تقدير حدود الثقة لنموذج الانحدار اللامعلمي الرتيب

أ.د . دجلة إبراهيم مهدي

جامعة بغداد / كلية الادارة والاقتصاد

م.م. أسماء نجم عبدالله

جامعة بغداد / كلية الادارة والاقتصاد

المستخلص

عند تطبيق الانحدار اللامعلمي على البيانات ، ربما يتوقع الباحث في حالات معينة بان دالة متوسط الاستجابة ستكون رتيبة(متزايدة أو متناقصة) وعلى الرغم من مرونة الانحدار اللامعلمي ، فان هناك بعض الحالات والتي تكون فيها الدالة المقدرة ليست رتيبة ، لذا سيتم معالجة القيم الشاذة (outlier) لجعل الدالة اللامعلمية المقدرة رتيبة ،لقد تم تقدير دالة الانحدار اللامعلمي باستخدام ممهد الـ loess أو Lowess(Locally weighted scttarplot smoothing) وهو مختصر لممهد الرسم المبعثر الموزون الموضـــعي ، ومن ثم نقوم بتقدير الطرائق الرتيبة لجعل الدالة (متزايدة) ،وسوف نستخدم الطريقة الرتيبة لمعالجة القيم الشاذة من خلال التقنية المقترحة من قبل الباحثيــن (Mukerjee & Stern)والتي تعتمد على المعدلات الاعتيادية البسيطة للحدود العليا والدنيا وهي تقنيـــة Simple Average Methods (SAT) لتعديل مقدرات kernel الأولية وربطنا مع الانحدار اللامعلمي الرتيب طرائق إعادة المعاينة الـ (bootstrap) وقارنا بين هذه الطرائق من خلال استخدام أسلوب المحاكاة باستخدام معيار متوسط مربعات الخطأ (MSE)، وعليه فان هدف البحث هو تقدير الأنواع المختلفة من حدود الثقة لمقدرات الانحدار اللامعلمية هذه وتم تقدير حدود الثقة باستخدام نموذج الانحدار الموضعي المتعدد الحدود و الطرائق المستعملة والطرائق المقترحة من خلال الدمج بين طريقة الانحدار الرتيب وطرائق إعادة المعاينة وتم تطبيق هذه الطرائق على بيانات حقيقية خاصة بتلوث الهواء.

المفتاح: الانحدار الموضعي المتعدد الحدود، الانحدار الرتيب، طريقة المعدلات البسيطة، الانحدار الرتيب ذو المربعات الصغرى ،خوارزمية (PAV)، طريقة إعادة المعاينة الـ (bootstrap).

ESTIMATE CONFIDENE INTERAVALS FOR MONOTONE NONPARAMETRIC REGRESSION

Assistant instructor :Asmaa. N.Abdulla

Prof: Dr. Dejla.I.Mahdi

تقدير حدود الثقة لنموذج الانحدار اللامعلمي الرتيب

أ.د . دجلة إبراهيم مهدي   م.م. أسماء نجم عبدالله

Abstract

Monotonic regression is a nonparametric method designed for application in which the expected value of a response variable increases or decreases in relation to one or more explanatory. One approach is to first apply nonparametric regression to data Estimation of a response variable as a function of two continuous predictor variables is considered and then monotone smooth the initial estimates to ‶iron out‶ violations to the assumed order. Here, such estimators are considered, where local polynomial regression is first used, followed by monotone method using simple averages(SAT) and we use bootstrap procedure and compare by this methods through Monte Carlo simulation using mean – squared error(MSE) . The primary focus of this work is to estimate different types of confidence intervals for these monotone nonparametric regression estimators Most of the confidence intervals use local polynomial regression and resampling methods bootstrap procedure and suggested method and combine between this methods, the methods are then applied for a real data in air pollution field.

Keywords: local polynomial regression ; monotone regression; simple averages methods; least squares isotonic regression; pool –adjacent – violators (PAV); bootstrap.

ملاحظة : البحث مستل من أطروحة دكتوراه

اولاً: المقدمة[13][12][10][6]:

Introduction

يعدّ الانحدار الموضعي المتعدد الحدود Local Polynomial Regression نوع من نماذج الانحدار اللامعلمي العام وهو من الطرائق القديمة لتمهيد البيانات، إذ برز الاهتمام في الانحدار الموضعي المتعدد الحدود في سبعينيات القرن الماضي ، مازال الباحثين مستمرين في تطويره من ناحية الخوارزميات التي تسهل إلى حد كبير في تقييم الاتجاهات للسلاسل الزمنية للبيانات والخاصة بالوفيات والبيئة والأحوال الجوية ، إذ يقوم الانحدار الموضعي بتقدير سطوح الانحدار بواسطة المطابقة الموضعية للدوال الخطية أو التربيعية ذات المتغيرات المستقلة ،وفي نموذج الانحدار الموضعي المتعدد الحدود تكون الاستجابة والمتنبأت مرتبطة بالعلاقة الآتية :

$$Y_i = \mu(x_i) + \varepsilon_i ....\ ....\ ..(1-1) \qquad i = 1,2,......., n$$

إذ إن :

