AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

تقدير دالة الانحدار اللامعلمي متعدد المتغيرات للتوزيعات غير الطبيعية وباستعمال ممهدات اللب

م.د. سرمد عبدالخالق صالح	م.د. عمر رمزي جاسم
sarmadsalih@uohamdaniya.edu.iq	omar.ramzi89@uohamdaniya.edu.iq
قسم الرياضيات - كلية التربية - جامعة الحمدانية، نينوى، العراق.	قسم المحاسبة - كلية الادارة والاقتصاد - جامعة الحمدانية، نينوى، العراق.
أ.د. قتيبة نبيل نايف	**أ.م.د. عماد حازم عبودي**
dr.qutaiba@coadec.uobaghdad.edu.iq	emadhazim@coadec.uobaghdad.edu.iq
قسم الاحصاء - كلية الادارة والاقتصاد - جامعة بغداد، بغداد، العراق	

المستخلص

تعتبر التوزيعات ذات الذيول الثقيلة من اهم التوزيعات الاحتمالية المختلطة التي لها تطبيقات واسعة في مختلف مجالات الحياة وخاصة في المجالات التي تتعلق بعلم الاقتصاد الذي يهتم بموضوع التلوث البيئي نظراً للآثار الاقتصادية المترتبة عليه، لذا تم في البحث تقدير دالة الانحدار اللامعلمي متعدد المتغيرات عندما يتبع حد الخطأ العشوائي للأنموذج توزيعات ذات الذيول الثقيلة والمتمثلة بتوزيع مصفوفة زائدي معمم و مصفوفة بسل محور معمم (مصفوفة زائدي معمم متماثل). اذ تم تحويل انموذج الانحدار اللامعلمي متعدد المتغيرات الى انموذج خطي بالاعتماد على ممهد متعدد الحدود الموضعي ومن خلال الاسلوب التقليدي تم الحصول على ممهد نداريا واتسون متعدد المتغيرات وممهد الخطي الموضعي متعدد المتغيرات مفترضين ان تكون معلمات الشكل والالتواء معلومة. وقد استنتجنا بان ممهد نداريا واتسون وممهد الخطي الموضعي لـ **k** من متغيرات الاستجابة عندما يكون حد الخطأ يتبع توزيع مصفوفة بسل محور معمم هو ذات الممهدات عندما يتبع حد الخطأ العشوائي للأنموذج توزيع مصفوفة الطبيعي. فضلاً عن التطبيق على بيانات حقيقية متمثلة ببيانات التلوث البيئي /تلوث الهواء للعام **(2018)** ومن خلال نتائج التطبيق نلاحظ تفوق ممهد نداريا واتسون متعدد المتغيرات وعندما يتوزع الخطأ للأنموذج توزيع مصفوفة زائدي معمم اعتماداً على معيار **(MSE)**.

معلومات البحث

تواريخ البحث

تاريخ تقديم البحث: 2022/12/3
تاريخ قبول البحث: 2023/2/25
تاريخ رفع البحث على الموقع: 2023/12/31

الكلمات المفتاحية

التلوث البيئي، انموذج الانحدار اللامعلمي متعدد المتغيرات، توزيع مصفوفة زائدي معمم، توزيع مصفوفة بسل محور معمم، ممهد متعدد الحدود الموضعي، دالة اللب الطبيعية، طريقة قاعدة الابهام لاختيار معلمة التمهيد.

للمراسلة:

م.د. سرمد عبدالخالق صالح

sarmadsalih@uohamdaniya.edu.iq

https://doi.org/10.55562/jrucs.v54i1.571

1. المقدمة Introduction

التلوث هو تعرض الغلاف الجوي لمواد كيميائية أو جزيئات فيزيائية أو مركبات بيولوجية تسبب الأذى والإضرار بالبشر والكائنات الحية الأخرى أو تضر بالبيئة الطبيعية. يتم تعريفه أيضًا على أنه أي مادة في الهواء يمكن أن تسبب ضررًا للإنسان والبيئة، ويمكن أن تكون على شكل جزيئات صلبة أو قطرات سائلة أو غازات.

يُعد انموذج الانحدار اللامعلمي متعدد المتغيرات احد نماذج الانحدار المستخدمة في تحليل مثل هذه البيانات كونه موضع اهتمام للإحصائيين والعلماء وذلك لمرونته وسهولة التعامل مع عناصره، اذ قدم [1] بعض الطرائق التمهيدية المهمة لتقدير دالة الانحدار اللامعلمي عندما حد الخطأ العشوائي يتوزع طبيعياً والمتمثلة بطريقة الدالة اللبية وطريقة الجار الأقرب وطريقة الشريحة التمهيدية اذ تم استخدام طريقة العبور الشرعي لاختيار معلمة التمهيد بالنسبة إلى طرائق الدالة اللبية والشريحة التمهيدية. أما

