

**دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة
اللامعلمية للبيانات العنقودية مع طرق أخرى**

أ.م.د. سجي محمد حسين	م.م. حلا كاظم عبيد
قسم الإحصاء - جامعة بغداد	قسم العلوم المالية والمصرفية - الجامعة العراقية

**Experimental study to compare Robust proposed methods for
estimating nonparametric function of cluster data with other
methods**

Ph.D. Saja Mohammad Hussein

hala kadeem obeid

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

أ.م.د. سجي محمد حسين

م.م. خلا كاظم عبيد

المستخلص

تهدف التحليلات الاحصائية الى تقييم تاثير المتغيرات التوضيحية على متغيرات الاستجابة خلال فترة معينة ولان النماذج المعلمية تخضع الى العديد من الافتراضات والقيود فقد تم اللجوء الى النماذج اللامعلمية. اذ تم تقدير الدالة اللامعلمية لنموذج الانحدار اللامعلمي للبيانات العنقودية التي تكون مرتبطة داخل العنقود الواحد حيث ان (Y_{ij}, X_{ij}) تمثل متغيرات استجابة والمتغيرات التوضيحية عبر فترات زمنية مختلفة على التوالي. ومما تجدر الاشارة اليه ان البيانات العنقودية تشبه البيانات الطولية من حيث اعتمادية البيانات داخل العنقود الواحد (القطاع بالنسبة للبيانات الطولية) على بعضها البعض. وتم التطرق في هذا البحث الى تقدير الدالة اللامعلمية للبيانات العنقودية غير المعلومة باستخدام طرق مقترحة حصينة التي تحاكي الاساليب الاحصائية ولكنها لا تتاثر بالقيم المتطرفة حيث يكون اداء الطرق الكلاسيكية ضعيفا بوجود هذه القيم^[8] ومقارنتها مع طرق أخرى عن طريق معيار المقارنة MSE و MAE باستخدام بعض الدوال ضمن تجربة المحاكاة حيث حققت الطرق المقترحة باستخدام دالة Huber لتقدير الدالة اللامعلمية للبيانات العنقودية اقل MSE و MAE من بقية الطرق الاخرى.

الكلمات المفتاحية: الانحدار اللامعلمي، البيانات العنقودية، مصفوفات الارتباط، دالة MSE, MAE, Huber.

ABSTRACT

The statistical analysis aims to assess the impact of the explanatory variables on the response variables during period because the parametric models are subject to many restrictions and borrowings have been resorting to nonparametric models, the nonparametric function was estimated to nonparametric regression design for cluster data that are linked with the same cluster that (Y_{ij}, X_{ij}) represents response variables and explanatory variables across different time periods, respectively. It should be noted that the data cluster is similar to longitudinal data in terms of reliability of the data with the same cluster (sector of the longitudinal data) on each other.

This paper addressed to estimate the unknown nonparametric data cluster function by using Robust proposed methods to mimic the statistical methods but they are not affected by other influential values, where the performance of classical methods is weak in presence of these values, and compare it with other methods by using criteria standard MSE and MAE , and some of the functions within the simulation experience, which achieved less from the rest of the other function methods by using Huber Function to assess the nonparametric function of the cluster data.

Key words: nonparametric regression, cluster data, correlation matrices, Huber Function, MSE, MAE.

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

المقدمة :

تكمّن فلسفة الاحصاء في محاولة نمذجة مختلف الظواهر بنماذج اقرب مايمكن الى الواقع وهذه النماذج هي على اشكال ,منها النماذج السببية التي تقوم على اساس السبب والنتيجة ومنها ما تسمى بنماذج الانحدار اللامعلمي ويتم صياغة نموذج الانحدار اللامعلمي وفق الاتي:

$$Y_{ij} = \theta (X_{ij}) + \varepsilon_{ij} \quad (1)$$

فان $\theta(\cdot)$ تكون اختيارية باستثناء خصائص قابلية الاشتقاق

ويمكن تعريف البيانات العنقودية بأنها مجموعة البيانات التي تكون على هيئة قياسات مكررة لدراسة (شخص، آلة، حيوان) وان تلك القياسات المكررة تكون غير مستقلة والقياسات لنفس (شخص، آلة، حيوان) تميل الى ان تكون اكثر شبيها من تلك القياسات للأشخاص المغايرين.^[2] استعملت البيانات العنقودية في الكثير من مجالات الطب الحيوي وعلم الوبئة والظواهر الزراعية والاقتصادية وغيرها .

مشكله البحث:

تكمّن مشكلة البحث في كيفية تقدير الدالة اللامعلمية للبيانات عنقوديه بوجود ارتباط بين العناقيد باستخدام الطرق الحصينة التي تحاكي الأساليب الإحصائية ولكنها تتأثر بالقيم المتطرفة حيث يكون أداء الطرق كلاسيكياً ضعيفاً بوجود هذه القيم.

هدف البحث :

يهدف البحث الى اقتراح طريقتين حصينتين باستخدام دالة Huber هي Robust seemingly unrelated estimator و kernel estimator و Robust generalized least squares smoothing spline estimator لتقدير الدالة اللامعلمية للبيانات العنقودية لتحقيق اقل MSE و MAE مقارنة مع طرق أخرى باستخدام المحاكاة من خلال عدة دوال باستعمال مصفوفات ارتباط مختلفة ومقارنتها مع حالة الاستقلالية.

الاستعراض المرجعي:

وفي عام (2000) بينت دراسة الذوات (RUCKSTUHL, A., WELSH, A.H. & CARROLL, R.J.) ان استخدام التقديرات الموضوعية لمقدرات quasi-likelihood تعطي تقديرات اسوأ مقارب محاذي في حالة تقدير الدالة اللامعلمية للبيانات المكررة او البيانات العنقودية في ظل افتراض استقلالية البيانات لذلك تم استخدام المقدرات التجميعية البسيطة التي تعطي افضل مقارب محاذي ولكن ليس في جميع الظروف^[7].

وفي عام (2001) قام كل من (Lin,X. Carroll,R.J.) بتقدير نموذج خطي شبه معلمي معمّم جزئياً للبيانات العنقودية باستخدام المعادلات التقديرية المعممة (GEE) والنماذج الهامشية (الحدية) تفترض ان المتوسط لمتغير النتيجة يعتمد على المتغيرات المعلمية وغير المعلمية لكل مستوى من العنقود وان الطريقة اللبية تسمح بتطور مصفوفة الارتباط داخل العنقود^[3] .