μ : تمثل متوسط سطح الانحدار

هدف ومشكلة البحث

هدف البحث هو تقدير الأنواع المختلفة من حدود الثقة لمقدرات الانحدار اللامعلمية وتم تقدير حدود الثقة باستخدام الطرائق اللامعلمية الرتيبة لتقدير دالة الانحدار اللامعلمي باستخدام ممهد الـ loess أو Lowess وهو مختصر لممهد الرسم المبعثر الموزون الموضعي(Locally weighted scttarplot smoothing) ، ومن ثم معالجة مشكلة القيم الشاذة الموجودة في دالة الانحدار اللامعلمي لجعل الدالة رتيبة (متزايدة) وربطنا مع الانحدار اللامعلمي الرتيب طريقة إعادة المعاينة إلـ (bootstrap) وقارنا بين هذه الطرائق من خلال استخدام أسلوب المحاكاة باستخدام معيار**MSE** وتم الدمج بين طريقة الانحدار الرتيب وإعادة المعاينة للحصول على أفضل المقدرات كطريقة مقترحة وتم تطبيق هذه الطرائق على بيانات حقيقية خاصة بتلوث الهواء.

ثانياً:الجانب النظري

(1-2)طرائق التقدير:

local polynomial Regression

[6]الانحدار الموضعي المتعدد الحدود(2-1-1)

يعتبر الـ loess أو Lowess(Locally weighted scttarplot smoothing) وهو مختصر ممهد الرسم المبعثر الموزون الموضعي والذي يعتبرها البعض كمترادفات ، واحد من عدة طرائق للنمذجة الحديثة والتي تبنى على الطريقة الكلاسيكية مثل انحدار المربعات الصغرى الخطية وغير الخطية والتي اقترحت من قبل الباحث Cleveland [4] وبعد ذلك تطورت على يد Cleveland و Devlin [5]، إذ إن الـ loess طريقة لامعلمية لتقدير سطوح الانحدار الموضعي وتسمح بمرونة أكثر من وسيلة النموذج الاعتيادي، وذلك لأننا نستطيع فيها معرفة الصيغة المعلمية الملائمة لسطح الانحدار إضافة إلى ذلك فان طريقة الـloess مناسبة عندما تكون هناك قيم شاذة (outlier) .

(1-1-1-2)الانحدار الموضعي المتعدد الحدود البسيط Local polynomial simple regression

ان الانحدار الموضعي المتعدد الحدود يعتبر طريقة مرنه لمطابقة نموذج الانحدار اللامعلمي المتعدد الحدود

$$\mu \setminus x_1, x_2, \dots\dots\dots \quad \dots x_p = f\left(x_1, x_2 \dots\dots\dots \quad x_p\right)$$

إن الهدف من الانحدار الموضعي المتعدد الحدود البسيط هو تقدير دالة الانحدار $\mu \setminus x$ عند قيمة النقطة المحورية (focal point) $x = x_0$.

حيث ينجز هذا الهدف بواسطة انحدار المربعات الصغرى الموزونة (WLS) لـ $y \setminus x$ للمشاهدات التي فيها قيم x تقترب إلى النقطة المحورية x_0 وحسب الخطوات التالية :-

1- حساب دالة kernel وهي تتمثل بـ $k\left(X_i - x_0\right)$ والتي تحقق اعلى وزن للمشاهدات القريبة الى النقطة المحورية x_0 وبعد ذلك تخفض تماثليا" عندما $\left|X_i - x_0\right|$ تكبر.

سوف نستخدم دالة(Gaussian- kernel) لحساب الأوزان وبفرض:

$$Z_i = \frac{(X_i - x_0)}{h}$$

إذ إن : Z_i تمثل المسافة المقاسة بين قيمة المتنبأ لـ ith من المشاهدات والقيمة المحورية x_0 .

وتتمثل دالة Gaussian-kernel بـ:

$$K_G(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

2- باستخدام أوزان kernel ، نقوم بانحدار WLS للمتعدد الحدود للدرجة dth لـ y/x الملائمة للمعادلة التالية :

$$Y_i = \beta_0 + \beta_1(X_i - x_0) + \beta_2(X_i - x_0)^2 + \dots\dots\dots\dots\dots + \beta_p(X_i - x_0)^P + \varepsilon_i$$

وعليه عندما تكون درجة التمهيد (d=1) فان الـ loess يميل إلى الانحدار الخطي الموضعي وكما يأتي:

$$Y_i = \beta_0 + \beta_1\left(X_i - x_0\right) + \varepsilon_i \dots\dots\dots \; \dots\dots\dots (2-1)$$

ومن نظرية المربعات الصغرى الموزونة القياسية نحصل على:

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y$$

إذ إن:

$$Y = [Y_1, \ldots\ldots\ldots \ldots, Y_n]^T$$

والذي هو متجه للاستجابات

ومصفوفةX من الدرجة n×(p+1) يمكن تمثيلها كما يلي:

$$X = \begin{bmatrix} 1 & (X_1 - x_0) \\ . & . \\ . & . \\ . & . \\ 1 & (X_n - x_0) \end{bmatrix}$$

حيث إن الانحدار الموضعي المتعدد الحدود يعمم لتقدير دالة kernel للمطابقة الموضعية عند النقطة المحورية x_0 باستخدام الأوزان W_i وان :

$$W_i = diag\left\{ k\left(\frac{X_1 - x_0}{h}\right), \ldots\ldots\ldots, \quad k\left(\frac{X_n - x_0}{h}\right) \right\}$$

وان المقدر الموضعي لدالة kernel المتعددة الحدود هو انحدار موزون للبيانات ، ومرتكز حول x_0 والهدف هو تقدير $g(x_0)$.

ويتم إيجاد عرض الحزمة (h) bandwidth أو الـ (S) Span للمقدار بواسطة الـ Visual trialوالـ error أو بواسطة الـ Cross-Validation أو بواسطة صيغة الـ minimization لتقدير الـ MSE ،إذ إنه في الـ Visual trial والـ error نحن نختار اقل قيمة لـ h أو S والتي تنتج عنها منحني الانحدار الممهد المعقول.

طريقة الـ (Cross- Validation).

إن الفكرة الرئيسية لهذه الطريقة هو اختيار قيمة (S)التي تقلل قيمة CV(S) وكل المشاهدات $\left(x_i, Y_i\right)$ تحذف من مجموعة البيانات ويتم التنبؤ بها بواسطة التمهيد للمشاهدات المتبقية n-1 وحسب الصيغة الاتية:

$$CV(S)=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\hat{g}_{-i}(x_i))^2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2-2)$$

وان:

$\hat{g}_{-i}(.)$ ترمز إلى المقدر الممهد عندما نقطة البيانات الفردية $\left(x_i, Y_i\right)$ المحذوفة من مجموعة البيانات ، وفقط البيانات المتبقية n-1 هي المستخدمة لحساب هذا التقدير ، إن هذا الحساب يعتمد على ترك واحدة خارجا "Leaving –out one لتقديرات الانحدار $\hat{g}_{-i}(.)$ والتي يمكن توضيحها بالمعادلة التالية :

$$\hat{g}_{-i}(x_i)=\frac{\hat{g}(x_i)-L_i(x_i)Y_i}{1-L_i(x_i)} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2-3)$$

وبتعويض المعادلة أعلاه في المعادلة (2-2) يصبح معيار (CV) كما يلي :

$$CV(S)=\frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i-\hat{g}(x_i))^2}{(1-L_i(x_i))^2} \qquad \ldots\ldots\ldots\ldots\ldots(2-4)$$

إما بالنسبة للـ **Generalized Cross- Validation**

نقوم باستبدال القيم لـ $L_i\left(x_i\right)$ بواسطة المعدل $\frac{df_{\text{mod}}}{n}$ وهذا يقودنا إلى :

$$GCV(S)=n\sum_{i=1}^{n}\frac{(Y_i-\hat{g}(x_i))^2}{(n-df_{\text{mod}})^2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(2-5)$$

إذ إن:

df_{mod}: يمثل درجة حرية النموذج ويساوي

$$df_{\text{mod}} = trace\ (L)$$

وان L تمثل مصفوفة التمهيد smoothing matrix إي إن :

$$L = X(X'WX)^{-1}X'W$$

وعليه يمكن إن نكتب هذا المعيار بالطريقة التالية :

$$GCV(S) = n^{-1}\sum_{i=1}^{n}\frac{(Y_i - \hat{g}(x_i))^2}{(1 - n^{-1}tr(L))^2} \qquad \ldots\ldots\ldots\ldots(2-6)$$

علما" إن هذا المعيار يستخدم لاختيار المعلمة التمهيدية وله فائدة هي سرعة الحساب ، إذ إن لطبيعة هذا المعيار في إجراء الخطوات المتسلسلة تجعل معيار GCV في بعض الأحيان يزداد في أنموذج ثم يبدأ يتناقص ثانية وهكذا إلى إن يتوقف ونختار معلمة التمهيد التي تقابل اصغر $GCV(S)$.

(2-1-2)مقدرات الانحدار الرتيب[12] [11][10][1] ويضم:

The Monotone Regression Estimators

بعد تنفيذ الانحدار اللامعلمي باستخدام الانحدار المتعدد الحدود الموضعي الذي ينتج التقديرات الأولية $g(x)$ الـ $\mu(x)$ يمكننا تطبيق التمهيد الرتيب (Monotone Smoothing) إذا كان من الممكن الافتراض بان الاستجابة إما متزايدة أو متناقصة فيما يتعلق بـ x .