بالنسبة إلى طريقة الجار الأقرب فقد استعمل أسلوب المسافة التقليدية لاختيار معلمة التمهيد، فضلاً عن استعمال اسلوب المحاكاة للمقارنة بين الطرق بنموذجين مختلفين وثلاث قيم للتباين. لكن هناك حالات يكون فيها خطأ الانموذج يتبع توزيع ذات النهاية الثقيلة وفي مثل هذه الحالة تكون التوزيعات المختلطة اكثر ملائمة ومن هذه التوزيعات توزيع مصفوفة زائدي معمم او توزيع مصفوفة بسل محور معمم، لذا قام [9] بتعميم التوزيع الزائدي متعدد المتغيرات المعمم الى توزيع مصفوفة زائدي معمم ودراسة بعض خصائصه وحالاته الخاصة، وتعود تسمية التوزيعات المختلطة أو الخليطة الى كونها ناتج من توزيع مصفوفة الطبيعي الملتوي مع توزيع مصفوفة كاوس المعكوس المعمم فضلاً عن تطبيقات التوزيع في التحليل البيزي لانموذج الانحدار الخطي الطبيعي متعدد المتغيرات بافتراض توزيع مصفوفة كاوس المعكوس المعمم كتوزيع اولي لمصفوفة التباين. و بما ان ظاهرة التلوث البيئي اصبحت مشكلة عالمية فقد عرض[3] مفهوم التلوث البيئي واشكاله وكيفية التقليل من خطورته، وذكر بان للتلوث البيئي اشكال وصور حسب درجتها و مصدرها، فقسم التلوث من حيث الدرجة الى التلوث المقبول والتلوث الخطر والتلوث المدمر، ومن حيث المصدر فقسم التلوث الى تلوث طبيعي وتلوث ناتج عن النشاط البشري والذي له تاثير سلبي في مجال الاقتصاد.

تناول المبحث الثاني تحويل انموذج الانحدار اللامعلمي متعدد المتغيرات الى انموذج خطي متعدد المتغيرات بالاعتماد على توسيع تايلر (Tayler Extension). اما في المبحث الثالث فتم وصف انموذج الانحدار اللامعلمي الزائدي المعمم متعدد المتغيرات. وتضمن المبحث الرابع الطرق التقليدية لانموذج الانحدار اللامعلمي الزائدي المعمم متعدد المتغيرات واشتمل المبحث الخامس على تطبيق ماتم التوصل اليه على بيانات حقيقية. اما في المبحث السادس والأخير فقد تم عرض أبرز الاستنتاجات التي تم الوصول اليها من خلال الدراسة.

2. ممهد متعدد الحدود الموضعي Localized Polynomial Smoother

يُعد هذا الممهد من الممهدات الجيدة المستخدمة في التمهيد اللامعلمي ويفضل على ممهدات اللب (Kernel Smoother) كما انه يستخدم في التصميم الثابت او العشوائي، ويمكن تمثيل انموذج الانحدار اللامعلمي متعدد المتغيرات بالصيغة الآتية: [10] [11]

$$Y_{im} = g_m(T_i) + \epsilon_{im} \tag{1}$$

$$\begin{bmatrix} y_{11} & \cdots & y_{1k} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nk} \end{bmatrix} = \begin{bmatrix} g_1(T_1) & \cdots & g_k(T_1) \\ \vdots & \ddots & \vdots \\ g_1(T_n) & \cdots & g_k(T_n) \end{bmatrix} + \begin{bmatrix} \epsilon_{11} & \cdots & \epsilon_{1k} \\ \vdots & \ddots & \vdots \\ \epsilon_{n1} & \cdots & \epsilon_{nk} \end{bmatrix} \tag{2}$$

اذ ان $g_m(T_i)$ تمثل مصفوفة دالة تمهيدية مجهولة (غير معروفة) ذو (m) من متغيرات الاستجابة و (T) من المتغيرات التوضيحية اللامعلمية وعادة مايكون متغيرا مستمرا، ونلاحظ من المعادلة (1) بان انموذج الانحدار اللامعلمي متعدد المتغيرات يكون غير خطي وعليه تم تحويله الى انموذج خطي بالاعتماد على توسيع تايلر وكالاتي:

نفترض بان الدالة (g) تمتلك مشتقة من الرتبة (u) عند النقطة (t) والنقاط (T_i) تقع في جوار النقطة (t) فان توسيع تايلر للدالة (g) تكون كالاتي:

$$g(T_i) = a_0 + a_1(t - T_i) + a_2(t - T_i)^2 + \cdots + a_u(t - T_i)^u \tag{3}$$

$$g(T_i) = \sum_{l=0}^{u} a_l(t - T_i)^l \qquad , \quad i = 1,2, \dots, n \tag{4}$$

نلاحظ بان المعادلة (4) اعلاه تمثل دالة الانحدار لـ متغير توضيحي لا معلمي ومتغير استجابة، وعليه تم تعميم المعادلة (4) لـ (k) من متغيرات الاستجابة وكالاتي: [11]

$$g_m(T_i) = \sum_{l=0}^{u} \sum_{m=1}^{k} a_{lm}(t - T_i)^l \quad , \quad i = 1,2, \dots, n \tag{5}$$

ومن المعادلة (5) اعلاه نلاحظ بان دالة الانحدار اصبحت خطية، وعليه يمكن كتابة الانموذج الخطي بصيغة المصفوفات وكالاتي:

$$Y = Ta + \epsilon \tag{6}$$

اذ ان:

$$Y=\begin{bmatrix} y_{11} & \cdots & y_{1k} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nk} \end{bmatrix}_{n\times k} \quad , \quad T=\begin{bmatrix} 1 & (t-T_1) & (t-T_1)^2 & \dots & (t-T_1)^u \\ \vdots & & \vdots & & \vdots \\ 1 & (t-T_n) & (t-T_n)^2 & \dots & (t-T_n)^u \end{bmatrix}_{n\times u+1}$$

$$a=\begin{bmatrix} a_{01} & a_{02} & \cdots & a_{0k} \\ & \vdots & \vdots & \vdots \\ a_{u1} & a_{u2} & \cdots & a_{uk} \end{bmatrix}_{u+1\times k} \quad , \quad \epsilon=\begin{bmatrix} \epsilon_{11} & \cdots & \epsilon_{1k} \\ \vdots & \ddots & \vdots \\ \epsilon_{n1} & \cdots & \epsilon_{nk} \end{bmatrix}_{n\times k}$$