وفي عام (2003) قامت الباحثة (WANG,N.) بايجاد طريقة لبية مقترحة تختلف عن الطرق اللبية القياسية اثبتت فيها ان مقدرات النموذج المعلمي وشبه المعلمي بوجود الارتباطات داخل العناقيد اكثر كفاءة من اهماله اوعدم وجوده مستخدماً بذلك التباين للمقارنة بين المقدرين عن طريق المحاكاة نظراً لتعقيد هيئة التحيز المحاذي الذي حال دون استخدام متوسط مربعات الخطأ^[9] .

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

في عام (2004) قدم كل من LIN,X. WANG,N. WELSH ,H.A. AND R.J.CARROLL بحثاً استخدموا طريقة التكافؤ بين مقدر النواة غير المتصل ظاهرياً *seemingly unrelated kernel estimator* ومقدر الطريقة الشرائحية من خلال علاقة المساواة بين عرض الحزمة (Bandwidth) ومعلمة التمهيد وتم حساب كلا المقدرين بالاعتماد على الارتباط داخل العناقيد من خلال المقارنة باستخدام كفاءة متوسط مربعات الخطأ للطريقة الشرائحية نسبة إلى مقدر النواة غير المتصل ظاهرياً [4] .

وفي عام 2007 قام Zeger, S.,L., بدراسة النموذج اللوجستي للبيانات العنقودية من ناحية الدراسات الاختبارية لنمو الاطفال لنموذج خطي كدراسة اصابة عدوى الاطفال التنفسية ونقص فيتامين A وبالاعتماد على النماذج الحدية (الهامشية) [11] .

وفي عام 2013 قام كل من Ma,S.,Song,Q.&Wang.L., بتطوير منهجية عامة لاختيار المتغيرات انيا وتقدير العناصر المجهولة للنماذج التجميعية الخطية جزئياً (additive partially linear models) باختصار (APLMS) للبيانات العنقودية واقترحوا في الخطوة الاولى المربعات الصغرى للحصول على التقدير المعلمي اما المركبات اللامعلمية فيتم تقديرها على اساس متعدد الحدود للشرائح التمهيدية polynomial spline smoothing [5] .

2- الجانب النظري:

الطرق المقترحة باستعمال دالة وزن Huber لمقدر M الحصينة:

اقترحت اولاً طريقة M من قبل (Huber) وتستند هذه الفكرة على تقليل الاخطاء لبعض الدوال بدلا من تقليل مجموع المربعات لها . ومقدرات M الحصينة تستخدم الطرق التكرارية مثل نيوتن التكرارية لتقدير الموقع ومصفوفة التباين المشترك [8] .
وان نظام التكرار لمقدرات M متعددة المتغيرات يفشل بدون نقاط ابتدائية جيدة وان M تصبح اكثر تحسس للمشاهدات الشاذة بتزايد عدد المتغيرات اي ان

$$\min \sum_i \sum_j e_{ij}^2 = \min \sum_i \sum_j (y_{ij} - \theta(x_{ij}))^2 \quad (2)$$

ان تقدير M المطورة من Huber تعتمد على فكرة ابدال مجموع مربعات الاخطاء e_{ij}^2 بدالة اخرى للأخطاء الهدف منها تقليل المقدر الاتي:

$$(**) \sum_i \sum_j p(y_{ij} - \theta(x_{ij}))^2$$

وان p دالة محدبة متماثلة ولتقليل المقدار السابق نأخذ المشتقة بالنسبة الى الدالة غير المعلومة $\theta(x_{ij})$

$$\sum_i \sum_j \psi(y_{ij} - \theta(x_{ij})) = 0$$

حيث ان ψ المشتقة الجزئية للدالة p بالنسبة للدالة غير المعلومة $\theta(x_{ij})$ و $\psi = p'$ وبهذا يكون هناك nm من المعادلات غير الخطية ويمكن حلها بواسطة طريقة المربعات الصغرى الموزونة التكرارية او طريقة

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى
 نيوتن أو Huber، ولكي تكون مقدرات M لها خاصية الـ scale invariant^(*) يتم إعادة كتابة المعادلة
 أعلاه. $\sum_i \sum_j p \left[\frac{(y_{ij} - \theta(x_{ij}))^2}{S} \right]$ والثانية تكون $\sum_i \sum_j \psi \left[\frac{y_{ij} - \theta(x_{ij})}{S} \right] = 0$
 ولحساب دالة الوزن فان :

$$W_{ij} = \frac{\psi \left[\frac{Y_i - \theta(x_{ij})}{S} \right]}{\left[\frac{Y_i - \theta(x_{ij})}{S} \right]} \quad (3)$$

$$S = 1.483 \times MAD \quad (4) \quad \text{حيث ان :}$$

تقدر قيمة المقدر المعياري S مرة واحدة قبل البدء بعملية التكرار وان

$$MAD(r_i) = \text{med}_i \{|r_i - \text{med } r_i|\}$$

حيث ان :

$$r_i = \frac{y_{ij} - \theta(x_{ij})}{S} \quad \text{حيث } \text{med } r_i = \text{قيمة الوسيط لـ } r_i \text{ وان :}$$

وباستخدام دالة Huber التالية

$$p(r) = \begin{cases} r^2 / 2 & , |r| \leq c \\ c|r| - c^2 / 2 & , |r| > c \end{cases}, \quad W(r) = \begin{cases} 1 & , |r| \leq c \\ c|r|^{-1} & , |r| > c \end{cases} \quad (5)$$

حيث ان c هو ثابت القطع ويأخذ القيمة $c = 1.345$ تهدف إلى كفاءة مقارنة إلى التوزيع الطبيعي القياسي.

وباستخدام دالة وزن Huber لمقدر M الحصين على الطريقتين الـ Seemingly unrelated kernel estimator^[1] وطريقة الـ generalized least squares smoothing spline estimator^[1] لنحصل على الطرائق المقترحة الآتية :

طريقة مقترحة أولى (Robust seemingly unrelated kernel estimator)

في هذه الفقرة تم الاعتماد على الأسلوب المقترح من قبل الباحثين Wang, B. & et al (2014)^[8] لتوظيف دالة وزن Huber لمقدر M وذلك للحصول على مقدرات حصينة . فعند توظيف دالة وزن Huber في الطريقة^{[1][4]} Seemingly unrelated kernel estimator سنحصل على الصيغة التالية :

$$\hat{\theta}_{knew}(X) = \{I + K_w(\tilde{V}^{-1} - \tilde{V}^d)\}^{-1} k_w \tilde{V}^{-1} Y^* \quad (6)$$

حيث ان

بدالة أخرى للأخطاء الهدف منها تقليل المقدار e_{ij}^2 تعتمد على فكرة إبدال مجموع مربعات الخطأ^(*) scale invariant الـ

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

Y^* : تمثل البيانات الزائفة التجريبية pseudo date .