وبالتحديد بالنسبة لـ x و x' والعلاقة $x \le x'$ تعني :

$$x_1 \le x_1', x_2 \le x_2', \ldots\ldots\ldots\ldots, x_p \le x_p'$$

وعليه فان الدالة تكون متزايدة إذا $\mu_{(x)} \le \mu_{(x')}$ عندما $x \le x'$ ، وبما $g(x)$ مقدرات أولية لـ $\mu(x)$ ، و سيتم استخدام طريقتين للتمهيد الرتيب لتقدير دالة الانحدار اللامعلمية الأولية عند نقاط البيانات المشاهدة .

(1-2-1-2) طريقة (Simple Average Method (SAT))[10][11]

تعتمد هذه الطريقة على الحدود العليا و الدنيا للمجاميع الجزئية للبيانات والشكل العام لهذه الطريقة هي :-

$$\mu^{\alpha}(x_i) = (1-\alpha)\mu_1(x_i) + \alpha\mu_2(x_i) \ldots\ldots\ldots \ \ldots\ldots\ldots (2-7)$$

وهي الصيغة العامة لطريقة Mukerjees monotone المتزايدة

$$\mu_1(x_i) = \min \{g(x') : x' \geq x\}$$

$$\mu_2(x_i) = \max \{g(x') : x' \leq x\}$$

وان حساب α بسيطاً جدا وكما يأتي:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n}[Z_X - M_1(x_i)][M_2(x_i) - M_1(x_i)]}{\sum_{i=1}^{n}[M_2(x_i) - M_1(x_i)]^2} \ldots\ldots\ldots\ldots (2-8)$$

(3-1-2) طرائق إعادة المعاينة

Resampling Methods

(1-3-1-2)طريقة الـ(Bootstrap):

اقترح Efron [8] طريقة الـ (Bootstrap) لإعادة المعاينة وهي طريقة تكرارية استخدمت لمعالجة دقة التقديرات في العينات الصغيرة حيث تقوم هذه الطريقة على مبدأ إيجاد تقديرات غير متحيزة من مجموعة التقديرات المتحيزة وذلك بتوليد مجموعة كبيرة من العينات المسحوبة بشكل عشوائي من بيانات العينة نفسها مع الإرجاع وبحجم العينة الأصلي نفسه .

لقد حاولEfron بهذه الطريقة حل مشكلة صغر حجم العينة بواسطة إعادة المعاينة وتوليد عدد كبير من العينات، و بعد إجراء (K) من التكرارات يتم الحصول على (K) تقدير لمعالم الأنموذج . وان معدل هذه التقديرات يدعى تقدير الـ Bootstrap .

علما بان طرائق إعادة المعاينة لا تعطي تقدير أفضل للمعالم ، و إنما بإعادة المعاينة نحصل على توزيع خطأ المعاينة في المقدرات ، لذا فان أسلوب الـBootstrap يستخدم كوسيلة لتصحيح التحيز , وسنقوم بتوضيح إعادة معاينة البواقي لأسلوب الـBootstrap في الانحدار .

(2-1-3-2-1) طريقة إعادة معاينة البواقي

يتم هذا الأسلوب من خلال توظيف المقدرات الأولية لمعالم الأنموذج الأصلي وبحسب المعادلة رقم(1-2) بغية الحصول على البواقي ، ومن ثم مركزتها (أي التقدير حول نقطة المتوسط) ومن ثم إعادة معاينتها ليتحول نموذج الانحدار إلى نموذج انحدار

جديد وعلى وفق الصيغة :

$$Y^* = X^*\beta + \varepsilon^* \text{.......... ....} (2-9)$$

ثم نطبق طريقة المربعات الصغرى الموزونة على معادلة رقم(2-33) لتقدير المعلمات وعلى وفق الصيغة:

$$\beta^* = (X^{*T}WX^*)^{-1}X^{*T}WY^*$$

ليكون لدينا بعد ذلك نموذج الانحدار التقديري باستخدام طريقة الـBootstrap بالشكل الأتي:

$$\hat{Y}^* = X^*\beta^*$$

ثم نحسب تقدير الخطأ العشوائي على وفق الصيغة الآتية:

$$\varepsilon_i^* = Y_i^* - \hat{Y}_i^* \text{........} (2-10)$$

ومن ثم نستخرج متوسط مربعات الخطأ باستخدام الصيغة:

$$BMSE = \frac{\sum_{i=1}^{n} \varepsilon_i^{*2}}{N-P} \text{.......... ....} (2-11)$$

الطريقة المقترحة :

بعد تقدير دالة الانحدار اللامعلمي باستخدام الانحدار الموضعي المتعدد الحدود الـ loess سيتم تطبيق إحدى طرائق الانحدار الرتيب وبعد ذلك تطبيق إحدى طرائق إعادة المعاينة في محاولة للدمج بينهما .