3. وصف انموذج الانحدار اللامعلمي الزائدي المعمم متعدد المتغيرات

Description of the multivariate generalized hyper parametric nonparametric regression model

نعلم ان انموذج الانحدار اللامعلمي متعدد المتغيرات وصف حسب المعادلة (6)، وبافتراض مصفوفة الاخطاء العشوائية تتبع توزيع مصفوفة الزائدي المعمم حيث يمكن ايجاد دالة كثافة الاحتمال بالاستعانة بمفهوم التوزيعات المختلطة من توزيع مصفوفة وسط-تباين الطبيعي (الملتوي) وتوزيع كاوس المعكوس المعمم وكالاتي:[5]

$$\epsilon|q\sim MN_{n,k}(\delta q\ , Z\Sigma\ , \varphi^{-1}) \quad , \quad q\sim GIG(\lambda,\psi,\nu)$$

اذ ان:

q : متغير عشوائي يتبع توزيع كاوس المعكوس المعمم.

Σ : مصفوفة التباين من الدرجة $(k\times k)$.

δ : مصفوفة الالتواء(الانحراف) من الدرجة $(n\times k)$.

φ :مصفوفة قطرية تمثل اوزان دالة اللب من الدرجة $(n\times n)$.

$$\varphi=\text{diag}\left(ker_h(t-T_i)\right) \quad , \quad ker_h(.)=\frac{1}{h}\,ker\left(\frac{.}{h}\right)$$

$ker_h(.)$: دالة اللب وتكون محددة ومستمرة ومتماثلة وتكاملها مساوٍ للواحد الصحيح. [4][8]

h: معلمة التمهيد وتكون موجبة واكثر اهمية من دالة اللب ومن الممكن اختيارها حسب خبرة الباحث. [4][8]

وعليه فان دالة كثافة احتمال مصفوفة الاخطاء المشروطة بالمتغير العشوائي (t) هي كالاتي:

$$f(\epsilon|q)=\frac{|\varphi|^{\frac{k}{2}}}{(2\pi\ q)^{\frac{nk}{2}}\ |\Sigma|^{\frac{n}{2}}}\ e^{-\frac{1}{2\,q}tr\,(\epsilon-\delta q)^T\varphi(\epsilon-\delta q)\Sigma^{-1}} \qquad (7)$$

وان دالة كثافة الاحتمال للمتغير العشوائي (q) يمكن أن تعرف بالشكل الاتي: [6]

$$P(q)=\frac{\left(\frac{\lambda}{\psi}\right)^{\frac{\upsilon}{2}}}{2\ K_\upsilon(\sqrt{\lambda\psi})}\ q^{\upsilon-1}\ exp\left[-0.5\ \left(\left(\frac{\psi}{q}\right)+\lambda q\right)\right] \qquad ,q>0 \qquad (8)$$

وبالاعتماد على مفهوم التوزيعات الخليطة فان دالة كثافة الاحتمال لمصفوفة الاخطاء غير المشروطة بالمتغير العشوائي (q) تكون بالشكل الاتي:

$$f(\epsilon)=\int_0^\infty f(\epsilon|q)\ P(q)\ \ dq$$

$$f(\epsilon) = \frac{\left(\frac{\lambda}{\psi}\right)^{\frac{nk}{4}} |\varphi|^{\frac{k}{2}} \; e^{tr\,(\epsilon)^T \varphi\, \delta\, \Sigma^{-1}}}{(2\pi)^{\frac{nk}{2}} \; |\Sigma|^{\frac{n}{2}} \; K_v\left(\sqrt{\lambda\psi}\right)} * K_{\frac{2v-nk}{2}}\left(\sqrt{\lambda\psi\left(1+\frac{tr\,\epsilon^T\varphi\,\epsilon\Sigma^{-1}}{\psi}\right)\left(1+\frac{tr\,\delta^T\varphi\,\delta\,\Sigma^{-1}}{\lambda}\right)}\right) * \left(1+\frac{tr\,\epsilon^T\varphi\,\epsilon\Sigma^{-1}}{\psi}\right)^{\frac{2v-nk}{4}} \left(1+\frac{tr\,\delta^T\varphi\,\delta\,\Sigma^{-1}}{\lambda}\right)^{\frac{nk-2v}{4}} \tag{9}$$

المعادلة (9) تمثل دالة كثافة احتمال توزيع مصفوفة زائدي معمم. اذ ان:

λ, ψ, v: تمثل معلمات الشكل.

$K_v(.)$: تمثل دالة بسل المعدلة من النوع الثالث ذات الرتبة (v) والتي تاخذ الصيغة الاتية: [5][6]

$$K_v(x) = 0.5 \int_0^{\infty} y^{v-1} \exp\left(-0.5\, x(y+y^{-1})\right) dy \qquad x > 0 \tag{10}$$

كما يعبر عن توزيع مصفوفة الاخطاء العشوائية وصفياً بالآتي:

$$\epsilon \sim MGH_{n,k}(0, \Sigma, \varphi^{-1}, \lambda, \psi, v, \delta)$$

$$vec(\epsilon) \sim MGH_{nk}(vec(0), \Sigma \otimes \varphi^{-1}, \lambda, \psi, v, vec(\delta))$$