$\hat{\theta}_{knew}(X)$: تمثل تقدير دالة النواة غير المتصل ظاهريا باستخدام طريقة دالة وزن Huber لمقدر M الحصينة.

$$V = A^{1/2} R(\alpha) A^{1/2}$$

A: مصفوفة الانحراف المعياري لـ Y و $R(\alpha)$ ومصفوفة الارتباط لـ y .

$$\tilde{V}^d = \text{diag}(V^d, V^d, \dots, V^d) ; V^d = \text{diag}(V^{-1}) = \text{diag}(V^{jj})$$

ان k_w عبارة عن مصفوفة $nm \times nm$

$$k_w = \{k_{wh}(X_{11}), \dots, k_{wh}(X_{nm})\}'$$

$$k_{wh}(x) = \{\sum_{i=1}^n \sum_{j=1}^m k_h(X_{ij} - x) V^{jj}\}^{-1} \{k_h(X_{11} - x), \dots, k_h(X_{nm} - x)\}'$$

$$k_h(s) = h^{-1} k_h(s/h)$$

وتتلخص طريقة توليف دالة وزن Huber لمقدر M باستعمال الخوارزمية التالية:

1- اجعل قيمة عداد التكرار $I = 0$ ثم نستعمل تقديرات $\hat{\theta}_k(X)$ في معادلة الـ Seemingly unrelated kernel estimator [4] كقيم اولية .

2- تقدير الانحراف المعياري للبواقي S وحسب الصيغة (4)

3- يتم اختيار دالة وزن حسب الصيغة (3) .

4- يتم حساب البيانات الزائفة pseudo data التجريبية بالصيغة التالية

$$Y^* = \hat{\theta}(X) + n \left(\frac{\psi(r_i)^* MAD}{2} \right)$$

5- وبتكرار الخطوات 3 و 4 الى حين ثبات تقدير $\theta_{knew}(X_{ij})$ للتكرار I فتكون الصيغة بالشكل

الاتي :

$$\hat{\theta}_{knew}(X) = \{I + K_w(\tilde{V}^{-1} - \tilde{V}^d)\}^{-1} k_w \tilde{V}^{-1} Y^*$$

[8] وتستمر عملية التكرار حتى يصبح

$$\|\hat{\theta}_{knew}(X_{ij})^{(I)} - \hat{\theta}_{knew}(X_{ij})^{(I-1)}\| \leq \hat{\theta}_{knew}(X_{ij})^{(I-1)} \quad (7)$$

طريقة مقترحة ثانية (Robust generalized least squares smoothing spline estimator)

في هذه الفقرة تم الاعتماد على الاسلوب المقترح من قبل الباحثين [8] Wang, B. & et al (2014) لتوليف دالة

وزن Huber لمقدر M وذلك للحصول على مقدرات حصينة . فعند توليف دالة وزن Huber في الطريقة

generalized least squares smoothing spline estimator [4][1] سنحصل على الصيغة التالية :

$$\hat{\theta}_{snew}(X) = (\tilde{V}^{-1} + n\lambda\psi)^{-1} \tilde{V}^{-1} Y^* \quad (8)$$

λ : معلمة تمهيد التي تتغير بتغير الدوال وليس بتغير العقد وتكون $\lambda > 0$

حيث ان C مصفوفة من الدرجة $n \times n - 2$

[

$$C_{n \times n-2} = [C_{ij}]$$

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

ومصفوفة الـ Q من الدرجة $n-2 \times n-2$

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

$$Q = [q_{ij}] \quad i, j = 2, \dots, N-1$$

$\hat{\theta}_{snew}(X)$: تمثل مقدر ممد الشريحة بطريقة المربعات الصغرى المعممة باستخدام دالة وزن Huber لمقدر M الحصين. ونتبع نفس خطوات الخوارزمية المذكورة في الطريقة المقترحة الاولى للوصول الى المقدر في

الصيغة اعلاه. وتستمر عملية التكرار الى ان يكون مقدر الشريحة الحصين كالاتي^[8]:

$$\|\hat{\theta}_{snew}(X_{ij})^{(I)} - \hat{\theta}_{snew}(X_{ij})^{(I-1)}\| \leq \hat{\theta}_{snew}(X_{ij})^{(I-1)} \quad (9)$$

انواع مصفوفة الارتباط:

وتأخذ مصفوفة الارتباط R عدة انواع:

1-الاستقلالية (*Independence*) (تعني افتراض قيمة معاملات الارتباط بين اي متغيرين توضيحيين داخل العنقود يساوي صفر)

$$\rho(Y_{it}, Y_{is}) = 0, \quad t \neq s \quad (10)$$

2-متبادلة (*Exchangeable*) (تعني افتراض قيمة معاملات الارتباط بين اي متغيرين توضيحيين متساوية)

$$\rho(Y_{it}, Y_{is}) = \alpha, \quad t \neq s \quad (11)$$

3-AR(1) (يعني ان معاملات الارتباط نقل او تتلاشى عبر الزمن)

$$\rho(Y_{it}, Y_{is}) = \alpha^{|t-s|}, \quad t \neq s \quad (12)$$

4-غيرمقيدة (*Unstructured*) (تعني ان قيمة معامل الارتباط بين اي متغيرين توضيحيين داخل العنقود تختلف عن المتغيرين الاخرين ضمن نفس العنقود)^[6]

$$\rho(Y_{it}, Y_{is}) = \alpha_{st} = \alpha_{ts}, \quad t \neq s \quad (13)$$

معيار المقارنة:

يستخدم معيار المقارنة للتعرف على افضل طريقة للتقدير , من خلال المقارنة باقل متوسط مربعات الخطا او اقل متوسط القيمة المطلقة للأخطاء والذان تكون صيغة كل منهما بالشكل التالي:

$$MSE[\hat{\theta}(X)] = \text{mean}[\theta(X) - \hat{\theta}(X)]^2 \quad (14)$$

$$MAE[\hat{\theta}(X)] = \text{mean}|\theta(X) - \hat{\theta}(X)| \quad (14)$$

3- الجانب التجريبي: تم استعمال اسلوب المحاكاة (Simulation) لمقارنة طرائق تقدير الدالة اللامعلمية للبيانات العنقودية والذي يمكننا من خلاله ان نحكي عدد كبير من الحالات المفترضة التي يمكن ظهورها ضمن الواقع العملي لذلك تصبح النتائج ذا شمولية اكثر.

توليد المتغيرات العشوائية:

يتم تنفيذ تجارب المحاكاة باستعمال عدد عناقيد ($n = 30, n = 60$) وحجم كل عنقود ($m = 3$) وتمثل القياسات المتكررة لكل عنقود وبذلك يكون لدينا اثنان من حجوم العينات وكالاتي: ($nm = 180$) و ($nm = 90$) وبواقع 500 تكرار ($Re plicates = 500$) لكل تجربة محاكاة.

