **.**

4- Buja , A. , Hastie,T.J. and Tibshirani , R.J. ,(1989) , **"Linear Smoothing and Addition Models"** , Annass of Statistic , Vol.17,PP.453-555 .

5- Crowley . P. & Maraum . p. & Mayes , P. , (2006) **" How hard is the euro core ? an evaluation of growth cycles using wavelet analysis** " , Bank of Finland Research .

6- Donoho , L.D , Johnstone , M.l. , (1994) , **" Ideal Spatial adaptation by wavelet Shrinkage "** Biometrika , 81 , 3 , pp . 425-550 .

7- Friedman , J.H.,(1999),**" Multivariat Adaptive Regression Spline"** , The Annals of Statitics , Vol.19,No.1,pp 1-141.

8- Green,O.J. and Silverman, B.W.,(1994) , "Nonparametric Regression and Generalized Linear Models :ARoughness Penalty Approach " , Chapman and Hall , London .

9- He , Shui , (2004) , **"Generalized Additive Models for Data With Concurvity: Statistical Issus and a novel model Fitting Approach "** , School of Public Health Library , University of Pittsburgh .

10- Li,Y. and Ruppert , D. (2008) ,**"On the asymptotic of Penalized Splines"** , Biometricka , 95,415-436 .

11- Rajeev Aggarwal, Jai Karan Singh,and Other , (2011) , **"Noise Reducation of Speech Signal Using Wavelet Transform with Modified Universal Threshold "** International Journal of Computer Application , Volume 20, No.5 , P14-19.

12- Ruppert , D. , Wand , M.P. ,and Carroll,R.J. (2003) , **" Semmiparametric Regression"** Cambrdge Univeristy Press , New York .

13- Zitong , Z . , & Qawik.T , (2016) **, "Choosing Wavelet Method , Filter , and Length for Function Brain Network Construction "** , Jornal List , Plos one ,V.11(6) .

Some Wavelet Filters To Estimate Non-Parametric GAM Models With Application And Simulation

Alaa Abdel Sattar Daoud Hamoudat & Bashar Abdel Aziz Al-Talib
College of Computer Science and Mathematics, University of Mosul

.

Abstract:In this paper, the problem of the unknown probabilistic distribution of data and employing the method of estimating the Non-parametric Generalized Additive Models (GAM) based on smoothing splines as smoothers, and dealing with this case by using an iterative approach. The Wavelet Shrinkage technique was used and relied upon to estimate the Wavelet Generalized Additive Models (WGAM) regression model, which was proposed to be used as a data smoothers through the use of some wavelets as filters in the calculation of the discrete wavelet transformation, and some statistical criteria were used to compare the estimation methods. This done through the use of simulation and analysis of real data as well, the efficiency of the proposed methods was tested on data collected from Ibn Sina Teaching Hospital, on cases of short stature, and wavelet shrinkage smoothers gave the best results compared to the usual GAM method and the wavelet helped on smoothing data, by obtaining the most efficient results.

Keyword: Generalized additive model estimation method, nonparametric ,introductory, Use of wavelet reduction technique