وعليه فان التوزيع الاحتمالي لمصفوفة المشاهداتY هو توزيع مصفوفة زائدي معمم كونه تركيبة خطية بدلالة مصفوفة الاخطاء العشوائية وان دالة كثافة احتماله تكون كالاتي:[9]

$$f(Y) = \frac{\left(\frac{\lambda}{\psi}\right)^{\frac{nk}{4}} |\varphi|^{\frac{k}{2}} e^{tr\,(Y-Ta)^T\, \varphi\, \delta\, \Sigma^{-1}}}{(2\pi)^{\frac{nk}{2}} \; |\Sigma|^{\frac{n}{2}} \; K_v\left(\sqrt{\lambda\psi}\right)} * K_{\frac{2v-nk}{2}}\left(\sqrt{\lambda\psi\left(1+\frac{tr\,(Y-Ta)^T\varphi(Y-Ta)\Sigma^{-1}}{\psi}\right)\left(1+\frac{tr\,\delta^T\varphi\,\delta\,\Sigma^{-1}}{\lambda}\right)}\right) * \left(1+\frac{tr\,(Y-Ta)^T\varphi\,(Y-Ta)\Sigma^{-1}}{\psi}\right)^{\frac{2v-nk}{4}} \left(1+\frac{tr\,\delta^T\varphi\,\delta\,\Sigma^{-1}}{\lambda}\right)^{\frac{nk-2v}{4}} \tag{11}$$

ويمكن التعبير عن هذا التوزيع وصفياً بالاتي:

$$Y \sim MGH_{(n,k)}\left(Ta, \Sigma, \varphi^{-1}, \lambda, \psi, v, \delta\right)$$

او

$$vec(Y) \sim MGH_{(nk)}\left(vec(Ta), \Sigma \otimes \varphi^{-1}, \lambda, \psi, v, vec(\delta)\right)$$

وتحت افتراض اذا كان خطأ الانموذج متماثلا اي تكون مصفوفة الالتواء مصفوفة صفرية نستنتج بان التوزيع الاحتمالي لمصفوفة المشاهدات (Y) هو توزيع مصفوفة زائدي معمم المتماثل (توزيع مصفوفة بسل محور معمم) ويكون كالاتي: [9]

$$f(Y) = \frac{\left(\frac{\lambda}{\psi}\right)^{\frac{nk}{4}} |\varphi|^{\frac{k}{2}}}{(2\pi)^{\frac{nk}{2}} |\Sigma|^{\frac{n}{2}} K_v\left(\sqrt{\lambda\psi}\right)} * K_{\frac{2v-nk}{2}}\left(\sqrt{\lambda\psi\left(1+\frac{tr\,(Y-Ta)^T\varphi(Y-Ta)\Sigma^{-1}}{\psi}\right)}\right) * \left(1+\frac{tr\,(Y-Ta)^T\varphi\,(Y-Ta)\Sigma^{-1}}{\psi}\right)^{\frac{2v-nk}{4}} \quad (12)$$

4. التقدير التقليدي لأنموذج الانحدار اللامعلمي الزائدي المعمم متعدد المتغيرات

Conventional estimation of multivariate generalized hyper parametric nonparametric regression model

اذا توفرت (n) من مشاهدات متغير الاستجابة والمتغيرات التوضيحية اللامعلمية ودالة اللب (Kernel function) فان دالة الامكان لـ (Y) غير المشروطة بالمتغير (q) من الصعب ايجاد مقدرات الامكان الاعظم منها لذا نلجا الى مفهوم التوزيعات المختلطة في ايجاد تلك المقدرات وعليه فان دالة الامكان لـ (Y) المشروطة بـ (q)تكتب بالشكل الاتي:

$$f(Y|a,\Sigma,q) = \frac{|\varphi|^{\frac{k}{2}}}{(2\pi\, q)^{\frac{nk}{2}} |\Sigma|^{\frac{n}{2}}}\, e^{-\frac{1}{2\,q}\,tr\,(Y-Ta-\delta q)^T\varphi(Y-Ta-\delta q)\Sigma^{-1}} \quad (13)$$

وباخذ اللوغارتم الطبيعي لطرفي المعادلة (13) اعلاه واخذ المشتقة الجزئية الاولى نسبة الى المصفوفة (a) نحصل على:

$$\frac{\partial \ln f(Y|a,\Sigma,q)}{\partial\, a^T} = -\frac{1}{2\,q}\,tr(-2\,T^T\varphi\,\mathrm{Y} + 2T^T\varphi\,\mathrm{T}\,a + 2T^T\varphi\,\delta\,\mathrm{q})\Sigma^{-1} \quad (14)$$

وبعد مساواة المعادلة (14) بالصفر واجراء التكامل نسبة الى المتغير العشوائي (q) نحصل على مقدر الامكان الاعظم لـ (a) وكالاتي:

$$\hat{a}_m = (T^T\varphi\,\mathrm{T})^{-1}T^T\varphi\,\mathrm{Y} - (T^T\varphi\,\mathrm{T})^{-1}T^T\varphi\,\delta * \frac{K_{v+1}\left(\sqrt{\lambda\psi}\right)}{K_v\left(\sqrt{\lambda\psi}\right)}\left(\frac{\lambda}{\psi}\right)^{-0.5} \quad (15)$$

وللحصول على ممهد ناداريا واتسون متعدد المتغيرات وممهد الخطي الموضعي متعدد المتغيرات نستخدم العلاقة:[7]

$$\hat{g}(t,u,h) = e^T\hat{a}_m \quad , \quad e^T = (1,0,0,\dots,0) \quad (16)$$

اذا كانت رتبة متعدد الحدود $(\mathrm{u}=0)$ نحصل على ممهد ناداريا واتسون متعدد المتغيرات وكالاتي:

$$\hat{g}(t,0,h) = \frac{1}{\sum_{i=1}^n ker_h(t-T_i)}\left[\sum_{i=1}^n ker_h(t-T_i)\,y_{i1}\,,\dots,\sum_{i=1}^n ker_h(t-T_i)\,y_{ik}\right] - \frac{\frac{K_{v+1}\left(\sqrt{\lambda\psi}\right)}{K_v\left(\sqrt{\lambda\psi}\right)}\left(\frac{\lambda}{\psi}\right)^{-0.5}}{\sum_{i=1}^n ker_h(t-T_i)}\left[\sum_{i=1}^n ker_h(t-T_i)\,\delta_{i1}\,,\dots,\sum_{i=1}^n ker_h(t-T_i)\,\delta_{ik}\right] \quad (17)$$