(2-5) حدود الثقة[9][13] Confidence Intervals

عليه تكون حدود الثقة عند 95% كما يأتي :

بما إن

$$\hat{g}_{(x_i)} = \sum_{i=1}^{n} L_i(x_i)\, y_i$$

فإن الانحراف المعياري لـ $\hat{g}(x)$ هو

$$\sigma(x) = \sigma\sqrt{\sum_{i=1}^{n} L_i^2(x)}$$

وإن تقدير $\sigma(x)$ هو :

$$S(x) = S\sqrt{\sum_{i=1}^{n} L_i^2(x)}$$

وعليه فان حدود الثقة (NPR) المعتمدة على توزيع t والتي تأخذ الشكل الأتي :

$$g(x) \mp t_{\alpha/2, v} SE^{g}(x) \ldots\ldots\ldots \quad \ldots\ldots\ldots. \quad \ldots..(2-12)$$

إذ إن : $t_{\alpha/2, v}$: تمثل القيمة الحرجة المرتبطة بالمنطقة $\alpha/2$ في Upper tail لتوزيع t مع درجات حرية v

إما بالنسبة لتقديرات الطرائق المستعملة فهي شبيه بتلك الموجودة في(2-12) ولكن تتمركز على التقديرات للطرائق المستعملة بدلاً من تقديرات الانحدار اللامعلمية

$$\hat{\mu}(x) \mp t_{\alpha/2, v} SE^{g}(x) \ldots\ldots\ldots \quad \ldots\ldots\ldots. \quad \ldots..(2-13)$$

وهكذا فان حدود الثقة لتقديرات الطرائق المستعملة تنطبق عليها المعادلة أعلاه في استخراجها في حالة حجم العينة اقل من (30).

إما حدود الثقة في حالة حجم العينة اكبر من (30) فيمكن توضيحها كما يأتي :

$$\hat{\mu}(x) \mp Z_{\alpha/2} SE^{g}(x) \ldots\ldots\ldots \quad \ldots\ldots\ldots. \quad \ldots..(2-14)$$

ثالثاً: الجانب التجريبي

(1-3)المحـــــاكاة

إن عملية المقارنة بين الطرائق التي تناولتها دراستنا بغية التوصل إلى أفضلها في التقدير كان احد الأهداف الرئيسة في الدراسة وان عملية تحديد أفضل طريقة تتطلب أيضاً تحديد الظروف التي تعمل فيها طرائق التقدير والتي أدت بنا إلى الحكم في تحديد أفضل طريقة ، ولغرض تطبيق أسلوب المحاكاة لابد من توفر الكثير من الشروط منها حجم العينة ودالة الوسط الحسابي الصحيحة واختيار درجة الانحدار اللامعلمي وعدد المتنبأت في النموذج وطبيعة الخطأ والذي يتضمن نوع التوزيع والتباين.

وباستخدام أسلوب محاكاة الـ (**Monte Carlo**) كونها توصف الأنظمة العشوائية التي تستخدم في استخراج القيم التي يصعب استخراجها بطريقة رياضية، وبصورة عامة تستخدم هذه الطريقة عينات أرقام عشوائية وتستخرج في ضوئها النتائج معتمدة على قيم أرقام العينات المستخدمة.

(2-3) توليد المشاهدات للتوزيعات المستخدمة : Generating Observation for the Used Distribution

فيما يأتي سندرج كيفية توليد مشاهدات تخضع للتوزيعات المفترضة للخطأ وذلك باستخدام الأرقام العشوائية المولدة من خلال الحاسبة الإلكترونية والتي تخضع للتوزيع الطبيعي حيث تكون هذه الأرقام مستقلة عن بعضها بعضاً.

(1-2-3)توليد مشاهدات تتبع التوزيع الطبيعي :

إن عملية توليد مشاهدات تخضع لهذا التوزيع بالاعتماد على طريقة (**Monte Carlo**).

إذ إن:

X_i : يمثل المتغيرات التوضيحية وتم توليد مشاهداته وإن هذا المتغير يتبع التوزيع الطبيعي.

$$X_i \sim N(\mu, \sigma^2) \qquad i = 1, 2, ..., n$$

تم توليد المتغيرات التوضيحية بالاعتماد على الدوال الجاهزة وكما يأتي:

بالاعتماد على لغة *Matlab (R2011) varsion7.11 فأن:

```
x1_data=normrnd(mu_x1,ck_x1,n_sample_size,1);
```

إما ε_i : فيمثل المتغير العشوائي وتم توليده بالاعتماد على التوزيعات الآتية:-

1- التوزيع الطبيعي القياسي.   (Standard Normal distribution)

2 - التوزيع الطبيعي الآسي.   (lognormal distribution)

إما عن سبب اختيارنا لهذه التوزيعات فهو من حقيقة كون الكثير من الظواهر تتبع تلك التوزيعات.