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

ويتم توليد المتغيرات التوضيحية X_{ij} بصورة مستقلة من التوزيع المنتظم $U[-1,1]$ والتي تتبع النموذج (1) أذ أن:

$$X_i = (x_{i1}, \dots, x_{ij})' \quad i = 1, 2, \dots, n$$

$$j = 1, 2, \dots, m$$

$$[E_i = (E_{i1}, \dots, E_{ij})']$$

أذ أن $E_{ij} \sim MVN[0, \Sigma]$ وتوليد بيانات ملوثة تتبع توزيع الخطأ التالي $E_{ij} \sim MVN[0, \alpha^2 \Sigma]$ و ان $\alpha^2 > 1$. اما بالنسبة لمتغيرات الاستجابة Y_{ij} في النموذج (1) تحدد وفق $\theta(X_{ij})$ المحددة بواسطة النماذج التالية بمختلف المنحنيات [3] [4] [9]:

$$1 - \theta(x) = \exp(x)$$

$$2 - \theta(x) = 2 \sin(2\pi x)$$

$$3 - \theta(x) = \sqrt{z(1-z)} \sin\{2\pi(1 + 2^{-3/5})/(z + 2^{-3/5})\}$$

$$4 - \theta(x) = \sin(8z - 4) + \exp\{-265(z - .5)^2\}$$

ومن الجدير بالذكر ان :

$$z = \frac{x + 2}{4}$$

فضلاً عن النماذج المذكورة اعلاه فقد تم استخدام انواع متعددة من مصفوفة الارتباط ومنها:

- | | |
|---|-------|
| 1 – (Independence) | case1 |
| 2 – (Exchangeable with Correlation ρ) | case2 |
| 3 – (Autoregressive with Correlation ρ) | case3 |
| 4 – (Unstructured with Correlation ρ) | case4 |

كما في المعادلات (10)(11)(12)(13) المذكورة في الجانب النظري حيث تم اخذ الحالة الثانية بـ

$p = 0.6$ وفي الحالة الثالثة بـ $p = 0.6$ وفي الحالة الرابعة $p_{12} = p_{23} = 0.8, p_{13} = 0.2$ [10]

مع استخدام نسب تلويث مختلفة 20%, 30%. وحالة عدم وجود التلويث ولجميع النماذج المذكورة والطرق التي تم التطرق لها في الجانب النظري .

تحليل النتائج:

الجدول رقم (1): جدول يمثل قيمة الـ MSE و MAE للطرق المذكورة في الجانب النظري عندما يكون عدد العناقيد $(n = 30)$ وحجم كل عنقود $(m = 3)$ ليكون حجم العينة الكلي $(nm = 90)$ ولجميع حالات الارتباط وينسب تلويث مختلفة للنموذج الاول

METHOD	CASE 1			CASE 2			CASE 3			CASE 4		
	0%	20%	30%	0%	20%	30%	0%	20%	30%	0%	20%	30%
KERNEL	0.8292	1.967		0.9413	2.074		0.937	2.075		0.943	2.081	
		2.562 (0.719)		2.744			2.738			2.748		
		(1.004) (1.158)		(0.769) (1.036)			(0.768) (1.037)			(0.77) (1.036)		
				(1.190)			(1.189)			(1.191)		

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

SPLINE	0.879 2.004 2.665 (0.749) (1.002) (1.172)	0.888 2.004 2.662 (0.751) (1.021) (1.171)	0.887 2.007 2.670 (0.750) (1.002) (1.173)	0.886 2.051 2.670 (0.749) (1.025) (1.173)
Robust kernel	0.845 1.870 2.541 (0.726) (0.991) (1.152)	0.948 2.077 2.740 (0.773) (1.037) (1.192)	0.948 2.075 2.738 (0.773) (1.038) (1.192)	0.952 2.077 2.745 (0.774) (1.038) (1.193)
Robust spline	0.887 2.0111 2.675 (0.752) (1.025) (1.176)	0.896 2.011 2.674 (0.756) (1.025) (1.176)	0.984 2.014 2.683 (0.753) (1.026) (1.178)	0.888 1.848 2.669 (0.753) (0.985) (1.175)

حيث القيمة بدون اقواس تمثل الـ MSE والقيمة التي بين الاقواس تمثل الـ MAE

1- يوضح الجدول (1) ان طريقة الـ *Seemingly unrelated kernel estimator* أفضل من غيرها في حالة عدم وجود تلويث ضمن حالة الـ CASE 1 بالاعتماد على معياري المقارنة MSE, MAE , فيما تحقق طريقة مقدر الـ *Robust generalized least squares smoothing spline estimator* اقل متوسط مربعات الخطأ واقل MAE في حالة نسبة التلويث 20% ضمن حالة الـ CASE 4 و حققت طريقة الـ *Robust Seemingly unrelated kernel estimator* اقل متوسط مربعات للخطأ واقل MAE في حالة نسبة التلويث 30% ضمن الـ CASE 1

الجدول رقم(2) : جدول يمثل قيمة الـ MSE و MAE للطرق المذكورة في الجانب النظري عندما يكون عدد العناقيد ($n = 60$) وحجم كل عنقود ($m = 3$) ليكون حجم العينة الكلي ($nm = 180$) ولجميع حالات الارتباط وينسب تلويث مختلفة للنموذج الاول

METHOD	CASE 1			CASE 2			CASE 3			CASE 4		
	0%	20%	30%	0%	20%	30%	0%	20%	30%	0%	20%	30%
KERNEL	1.07384	2.3451		1.07284	2.394		1.0932	2.395		1.0959	2.528	
	2.9845			3.0312			3.0322			3.154		
	(0.829)	(1.123)		(0.829)	(1.138)		(0.836)	(1.136)		(0.836)	(1.166)	
	(1.1986)			(1.223)			(1.222)			(1.253)		
SPLINE	1.099	2.403	3.043	1.0947	2.3933	0.605	1.0947	2.395		1.090	2.406	
	(0.842)	(1.136)	(1.217)	(0.840)	(1.135)		3.059			3.047		
				(1.224)			(0.839)	(1.135)		(0.840)	(1.136)	
							(1.223)			(1.219)		
Robust kernel	1.077	2.361		1.103	2.402		1.1022	2.401		1.165	2.406	
	2.9842			3.050			3.040			4.046		
	(0.8311)	(1.128)		(0.842)	(1.142)		(0.841)	(1.140)		(0.860)	(1.141)	
	(1.1980)			(1.231)			(1.225)			(1.53)		
Robust spline	1.1082	2.394	3.045	1.102	2.396		1.102	2.403		1.098	2.394	
	(0.847)	(1.137)		3.053			3.053			3.048		
	(1.214)			(0.844)	(1.136)		(0.843)	(1.137)		(0.8422)		
				(1.218)			(1.214)			(1.137)(1.214)		