نلاحظ من المعادلة (17) بان الحد الاول من المعادلة عبارة عن ممهد ناداريا واتسون لـ (k) من متغيرات الاستجابة عندما حد الخطأ للانموذج يتبع توزيع مصفوفة الطبيعي ومصفوفة بسل محور معمم.

اما اذا كانت رتبة متعدد الحدود $(\mathrm{u}=1)$ نحصل على ممهد الخطي الموضعي متعدد المتغيرات وكالاتي:

$$\hat{g}(t,1,h) = [1,0]\,\hat{a}_m \quad (18)$$

بما ان:

$$T^T \Lambda \, \mathrm{T} = \begin{bmatrix} \sum_{i=1}^{n} ker_h(t-T_i) & \sum_{i=1}^{n} ker_h(t-T_i)\,(t-T_i) \\ \sum_{i=1}^{n} ker_h(t-T_i)\,(t-T_i) & \sum_{i=1}^{n} ker_h(t-T_i)\,(t-T_i)^2 \end{bmatrix}$$

وبالاعتماد على:[2]

$$\hat{s}_r(t,h) = n^{-1} \sum_{i=1}^{n} ker_h(t-T_i)\,(t-T_i)^r \tag{19}$$

$$(T^T \Lambda \, \mathrm{T})^{-1} = \frac{\begin{bmatrix} \hat{s}_2(t,h) & -\hat{s}_1(t,h) \\ -\hat{s}_1(t,h) & \hat{s}_0(t,h) \end{bmatrix}}{n\left[\hat{s}_2(t,h)\,\hat{s}_0(t,h) - \left(\hat{s}_1(t,h)\right)^2\right]}$$

فان ممهد الخطي الموضعي متعدد المتغيرات يكون كالاتي:

$$\begin{aligned} \hat{g}(t,1,h) = {} & \frac{n^{-1}}{\left[\hat{s}_2(t,h)\,\hat{s}_0(t,h) - \left(\hat{s}_1(t,h)\right)^2\right]} \\ & * \left[\sum_{i=1}^{n} ker_h(t-T_i)\,y_{i1}\,(\hat{s}_2(t,h) - (t-T_i)\hat{s}_1(t,h)), \ldots, \sum_{i=1}^{n} ker_h(t-T_i)\,y_{ik}\,(\hat{s}_2(t,h) \right. \\ & \left. - (t-T_i)\hat{s}_1(t,h)\right] - \frac{\frac{K_{v+1}\left(\sqrt{\lambda\psi}\right)}{K_v\left(\sqrt{\lambda\psi}\right)}\left(\frac{\lambda}{\psi}\right)^{-0.5} n^{-1}}{\left[\hat{s}_2(t,h)\,\hat{s}_0(t,h) - \left(\hat{s}_1(t,h)\right)^2\right]} \\ & * \left[\sum_{i=1}^{n} ker_h(t-T_i)\,\delta_{i1}\,(\hat{s}_2(t,h) - (t-T_i)\hat{s}_1(t,h)), \ldots, \sum_{i=1}^{n} ker_h(t-T_i)\,\delta_{ik}\,(\hat{s}_2(t,h) \right. \\ & \left. - (t-T_i)\hat{s}_1(t,h)\right] \end{aligned} \tag{20}$$

من المعادلة (20) نلاحظ بان الحد الاول يمثل ممهد الخطي الموضعي لـ (k) من متغيرات الاستجابة عندما يتبع حد الخطأ الانموذج توزيع مصفوفة الطبيعي ومصفوفة زائدي معمم متماثل (مصفوفة بسل المحور المعمم).