(2-3-3) توليد الأخطاء العشوائية التي تتبع التوزيعات الإحصائية الآتية:-

تم الاعتماد في توليد الأخطاء على الدوال الجاهزة في لغة((R2011) (Matlab varsion7.11) وكما مبين فيما يأتي:

```
R = normrnd(mu,sigma)

R = normrnd(mu,sigma,m,n,...)

R = normrnd(mu,sigma,[m,n,...])
```

Lognormal random numbers

و ثبتت كل توزيعات الخطأ لكي يكون لها التباين نفسه كما يأتي:

```
er=normrnd(0,1.025,n_sample_size,1);

er= lognrnd(0,1.025,n_sample_size,1);
```

أي إن المقارنة ستكون على أساس نوع التوزيع وليس على أساس التباين،إما بالنسبة لإحجام المشاهدات فقد تناولت الدراسة أربعة إحجام (**10,25,50,100**).

النموذج المستعمل في حالة المتغير التوضيحي الواحد:

$$Y_i = 0.9 + 0.035 X_1 + \varepsilon_i ....(3-1)$$

(3-3) نتـــــائج المحـــــاكاة:

تم تكرار تجربة المحاكاة بمقدار (1000) تجربة لكل أنموذج من نماذج الانحدار الافتراضية ولكل توزيع من توزيعات الخطأ العشوائي, وقد جرى إعداد جداول للنتائج التي جرى الحصول عليها خلال عملية المحاكاة لمقارنة الطرائق الرتيبة للانحدار اللامعلمي إذ تم احتساب MSE لكل أنموذج ولكل توزيع وبحسب الصيغة الاتي:-

$$MSE = \frac{1}{n} E\sum_{i=1}^{n}\left[g^*(x_i) - g(x_i)\right]^2 ....(3-2) \qquad i = 1,2,......,n$$

جدول رقم (3-1)

يبين قيم متوسط مربعات الخطأ (MSE)وعندما(d=1,p=1)

Distribution	Sample size	m_had_x	Boot	Sugg
Normal	10	1.1237	1.3890	1.1131
	25	1.0893	1.1482	1.0825
	50	1.0669	1.1161	1.0516
	100	1.0350	1.1108	1.0248
Lognormal	10	11.0819	12.4818	10.8098
	25	8.2454	8.1703	7.0351
	50	5.3117	5.1742	5.0866
	100	4.9869	4.8568	4.6024

إذ إن:

m_had_x: طريقة (M&S)

Boot: طريقةbootstrap

Sugg: الطريقة المقترحة

تفسيرات الجدول رقم (1) الخاص بقيم (MSE) لكل الطرائق ولجميع أحجام العينات والتوزيعات المستعملة

- أظهرت النتائج إن الطريقة المقترحة تمتلك اقل**(MSE)** إي أنها أفضل طريقة ولجميع أحجام العينات المستعملة وللتوزيعين الطبيعي والطبيعي الآسي .
- تناقص قيم **(MSE)** مع تزايد حجم العينة ولجميع الطرائق في حالة التوزيع الطبيعي والطبيعي الآسي.

- نلاحظ إن قيم **(MSE)** في حالة التوزيع الطبيعي اقل من قيم **(MSE)** في حالة التوزيع الطبيعي الآسي.
- نلاحظ إن اكبر قيم للـ **(MSE)**هي عند حجم العينة(10) في حالة التوزيع الطبيعي الآسي ولجميع الطرائق المستعملة.

رابعاً: الجانب التطبيقي

(1-4): جمع البيانات*:-

وقد تم الحصول على البيانات من وزارة البيئة وهي تشمل على عدد المحطات ومواقعها والتراكيز الملوثة التي تقيسها الأجهزة في كل محطة. وان مدة عمل كل محطة هو ثمانية ساعات يومياً والمحطات هي محطة ساحة الأندلس، محطة الجادرية، محطة كراج العلاوي

(2-4): بعض ملوثات الهواء الرئيسة :-

(أولاً) مجموع الدقائق العالقة (TSP) :-

مجموع الدقائق العالقة Total Suspended Particulatesهي أي مواد مشتتة أو منتشرة في الهواء قد تكون صلبة أو سائلة أو غازية ، وهي الجسيمات المادية التي تتراوح حجومها الجزيئية من أجزاء الميكرون إلى أقل من 100 ميكرون ، والتي تطلق إلى الجو من مصادر قد تكون طبيعية أو بفعل أنشطة الإنسان المختلفة وقد تحتوي على الكثير من المركبات العضوية واللاعضوية وبعض العناصر الثقيلة ومن أهمها عنصر الرصاص المنبعث في أكثر الأحيان من عوادم السيارات ، ويتم حساب التركيز ألوزني للدقائق العالقة بوحدات مايكرو غرام/م^3 من وزن الدقائق المتجمعة على ورقة الترشيح وحجم الهواء المار خلال الجهاز ودرجة تدفق الهواء كما في المعادلة الآتية