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

2- يوضح الجدول (2) أن طريقة Seemingly unrelated kernel estimators أفضل من غيرها في حالة عدم وجود تلويث ضمن حالة الـ CASE2 و كانت قيمة كل من MAE , MSE أقل ما يمكن، فيما تحقق طريقة Seemingly unrelated kernel estimators أقل متوسط مربعات الخطأ وأقل MAE في حالة نسبة التلويث 20% ضمن حالة الـ CASE 1 و حققت طريقة مقدر Robust Seemingly unrelated kernel estimators أقل متوسط مربعات للخطأ وأقل MAE في حالة نسبة التلويث 30% ضمن الحالة الـ CASE 1 الجدول رقم (3) يمثل قيمة الـ MSE و MAE للطرق المذكورة في الجانب النظري عندما يكون عدد العناقيد ($n = 30$) وحجم كل عنقود ($m = 3$) ليكون حجم العينة الكلي ($nm = 90$) ولجميع حالات الارتباط وينسب تلويث مختلفة للنموذج الثاني .

METHOD	CASE 1 0% 20% 30%			CASE 2 0% 20% 30%			CASE 3 0% 20% 30%			CASE 4 0% 20% 30%		
KERNEL	2.902	1.917		3.229	2.843		3.215	2.850		3.258	2.895	
	2.625			3.561			3.527			4.216		
	(1.259)	(1.012)		(1.353)	(1.286)		(1.351)	(1.299)		(1.361)	(1.302)	
	(1.187)			(1.439)			(1.430)			(1.598)		
SPLINE	2.670	1.903		2.647	1.876		2.648	1.88		2.630	1.906	
	2.4023			2.4016			2.406			2.4017		
	(1.203)	(1.011)		(1.200)	(1.004)		(1.199)	(1.007)		(1.197)	(1.013)	
	(1.127)			(1.127)			(1.128)			(1.127)		
Robust kernel	5224.3	18.33		1344.6	16.308		1276.6	29.00		1496.3	15.86	
	41.713			20.30			22.484			19.83		
	(64.74)	(2.649)		(32.6)	(2.797)		(31.9)	(3.670)		(34.6)	(2.820)	
	(3.807)			(2.780)			(2.984)			(2.804)		
Robust spline	4.597	1.983		3.305	1.985		3.406	1.906		2.707	2.017	
	4.931			4.628			5.504			3.557		
	(1.692)	(1.045)		(1.394)	(1.017)		(1.423)	(1.015)		(1.227)	(1.046)	
	(1.348)			(1.312)			(1.397)			(1.269)		

3- يوضح الجدول (3) إن طريقة generalized least squares smoothing spline estimator أفضل من غيرها في حالة عدم وجود تلويث ضمن حالة الـ CASE 4 و كانت قيمة كل من MAE , MSE أقل ما يمكن، فيما تحقق طريقة generalized least squares smoothing spline estimator أقل متوسط مربعات الخطأ وأقل MAE في حالة نسبة التلويث 20% و 30% ضمن حالة الـ CASE2. الجدول رقم (4) يمثل قيمة الـ MSE و MAE للطرق المذكورة في الجانب النظري عندما يكون عدد العناقيد ($n = 60$) وحجم كل عنقود ($m = 3$) ليكون حجم العينة الكلي ($nm = 180$) ولجميع حالات الارتباط وينسب تلويث مختلفة للنموذج الثاني

METHOD	CASE 1 0% 20% 30%			CASE 2 0% 20% 30%			CASE 3 0% 20% 30%			CASE 4 0% 20% 30%		
KERNEL	1.140	2.067		1.646	2.842		1.689	2.8404		1.7218	2.918	
	2.551			2.631			2.6228			2.939		
	(0.845)			(1.045)			(1.0572)			(1.070)	(1.294)	
	(1.070)	(1.176)		(1.276)	(1.174)		(1.1275)	(1.173)		(1.277)		

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

SPLINE	0.9618 1.729 2.557 (0.760) (0.99) (1.171)	0.943 1.695 2.533 (0.753) (0.984) (1.170)	0.9463 1.696 2.5362 (0.753) (0.983) (1.169)	0.9310 1.738 2.5078 (0.747) (0.996) (1.1687)
Robust kernel	595.03 17.49 2.546 (22.109) (3.67) (1.184)	5.7799 3.758 2.649 (1.999) (1.50) (1.174)	516.95 4.304 2.639 (18.90) (1.619) (1.171)	1.937 4.469 2.649 (1.104) (1.624) (1.174)
Robust spline	2.238 1.732 2.564 (1.227) (1.012) (1.177)	0.946 1.693 2.523 (0.757) (0.982) (1.176)	0.9591 1.707 2.527 (0.7525) (0.986) (1.175)	0.952 1.697 2.4987 (0.7617) (0.987) (1.1680)

4- يوضح الجدول (4) أن طريقة generalized least squares smoothing spline أفضل من غيرها في حالة عدم وجود تلويث ضمن حالة الـ CASE 4 وكانت قيمة كل من MAE , MSE أقل ما يمكن، فيما تحقق طريقة Robust generalized least squares smoothing spline أقل متوسط مربعات الخطأ في حالة نسبة التلويث 20% وأقل MAE ضمن حالة الـ CASE 2 وحققت طريقة Robust generalized least squares smoothing spline أقل متوسط مربعات الخطأ في حالة نسبة التلويث 30% وأقل MAE ضمن حالة الـ CASE4.

الجدول رقم (5): جدول يمثل قيمة الـ MSE و MAE للطرق المذكورة في الجانب النظري عندما يكون عدد العناقيد ($n = 30$) وحجم كل عنقود ($m = 3$) ليكون حجم العينة الكلي ($nm = 90$) ولجميع حالات الارتباط وينسب تلويث مختلفة للنموذج الثالث.