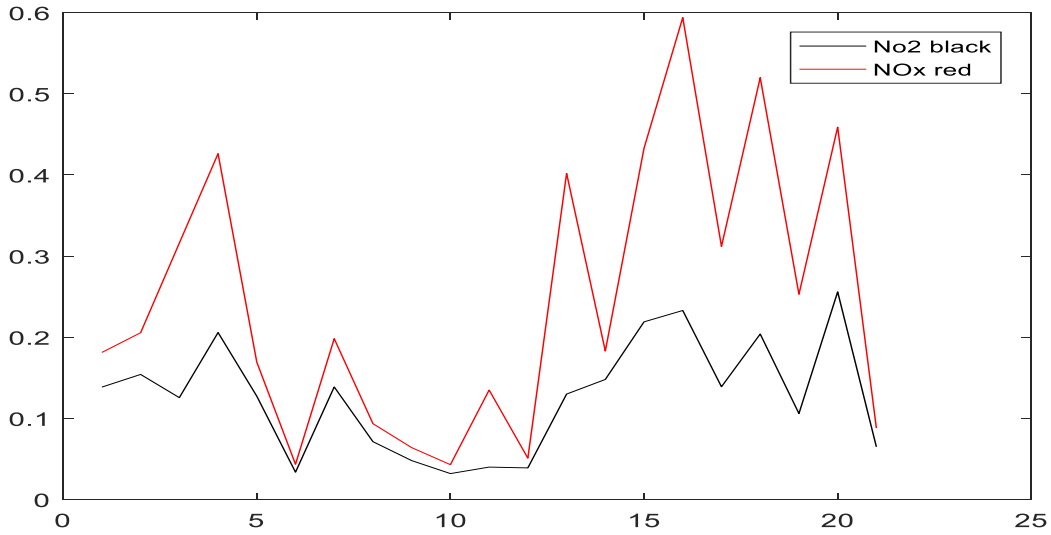
5. الجانب العملي Practical aspect

تم في هذا المبحث تطبيق ماتم التوصل اليه في المبحث (4) على بيانات حقيقية تتعلق بالتلوث البيئي/ تلوث الهواء اذ تم اخذ البيانات (للفترة من 3 حزيران لغاية 17 تموز) من وزارة الصحة والبيئة العراقية / محطة الوزيرية لسنة 2018.

5.1. تحديد المتغيرات الاساسية وتهيئة البيانات

يعرف التلوث على انه ادخال الملوثات التي تسبب تغيراً في البيئة الطبيعية، فقد يكون التلوث على شكل مواد صلبة او سائلة او غازية او على شكل طاقة مثل الحرارة والضوضاء والاشعاع، فهنالك اشكال رئيسية للتلوث مثل تلوث الهواء والتلوث الاشعاعي والقاء النفايات والتلوث الحاصل في التربة والمياه. فكانت تقديرات منظمة الصحة العالمية في عام (2016) ان (58%) من حالات الوفاة المبكرة كانت نتيجة تلوث الهواء. لذا سيتم في هذا البحث دراسة تأثير غاز الاوزون (O3) وهو غاز سام ذو لون ازرق مكون من ثلاث ذرات من الاكسجين ونسبته ضئيلة في الغلاف الجوي كمتغير توضيحي لا معلمي مستمر (t) على غاز ثاني أكسيد النيتروجين (NO2) وغيره من أكسيد النيتروجين (NOx) المنتجان من احتراق المركبات العضوية ومن عوادم الشاحنات والسيارات وبعض العمليات الصناعية بدرجة حرارة عالية كمتغيرات استجابة (y_1, y_2) على التوالي ويعد (NO2)

و(NOx) من اهم ملوثات الهواء التي تسبب ضراراً على الانسان والتي تعتبر اهم المشاكل الرئيسة للتلوث البيئي فقد يتسببان في تهيج الاغشية المخاطية للعين والمجاري التنفسية واضرار في الرئة. والشكل الاتي يبين غاز اكسيد وثاني اكسيد النتروجين.



شكل (1): سلوك غاز ثاني اكسيد النتروجين واكسيد النتروجين.

قبل اجراء عملية مطابقة البيانات لانموذج الدراسة تبين بان البيانات ملتوية نحو اليمين اذ كانت قيمة معامل الالتواء لغاز ثاني اكسيد النتروجين (y_1)(0.2945) والتفرطح (1.9974) وقيمة معامل الالتواء لغاز اكسيد النتروجين(y_2) (0.5123) والتفرطح (2.0705) وفي حالة تحويل مصفوفة متغيرات الاستجابة الى متجه (vector operator) اي عملية تكديس المصفوفة تبين قيمة معامل الالتواء (1.1817) والتفرطح (3.7154).

ولغرض معرفة ملائمة بيانات التلوث البيئي/ تلوث الهواء للانموذج المستخدم تم استخدام احجام عينات مختلفة من خلال اختبار حسن المطابقة (Chi-square test) وبالاعتماد على عينات مختلفة لمعلمات الشكل. والجدول ادناه يبين قيم الاختبار لعينات مختلفة لمعلمات الشكل التي ادت الى مطابقة البيانات تحت مستوى معنوية ($\alpha = 0.01$).

جدول (1): قيم اختبار مربع كاي لمطابقة البيانات

معلمات الشكل $(\lambda, \psi, \upsilon)$	Chi2-calculate	Chi2-tab. $(\alpha = 0.01)$
$(0.5, 1, 3)$	6.4512	9.2103
$(2, 3, 4)$	6.9215	9.2103
$(5, 4.5, 6)$	7.0214	9.2103
$(7, 6, 7.5)$	7.9548	9.2103
$(10, 8, 9)$	8.9545	9.2103

5.2 تقدير دالة الانحدار اللامعلمي متعدد المتغيرات للتوزيعات غير الطبيعية:

سيتم في هذا المبحث استعمال ممهدات اللب والمتمثلة بـ ممهد ندارايا واتسون متعدد المتغيرات والممهد الخطي الموضعي متعدد المتغيرات وبالاعتماد على العينات التي ادت الى المطابقة تحت مستوى معنوية ($\alpha = 0.01$) وبالنسبة لدالة اللب الطبيعية $((2\pi)^{-0.5}\exp(-x^2/2))$ وقاعدة الابهام لاختيار معلمة التمهيد $(h = 1.06\,\hat{\sigma}\; n^{-\frac{1}{5}})$ وعندما يتوزع خطأ الانموذج توزيعي مصفوفة زائدي معمم ومصفوفة بسل محور معمم، والجدولان ادناه يبينان قيم متوسط مربعات الخطأ لممهدات اللب:

جدول (2): قيم متوسط مربعات الخطأ لممهدات اللب تحت دالة اللب الطبيعية وطريقة قاعدة الابهام عندما يتوزع خطأ الانموذج توزيع مصفوفة زائدي معمم