وزن الغبار الذي تم جمعه بالمايكرو غرام

التركيز بالمايكرو غرام / م^3 =----------------------------------

معدل تدفق الهواء بالمتر المكعب / دقيقة × زمن التشغيل بالدقيقة

ثانياً (العناصر الثقيلة):

تعد العناصر الثقيلة من الملوثات البيئية المهمة والتي ازداد الاهتمام بها خلال العشرين سنة الماضية بسبب التلوث البيئي لبعضها والأهمية الحياتية للبعض الآخر لإدامة النمو في جسم الكائن الحي ، وهي تتصف بخاصية التجمع في جسم الكائنات الحية وعدم قابليتها على التكسر الجيولوجي والبكتريولوجي في البيئة وتدخل إلى جسم الإنسان عن طريق الماء والهواء والسلسة الغذائية،إذ يؤدي استمرار انبعاثها إلى الهواء من مصادرها المختلفة (الطبيعية والصناعية) إلى زيادة تراكيزها في

الغلاف الجوي ، وتم الاعتماد على جهاز(460GFG) لقياس تراكيز الكثير من العناصر الثقيلة مثل الحديد (Fe) والنيكل (Ni) والخارصين (Zn) .

(*) الواقع البيئي لنوعية الهواء (2009,2010)/ وزارة البيئة / دائرة التخطيط والمتابعة الفنية/ قسم نوعية الهواء

جدول رقم (4-1) يبين تراكيز الدقائق العالقة والنيكل للنصف الأول من عام 2009 والنصف الثاني من عام 2010

X_1	Y
41.37	627
17.6	824
128	653
99.4	638
132.6	523
29	485
1.6	313
69.3	559
76.3	511
89.4	495
61.4	602
202.5	807

جدول رقم (4-2) يبين حدود الثقة للدقائق العالقة والنيكل لنموذج الانحدار الخطي الموضعي المتعدد الحدود للطرق المستعملة وبدالة وزن Gaussian وعرض حزمة(0.3)

$g(x_i)$	Upper	Lower	m_had_x	Upper	Lower	boot	Upper	Lower	Sugg	Upper	Lower
611.51	1354.47	-131.44	630.478	1367.58	-106.63	597.29	1340.25	-145.66	598.10	1341.06	-144.85
626.96	1369.92	-115.99	630.47	1367.5	-106.63	597.86	1340.82	-145.09	598.10	1341.06	-144.85
619.94	1362.90	-123.01	631.97	1369.08	-105.13	598.07	1341.03	-144.88	598.10	1341.06	-144.85
645.23	1388.19	-97.72	631.97	1369.08	-105.13	597.33	1340.29	-145.62	598.10	1341.06	-144.85
620.64	1363.60	-122.31	631.97	1369.08	-105.13	598.05	1341.01	-144.90	598.10	1341.06	-144.85
612.21	1355.17	-130.74	631.97	1369.08	-105.13	598.29	1341.25	-144.66	598.17	1341.13	-144.79
589.03	1331.99	-153.92	631.97	1369.08	-105.13	598.96	1341.92	-143.99	598.17	1341.13	-144.79
610.10	1353.06	-132.85	631.97	1369.08	-105.13	598.35	1341.31	-144.60	598.17	1341.13	-144.79
630.47	1373.43	-112.48	631.97	1369.08	-105.13	597.76	1340.72	-145.19	598.17	1341.13	-144.79
633.99	1376.95	-108.96	631.97	1369.08	-105.13	597.66	1340.62	-145.29	598.17	1341.13	-144.79
623.45	1366.41	-119.50	631.97	1369.08	-105.13	597.96	1340.93	-144.99	598.17	1341.13	-144.79
646.63	1389.59	-96.326	631.97	1369.08	-105.13	598.31	1341.27	-144.64	598.31	1341.27	-144.64

(2-1-5)الجانب التطبيقي:-

اولاً:

من خلال البيانات المقدمة لنا من شعبة مراقبة نوعية الهواء والضوضاء والمستخدمة في الجانب التطبيقي استنتجنا ما يأتي:

1- قلة المواقع التي تؤخذ منها النماذج لغرض قياس كميات الغبار المتساقط والدقائق العالقة الكلية إذ اقتصرت على ثلاثة مواقع فقط وهي مواقع قليلة جداً لاتعكس واقع كميات الغبار المتساقط والدقائق العالقة الكلية في المحافظة كونها تقع في مناطق تجارية وسكنية.

2- من خلال دراسة القراءات التي تم تسجيلها في المواقع التي تؤخذ منها نماذج قياس الدقائق العالقة الكلية فإن هناك زيادة في جميع هذه القراءات عن الحد المقبول والبالغ (350) مايكرو غرام/3 م.

3- لم تقم الشعبة بقياس الغازات الملوثة للهواء مثل غاز أول اوكسيد الكاربون وثاني اوكسيد الكبريت والتي تعتبر من الغازات السامة وذلك بسبب قلة الكادر المؤهل وعطل جهاز (460GFG) الخاص بفحص الغازات الملوثة .