METHOD	CASE 1 0% 20% 30%	CASE 2 0% 20% 30%	CASE 3 0% 20% 30%	CASE 4 0% 20% 30%
KERNEL	0.9754 2.046 2.683 (0.775) (1.028) (1.155)	0.973 2.1248 2.7401 (0.771) (1.034) (1.167)	0.9714 2.1297 2.7272 (0.770) (1.0412)(1.167)	0.975 2.134 2.750 (0.772) (1.020)(1.77)
SPLINE	0.9712 2.046 2.684 (0.7723)(1.026)(1.156)	0.970 2.0554 2.677 (0.7724)(1.029)(1.154)	0.967 2.0611 2.7304 (0.7698)(1.030) (1.171)	0.961 2.0479 2.685 (0.767) (1.026) (1.157)
Robust kernel	0.977 2.064 2.629 (0.779) (1.024) (1.152)	0.9785 2.114 2.740 (0.774) (1.032) (1.169)	0.979 2.073 2.697 (0.774) (1.027) (1.157)	0.998 2.063 2.729 (0.782)(1.021)(1.166)
Robust spline	0.977 2.057 2.687 (0.773)(1.034)(1.158)	0.977 2.060 2.597 (0.774) (1.035) (1.143)	0.975 2.066 2.737 (0.772)(1.036)(1.174)	0.969 2.0410 2.689 (0.770)(1.020)(1.160)

5- يوضح الجدول (5) إن طريقة generalized least squares smoothing spline أفضل من غيرها في حالة عدم وجود تلويث ضمن حالة الـ CASE 4 وكانت قيمة كل من MAE , MSE أقل ما يمكن في ما تحقق طريقة مقدر Robust generalized least squares smoothing spline أقل متوسط مربعات الخطأ

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

وأقل MAE في حالة نسبة التلويث 20% ضمن حالة الـ CASE 4 و حققت طريقة الـ Robust generalized least squares smoothing spline أقل متوسط مربعات للخطأ وأقل MAE في حالة نسبة التلويث 30% ضمن حالة الـ CASE 2.

الجدول رقم (6) : جدول يمثل قيمة الـ MSE و MAE للطرق المذكورة في الجانب النظري عندما يكون عدد العناقيد ($n = 60$) وحجم كل عنقود ($m = 3$) ليكون حجم العينة الكلي ($nm = 180$) ولجميع حالات الارتباط وينسب تلويث مختلفة للنموذج الثالث.

METHOD	CASE 1 0% 20% 30%	CASE 2 0% 20% 30%	CASE 3 0% 20% 30%	CASE 4 0% 20% 30%
KERNEL	1.055 3.658 2.727 (0.810) (1.311) (1.216)	1.039 3.616 2.705 (0.804) (1.345) (1.218)	1.0382 3.652 2.702 (0.803) (1.34) (1.214)	1.060 3.640 2.784 (0.813) (1.335) (1.230)
SPLINE	1.0578 3.655 2.723 (0.810) (1.310) (1.216)	1.0571 3.646 2.738 (0.810) (1.310) (1.216)	1.0578 3.642 2.740 (0.811) (1.308) (1.214)	0.612 3.642 2.740 (0.625) (1.310) (1.214)
Robust kernel	0.619 3.624 2.667 (0.632) (1.34) (1.209)	1.0411 3.658 2.698 (0.804) (1.33) (1.224)	1.0393 3.647 2.694 (0.803) (1.337) (1.225)	1.042 3.657 2.696 (0.804) (1.340) (1.225)
Robust spline	1.0589 3.654 2.729 (0.811) (1.310) (1.217)	1.0588 3.639 2.736 (0.811) (1.311) (1.214)	1.0586 3.634 2.666 (0.811) (1.312) (1.206)	1.059 3.554 2.737 (0.811) (1.303) (1.214)

6- يوضح الجدول (6) أن طريقة generalized least squares smoothing spline أفضل من غيرها في حالة عدم وجود تلويث ضمن حالة الـ CASE 4 وكانت قيمة كل من MSE , MAE أقل ما يمكن , فيما تحقق طريقة مقدر Robust generalized least squares smoothing spline أقل متوسط مربعات الخطأ في حالة نسبة التلويث 20% ضمن حالة الـ CASE 4 وأقل MAE و حققت طريقة الـ Robust generalized least squares smoothing spline أقل متوسط مربعات للخطأ وأقل MAE في حالة نسبة التلويث 30% ضمن حالة الـ CASE 3.

الجدول رقم (7) : جدول يمثل قيمة الـ MSE و MAE للطرق المذكورة في الجانب النظري عندما يكون عدد العناقيد ($n = 30$) وحجم كل عنقود ($m = 3$) ليكون حجم العينة الكلي ($nm = 90$) ولجميع حالات الارتباط وينسب تلويث مختلفة للنموذج الرابع.

METHOD	CASE 1 0% 20% 30%	CASE 2 0% 20% 30%	CASE 3 0% 20% 30%	CASE 4 0% 20% 30%
KERNEL	0.880 2.180 2.515 (0.748) (1.083) (1.154)	1.024 1.962 2.794 (0.807) (1.03) (1.213)	1.007 2.175 2.784 (0.802) (1.083) (1.212)	1.171 2.165 2.802 (0.870) (1.077) (1.216)
SPLINE	0.860 2.014 2.508 (0.740) (1.048)	0.866 1.968 2.510 (0.743) (1.025)	0.883 2.006 2.518 (0.750) (1.042)	0.876 2.017 2.508 (0.748) (1.050)

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

	(1.148)	(1.149)	(1.151)	(1.148)
Robust kernel	0.899 1.872 2.620 (0.756) (1.019) (1.164)	1.0514 2.194 3.245 (0.820) (1.093) (1.274)	1.010 2.191 3.040 (0.804) (1.095) (1.257)	1.026 2.204 3.871 (0.8105) (1.098) (1.327)
Robust spline	0.877 1.971 2.515 (0.749) (1.032) (1.154)	0.885 2.032 2.517 (0.752) (1.056) (1.156)	0.9022 1.859 2.5455 (0.759) (1.011) (1.162)	0.894 2.019 2.497 (0.756) (1.048) (1.145)

7- يوضح الجدول (6) ان طريقة generalized least squares smoothing spline افضل من غيرها في حالة عدم وجود تلويث ضمن حالة الـ CASE 1 وكانت قيمة كل من MAE , MSE اقل ما يمكن، فيما تحقق طريقة Robust generalized least squares smoothing spline اقل متوسط مربعات الخطأ و اقل MAE في حالة نسبة التلويث 20% ضمن حالة الـ CASE 3 وحقت طريقة الـ Robust generalized least squares smoothing spline اقل متوسط مربعات للخطأ و اقل MAE في حالة نسبة التلويث 30% ضمن الحالة الـ CASE 4

الجدول رقم (8): جدول يمثل قيمة الـ MSE و MAE للطرق المذكورة في الجانب النظري عندما يكون عدد العناقيد ($n = 60$) وحجم كل عنقود ($m = 3$) ليكون حجم العينة الكلي ($nm = 180$) ولجميع حالات الارتباط وينسب تلويث مختلفة للنموذج الرابع.