الرتب	ممهد الخطي الموضعي متعدد المتغيرات	ممهد ناداريا واتسون متعدد المتغيرات	معلمات الشكل$(\lambda, \psi, \upsilon)$	
5	0.6021	0.5895	$(0.5, 1, 3)$	**1**
4	0.5487	0.4658	$(2, 3, 4)$	**2**
3	0.4512	0.4015	$(5, 4.5, 6)$	**3**
2	0.4136	0.3542	$(7, 6, 7.5)$	**4**
1	**0.3642**	**0.2954**	$(10, 8, 9)$	**5**

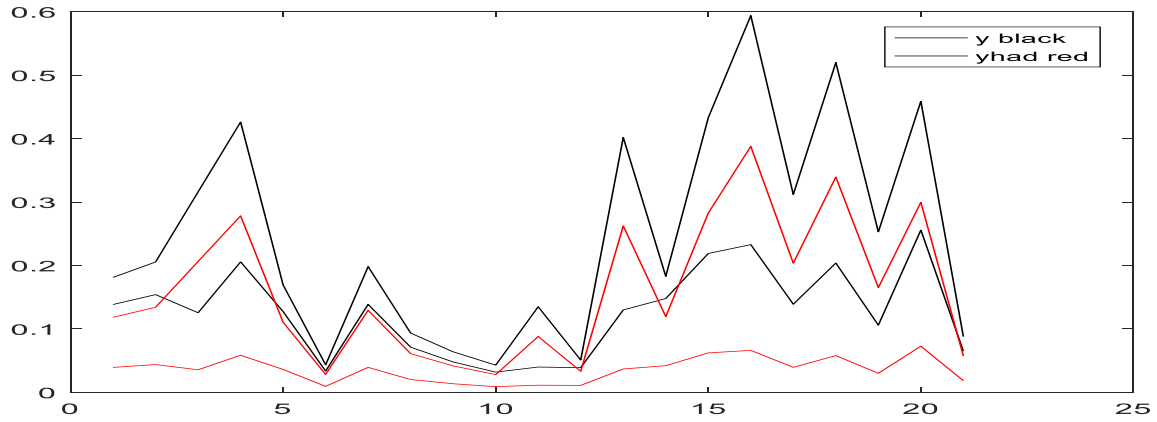
الجدول (3): قيم متوسط مربعات الخطأ لممهدات اللب تحت دالة اللب الطبيعية وطريقة قاعدة الابهام عندما يتوزع خطأ الانموذج توزيع مصفوفة بسل محور معمم

معلمات الشكل $(\lambda, \psi, \upsilon)$		ممهد نادارايا واتسون متعدد المتغيرات	ممهد الخطي الموضعي متعدد المتغيرات	الرتب
1	$(0.5, 1, 3)$	0.7890	0.7918	5
2	$(2, 3, 4)$	0.5411	0.6071	4
3	$(5, 4.5, 6)$	0.4121	0.5124	3
4	$(7, 6, 7.5)$	0.3995	0.4522	2
5	$(10, 8, 9)$	**0.3125**	**0.4055**	**1**

من خلال الجدولين (2) و (3) تبين كلما زادت قيم معلمات الشكل (العينات) تناقصت قيم معيار متوسط مربعات الخطأ بالنسبة لكلا الممهدين وعندما يتوزع خطأ الانموذج توزيع مصفوفة زائدي معمم ومصفوفة بسل محور معمم، وافضل ممهد كان ممهد نادارايا واتسون متعدد المتغيرات عند العينة الخامسة $(\lambda = 10, \psi = 8, \upsilon = 9)$ وعندما يتوزع خطأ الانموذج توزيع مصفوفة زائدي معمم وكالاتي:

$$\boldsymbol{Nadarya - Watson\ Smoother} = \hat{g}(t, 0, h) = [0.28451 \quad 0.65310]$$

نلاحظ من ممهد نادارايا واتسون متعدد المتغيرات فان تأثير غاز الاوزون (O3) كان بنسبة (28%) على غاز ثاني اكسيد النتروجين (NO2) ونسبة (65%) على غاز اكسيد النتروجين (NOx). والشكل (2) يبين القيم الحقيقية والممهدة لغاز (NO2) و(NOx).



شكل (2): سلوك المشاهدات الحقيقية والممهدة لغاز (NO2) و(NOx).

6. الاستنتاجات

توصل البحث الى اهم الاستنتاجات النظرية والتطبيقية منها:

1. عندما يكون خطأ الانموذج يتبع توزيع مصفوفة زائدي معمم متماثل (مصفوفة بسل محور معمم) فان ممهد نادارايا واتسون والخطي الموضعي متعدد المتغيرات تكون مماثلة عندما الخطأ يتبع توزيع مصفوفة الطبيعي وهذه الممهدات تكون كالاتي:

$$\hat{g}(t, 0, h) = \frac{1}{\sum_{i=1}^{n} ker_h(t - T_i)} \left[\sum_{i=1}^{n} ker_h(t - T_i)\, y_{i1}\, , \dots , \sum_{i=1}^{n} ker_h(t - T_i)\, y_{ik} \right]$$

$$\hat{g}(t, 1, h) = \frac{n^{-1}}{\left[\hat{s}_2(t, h)\, \hat{s}_0(t, h) - \left(\hat{s}_1(t, h) \right)^2 \right]}$$

$$* \left[\sum_{i=1}^{n} ker_h(t - T_i)\, y_{i1}\, (\hat{s}_2(t, h) - (t - T_i)\hat{s}_1(t, h)), \dots , \sum_{i=1}^{n} ker_h(t - T_i)\, y_{ik}\, (\hat{s}_2(t, h) - (t - T_i)\hat{s}_1(t, h) \right]$$

2. تناقص قيم معيار متوسط مربعات الخطأ عند تزايد قيم معلمات الشكل ولخطأ الانموذج الملتوي (توزيع مصفوفة زائدي معمم) والمتماثل (مصفوفة بسل محور معمم) ولكلا الممهدين.