ثانياً:

لاحظنا من جداول التطبيق العملي أن المدى (الفرق بين الحد الأعلى والحد الأدنى) زاد بالنسبة للطريقة المقترحة الثالثة مقارنة بالطرق المستعملة الأخرى وهذا مايعطي مرونة إحصائية .

وهذا يعد تفسيراً ايجابياً من الناحية الاحصائيه عند مقارنة التقدير على شكل نقطـة (point estimator) والتقدير على شكل فترات (Confidence Intervals) إذ إن الثاني يعطي أكثر مرونة تقديرية من الأول ولاسيما ونحن نعمل في الجانب الإحصائي اللامعلمي.

(2-5)التوصــــيات :

يمكن إجمال التوصيات الآتية على ضوء الاستنتاجات التي توصلنا إليها من خلال هذه الدراسة

1- من خلال ما لاحظناه من زيادة في قراءات قياس الدقائق العالقة نوصي باتخاذ الإجراءات اللازمة ومفاتحة الجهات ذات العلاقة للسيطرة على المصادر الملوثة والتي لها تأثير مباشر في نوعية الهواء مثل معامل الاسبست التي تتطاير منها كميات كبيرة من الدقائق العالقة.

2- نوصي أيضا بتصليح الأجهزة والمعدات المتخصصة بقياس الملوثات والاستفادة منها.

3- استعمال صيغ أخرى من دوال أوزان (Kernel) مثل (Epanechnikov)، (Biweight)، (Double Exbonential) ، (Gaussian)، (Uniform).

4- نوصي بتكثيف أجهزة القياس بالأماكن الصناعية والمكتظة بالسكان عما هو عليه بالمناطق الزراعية.

5- استعمال صيغ دوال الانحدار اللامعلميه مثل (Spline)،(Piecewise) و(Additive).

6- دمج طريقة (Jackknife) بعد طريقة (Bootstrap) .

7- في الجانب التطبيقي الخاصة بالفصل الرابع ، استعمال متغيرات أخرى غير مذكورة في دراستنا التطبيقية مثل درجة الحرارة ، الرطوبة ، إعداد المرضى .....وغيرها .

References

1- المصادر العربيـــة Arabic References

1. الجواد ، ياسمين عبد الرحمن، (2007) " **تقدير دالة الانحدار اللامعلمي باستخدام بعض الطرائق اللامعلمية الرتيبة مع تطبيق عملي للمقارنة بينها"** ، رسالة ماجستير ، كلية الإدارة والاقتصاد ، جامعة بغداد .

2. الساعدي ، سمانة عزيز ، (1993) " **دراسة تطبيقية لأسلوبي الـ Jackknife و الـ Bootstrap لنماذج الانحدار الخطية و غير الخطية "**، رسالة ماجستير في الإحصاء، كلية الإدارة و الاقتصاد ، الجامعة المستنصرية.

3. كاظم ، مريم حسون ، (2003)" **البوتستراب في تحليل نماذج الانحدار مع تطبيق عملي"** ، أطروحة دكتوراه ، كلية الإدارة والاقتصاد ، جامعة بغداد .

2- المصادر الاجنبية Foreign References

4. Cleveland,W.S.(1979). **Robust locally weighted regression and smoothing scatter- plots.** Journal of the American statistical association. Vol:74, no: 368, p:829-836.

5. Cleveland, W. S., Devlin , S.J., Grosse, E.(1988).**Regression by local fitting : Methods, Properties, And Computational Algorithms**. Journal of Econometrics, vol: 37, p:87-114.

6. Dette,H.,Neumeyer. N, Pliz , K.F,(2003**). A simple nonparametric estimator of monotone regression function.**

7. Dette,H., , Pliz , K.F,(2004). **A comparative study of monotone nonparametric kernel estimates**. Holger .dette@ruhr- uni-bochum.de/ mathematics/ preprint.htm.

8. Efron, B.(1982) . **The Jacckknife ,The Bootstrap And Other Resampling Plans.** Society for industrial and applied mathematics.philadeliphia , Pennsylvania.

9. Hwang, J. T. G., Peddada, S. D. (1994**). Confidence interval estimation subject to order restrictions**. Ann. Stat. 22:67–93.

10. Mukerjee, H. (1988). **Monotone nonparametric regression.** Ann. Stat. vol:16,p:741–750.

11. Mukerjee, H., Stern, S. (1994). **Feasible nonparametric estimation of multiargument monotone functions**. J. Am. Stat. Assoc. 89:77–80.

12. Strand, M. (2003). **Comparison of methods for monotone nonparametric multiple regression.**Comm. Stat. Simulat and Comput.vol: 32,no:(1),p: 165–178.

13. Strand, M. , Zhang, yu.,Swihart,B.j.(2010). **Monotone nonparametric regression and confidence intervals.**Comm. Stat. Simulat and Comput. Vol:39, no: (4),p:828-845.