METHOD	CASE 1 0% 20% 30%	CASE 2 0% 20% 30%	CASE 3 0% 20% 30%	CASE 4 0% 20% 30%
KERNEL	0.989 2.639 2.970 (0.8017) (1.209) (1.248)	1.037 2.869 3.042 (0.821) (1.127) (1.269)	1.036 2.821 3.043 (0.820) (1.257) (1.269)	1.0415 2.863 3.126 (0.8218) (1.272) (1.304)
SPLINE	0.9657 2.492 2.955 (0.7955) (1.172) (1.245)	0.9546 2.456 2.928 (0.792) (1.165) (1.238)	0.9555 2.4599 2.930 (0.793) (1.165) (1.238)	0.9685 2.450 2.975 (0.796) (1.169) (1.251)
Robust kernel	1.002 3.382 2.968 (0.805) (1.439) (1.256)	1.0313 2.849 3.064 (0.8100) (1.268) (1.265)	1.05155 2.869 3.0709 (0.8108) (1.260) (1.267)	1.066 2.833 3.087 (0.8814) (1.257) (1.270)
Robust spline	0.969 2.933 2.954 (0.798) (1.322) (1.248)	0.9548 2.566 2.9314 (0.794) (1.208) (1.243)	0.957 2.580 2.939 (0.796) (1.212) (1.245)	0.9571 2.421 2.939 (0.795) (1.118) (1.246)

8- يوضح الجدول (8) أن طريقة generalized least squares smoothing spline أفضل من غيرها في حالة عدم وجود تلويث ضمن حالة الـ CASE 2 وكانت قيمة كل من MAE , MSE اقل ما يمكن ، فيما تحقق طريقة Robust generalized least squares smoothing spline اقل متوسط مربعات الخطأ في حالة نسبة التلويث 20% و اقل MAE ضمن حالة الـ CASE 4 وحقت طريقة generalized least squares smoothing spline estimators اقل متوسط مربعات للخطأ في حالة نسبة التلويث 30% و اقل MAE ضمن الحالة الـ CASE 2.

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

الاستنتاجات:

- (أ) حققت طريقة الـ generalized least squares smoothing spline estimator أقل MSE, MAE في حالة عدم وجود التلويث مقارنة مع طريقة الـ Seemingly unrelated kernel estimator وذلك ضمن حالة الارتباط CASE4 .
- (ب) حققت طريقة الـ Robust generalized least squares smoothing spline estimator أقل MSE, MAE في حالتي التلويث بنسبة 20%, 30% مقارنة مع بقية الطرق وذلك في حالات متنوعة من الارتباط .
- (ج) لم تكن طريقة الـ Robust Seemingly unrelated kernel estimator جيدة مقارنة بطريقة الـ Seemingly unrelated kernel estimator الاعتيادية .

التوصيات:

- (أ) يوصى باستخدام طرق حصينة أخرى مثل طريقة (S) الحصينة وطريقة (R).
- (ب) نوصي بدراسة الحالة عندما يكون هناك ارتباط بين الأخطاء وتطبيق نفس الطرق عليها.

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

المصادر العربية:

- 1- حسين, سجي محمد وكاظم ,حلا (2016) ((مقارنة طريقتي تقدير الدالة اللامعلمية لبيانات عنقودية عن كريات الدم البيضاء لمرضى اللوكيميا)) مجلة العلوم الاقتصادية والادارية (مقبول النشر).

المصادر الإلكترونية:

- 2- https://catalyst.harvard.edu/.../Fitzmaurice_BSP-Worksh

المصادر الأجنبية:

- 3- LIN,X.&CARROLL,R.J.(2001).((Semi parametric regression for clustered data using generalized estimation equations)) J.Am.Statist.Assoc.VOL.96,No.455,1045-1056.
- 4- Lin,X.,Wang,N.,Welsh,A.H. and Carroll ,R.J.(2004).((Equivalent kernels of smoothing spline in nonparametric regression for clustered/longitudinal data)). Biometrika VOL.91,No.1,177-193.
- 5- Ma, S., Song, Q. and Wang,L.,2013,((Simultaneous variable selection and estimation in semi parametric modeling of longitudinal/clustered data))Bernoulli VOL.19, No.1,252-274
- 6- Nicoleta Breaz (2004)((the cross-validation method in the smoothing spline regression)) universities Apuleius.
- 7- Ruckstuhl,A.,Welsh,A.H.&Carroll,R.J.(2000).((Nonparametric function estimation of the relationship between two repeatedly measured variables)).Statist. Sinica , No.10,51-71.
- 8- Wang.B., Wenzhong ,shi .,&zelang Miao(2014) ((Comparative Analysis for Robust Penalized Spline Smoothing Methods))
- 9- Wang,N.(2003)((Marginal nonparametric Kernel regression accounting for within-subject correlation)),Biometrika,VOL.90,43-52.MR 1966549
- 10- Wang,Y.G.& Zhao,Y.(2008)((Weighted rank regression for clustered data analysis)). Biometrics.VOL.64,No. 1 ,39-45
- 11- Zeger,S.,L.,2007((Modeling Regression and Association with Clustered Data))

ملحق رقم (1)

برنامج المحاكاة

```
clc
clear
itn=500;
n1=24;
n2=6;
n=n1+n2;
m=3;
mu1 = [0,0,0];
mu2 = [0,0,0];
sigma1 = [1 0 0;0 1 0;0 0 1];
sigma2 = 7.*[1 0 0;0 1 0;0 0 1];
corry0=[1 0 0;0 1 0;0 0 1];
corry1=[1 0.6 0.6;0.6 1 0.6;0.6 0.6 1];
corry2=[1 0.6 0.36;0.6 1 0.6;0.36 0.6 1];
corry3=[1 0.8 0.2;0.8 1 0.8;0.2 0.8 1];
Kmse00(itn,:)=0;
Kmae00(itn,:)=0;
Smse00(itn,:)=0;
Smae00(itn,:)=0;
mKmse00(itn,:)=0;
mKmae00(itn,:)=0;
mSmse00(itn,:)=0;
mSmae00(itn,:)=0;
Kmse01(itn,:)=0;
Kmae01(itn,:)=0;
Smse01(itn,:)=0;
Smae01(itn,:)=0;
mKmse01(itn,:)=0;
mKmae01(itn,:)=0;
mSmse01(itn,:)=0;
mSmae01(itn,:)=0;
Kmse02(itn,:)=0;
Kmae02(itn,:)=0;
Smse02(itn,:)=0;
Smae02(itn,:)=0;
mKmse02(itn,:)=0;
mKmae02(itn,:)=0;
mSmse02(itn,:)=0;
mSmae02(itn,:)=0;
```



```

Kmse03(itn,:)=0;
Kmae03(itn,:)=0;
Smse03(itn,:)=0;
Smae03(itn,:)=0;
mKmse03(itn,:)=0;
mKmae03(itn,:)=0;
mSmse03(itn,:)=0;
mSmae03(itn,:)=0;
    xx=-1+2.*rand(n,m);
    e1 = mvnrnd(mu1,sigma1,n1);
    e2 = mvnrnd(mu2,sigma2,n2);
    e=[e1;e2];
    %fx1=exp(xx);
    %fx2=2*sin(2*pi*xx);
    %z=(xx+2)/4;fx3=sqrt(z.*(1-z)).*sin(2*pi*(1+2^(-3/5)).*(z+2^(-3/5)).^(-1));
    z=(xx+2)/4;fx4=sin(8*z-4)+2*exp(-256*(z-0.5).^2);
    yy=fx4+e;
    yy1=[yy(1:n/m)];
    yy3=[yy(((n/m)+1):(2*(n/m)))];
    yy5=[yy((2*(n/m)+1):n)];
    subj=1:m;
    subj=repmat(subj,n,1);
    subj=subj(:);
    x=reshape(xx',n.*m,1);
    y=reshape(yy',n.*m,1);
    data=[subj,x,y];
    dpoly=2;
    dy1=std(yy1);
    dy2=std(yy3);
    dy3=std(yy5);
    dy=[dy1 dy2 dy3];
    sdy=diag(dy);
    dsd01=repmat(dy1,n);
    dsd1=diag((dsd01));
    dsd02=repmat(dy2,n);
    dsd2=diag((dsd02));
    dsd03=repmat(dy3,n);
    dsd3=diag((dsd03));
    dsdy=[dsd1' dsd2' dsd3'];
    dsdy=diag(dsdy);
    eyey=eye(n);