3. تفوق ممهد نداريا واتسون متعدد المتغيرات على الممهد الخطي الموضعي متعدد المتغيرات وعندما يتوزع الخطأ توزيع مصفوفة زائدي معمم.
4. نسبة تأثير غاز الاوزون (O_3) على غاز ثاني اكسيد النتروجين (NO_2) وغاز اكسيد النتروجين (NOx) هي (28% و65 %) على التوالي.

المصادر العربية والاجنبية

[1] الصفاوي، صفاء يونس و متي، نور صباح (2011)، "تقدير دوال الانحدار اللامعلمي باستخدام بعض أساليب التمهيد"، المجلة العراقية للعلوم الاحصائية، عدد خاص بوقائع المؤتمر العلمي الرابع لكلية علوم الحاسوب والرياضيات، (20)، 373-392.

[2] حمود، مناف يوسف و عاشور، مروان عبدالحميد (2012)، "مقارنة بضعة مقدرات لا خطية لتقدير دالة الانحدار"، مجلة العلوم الاقتصادية والادارية، المجلد 18، العدد 68، 359-372.

[3] غرابية، خليف مصطفى (2010)، "التلوث البيئي: مفهومه واشكاله وكيفية التقليل من خطورته"، مجلة الدراسات البيئية، العدد (3)، 121-133.

[4] Heidenreich, et al. (2013), "Bandwidth selection for kernel density estimation: a review of fully automatic selectors". ASTA Advances in Statistical Analysis 97, no. 4 ,403-433.

[5] Gallaugher, M.P. & McNicholas, P.D. (2019), "Three skewed matrix variant distribution", Statistics & Probability Letters, Vol. 145, p.p. 103-109.

[6] Koudou, A. E. & Ley, C. (2014), "Characterizations of GIG laws: a survey complemented with two new result," Proba. Surv. , Vol. 11, p.p. 161-176.

[7] Ruppert, D. & Wand, M. P. (1994), "Multivariate locally weighted least squares regression", The Annals of Statistics, Vol. 22, No. 3, p.p. 1346-1370.

[8] Schucany, W. R. & Sommers, J. (1977), "Improved of kernel type density estimators", JASA, Vol.72, No. 353, p.p. 420-423.

[9] Thabane, L. & Haq, M. S. (2004), "On the matrix-variate generalized hyperbolic distribution and its Bayesian applications", Journal of Theoretical and Applied Statistical Science, Vol. 38, No. 6, p.p.511-526.

[10] Xiang, D., Qiu, P. & PU, X. (2013), "Nonparametric regression analysis of multivariate longitudinal data" , Statistica Sinica, Vol. 23, p.p. 769-789.

[11] You, J., Zhou, Y. and Chen, G. (2013), "Statistical inference for multivariate partially linear regression models", The Canadian journal of statistics, Vol.41, No.1, pages1-22.

Journal of AL-Rafidain University College for Sciences (2023); Issue 54; 1-10

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

Estimation of a Multivariate Non-parametric Regression Function of a Non-normal Distributions by Using Kernel Smoothing

Dr. Sarmad A. Salih	Dr. Omar R. Jasim
sarmadsalih@uohamdaniya.edu.iq	omar.ramzi89@uohamdaniya.edu.iq
Department of Mathematics - College of Education - Al-Hamdaniya University, Nineveh, Iraq.	Accounting Department - College of Administration and Economics - Al-Hamdaniya University, Nineveh, Iraq.
Qutaiba N. Nayef	**Emad H. Abody**
dr.qutaiba@coadec.uobaghdad.edu.iq	emadhazim@coadec.uobaghdad.edu.iq
Statistics Department - College of Administration and Economics - University of Baghdad, Baghdad, Iraq	

Article Information

Article History:
Received: December, 3, 2022
Accepted: February, 25, 2023
Available Online: December, 31, 2023

Keywords:
Environmental Pollution, Multivariate Nonparametric Regression Model, Generalized Hyperbolic Matrix Distribution, Generalized Bessel Matrix Distribution, Local Polynomial Smoother, Normal Core function, Rule of Thumb Method for Smoothing Parameter Selection.

Correspondence:
Dr. Sarmad A. Salih
sarmadsalih@uohamdaniya.edu.iq

Abstract

Distributions with heavy tails are regarded as one of the most significant mixed probability distributions that have wide applications in various fields of life, particularly in fields related to economics, which is concerned with the issue of environmental pollution due to its economic effects. Therefore, the multivariate nonparametric regression function was estimated in the research when it followed the error limit. The model has random distributions with heavy tails, which are represented by the distribution of the generalized hyperbolic matrix and the generalized basil matrix (generalized hyperbolic symmetric matrix). The multivariate nonparametric regression model was converted to a linear model based on the local polynomial smoother, and through the traditional method, the multivariate Nadaraya-Watson smoother and the multivariate local linear smoother were obtained, presuming that the shape and torsion parameters are known. We concluded that the Nadaraya-Watson smoother and the local linear smoother for (k) of the response variables when the error term follows the distribution of a generalized Bessel matrix are the same as the smoothers when the random error term of the model follows the distribution of the normal matrix. In addition to the application on actual data, which is represented by environmental pollution and air pollution data for the year (2018), and through the results of the application, we also observe the superiority of the multivariate Nadaraya-Watson paver, and when the error distribution of the model is a generalized hyperbolic matrix distribution based on the MSE criterion.

https://doi.org/10.55562/jrucs.v54i1.571