```

```
cdsdy0=kron(eyey,corry0);
V0=dsdy*cdsdy0*dsdy;
Vd10=pinv(diag(diag(V0)));
Vd20=diag(diag(pinv(V0)));
[Kyhat00,B00]=glpfitsim(data,V0,Vd10,Vd20,[0,dpoly]);
[mKyhat00] = Ro(B00,Kyhat00);
[Syhat00,A00]=ssfitsim(data,[],Vd20);
[mSyhat00] = Ro(A00,Syhat00);
Kmse00(it,:)=mean((y-Kyhat00).^2);
Kmae00(it,:)=mean(abs(y-Kyhat00));
Smse00(it,:)=mean((y-Syhat00).^2);
Smae00(it,:)=mean(abs(y-Syhat00));
mKmse00(it,:)=mean((y-mKyhat00).^2);
mKmae00(it,:)=mean(abs(y-mKyhat00));
mSmse00(it,:)=mean((y-mSyhat00).^2);
mSmae00(it,:)=mean(abs(y-mSyhat00));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
cdsdy1=kron(eyey,corry1);
V1=dsdy*cdsdy1*dsdy;
Vd11=inv(diag(diag(V1)));
Vd21=diag(diag(inv(V1)));
[Kyhat01,B01]=glpfitsim(data,V1,Vd11,Vd21,[0,dpoly]);
[mKyhat01] = Ro(B01,Kyhat01);
[Syhat01,A01]=ssfitsim(data,[],Vd21);
[mSyhat01] = Ro(A01,Syhat01);
Kmse01(it,:)=mean((y-Kyhat01).^2);
Kmae01(it,:)=mean(abs(y-Kyhat01));
Smse01(it,:)=mean((y-Syhat01).^2);
Smae01(it,:)=mean(abs(y-Syhat01));
mKmse01(it,:)=mean((y-mKyhat01).^2);
mKmae01(it,:)=mean(abs(y-mKyhat01));
mSmse01(it,:)=mean((y-mSyhat01).^2);
mSmae01(it,:)=mean(abs(y-mSyhat01));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
cdsdy2=kron(eyey,corry2);
V2=dsdy*cdsdy2*dsdy;
Vd12=inv(diag(diag(V2)));
Vd22=diag(diag(inv(V2)));
[Kyhat02,B02]=glpfitsim(data,V2,Vd12,Vd22,[0,dpoly]);
[mKyhat02] = Ro(B02,Kyhat02);
[Syhat02,A02]=ssfitsim(data,[],Vd22);
[mSyhat02] = Ro(A02,Syhat02);
```

```

Kmse02(it,:)=mean((y-Kyhat02).^2);
Kmae02(it,:)=mean(abs(y-Kyhat02));
Smse02(it,:)=mean((y-Syhat02).^2);
Smae02(it,:)=mean(abs(y-Syhat02));
mKmse02(it,:)=mean((y-mKyhat02).^2);
mKmae02(it,:)=mean(abs(y-mKyhat02));
mSmse02(it,:)=mean((y-mSyhat02).^2);
mSmae02(it,:)=mean(abs(y-mSyhat02));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
cdsdy3=kron(eyey,corry3);
V3=dsdy*cdsdy3*dsdy;
Vd13=pinv(diag(diag(V3)));
Vd23=diag(diag(pinv(V3)));
[Kyhat03,B03]=glpfitsim(data,V3,Vd13,Vd23,[0,dpoly]);
[mKyhat03] = Ro(B03,Kyhat03);
[Syhat03,A03]=ssfitsim(data,[],Vd23);
[mSyhat03] = Ro(A03,Syhat03);
Kmse03(it,:)=mean((y-Kyhat03).^2);
Kmae03(it,:)=mean(abs(y-Kyhat03));
Smse03(it,:)=mean((y-Syhat03).^2);
Smae03(it,:)=mean(abs(y-Syhat03));
mKmse03(it,:)=mean((y-mKyhat03).^2);
mKmae03(it,:)=mean(abs(y-mKyhat03));
mSmse03(it,:)=mean((y-mSyhat03).^2);
mSmae03(it,:)=mean(abs(y-mSyhat03));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Koutput00=[mean(Kmse00) mean(Kmae00)];
Soutput00=[mean(Smse00) mean(Smae00)];
mKoutput00=[mean(mKmse00) mean(mKmae00)];
mSoutput00=[mean(mSmse00) mean(mSmae00)];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Koutput01=[mean(Kmse01) mean(Kmae01)];
Soutput01=[mean(Smse01) mean(Smae01)];
mKoutput01=[mean(mKmse01) mean(mKmae01)];
mSoutput01=[mean(mSmse01) mean(mSmae01)];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
Koutput02=[mean(Kmse02) mean(Kmae02)];
Soutput02=[mean(Smse02) mean(Smae02)];

```

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

```
mKoutput02=[mean(mKmse02) mean(mKmae02)];
mSoutput02=[mean(mSmse02) mean(mSmae02)];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
Koutput03=[mean(Kmse03) mean(Kmae03)];
Soutput03=[mean(Smse03) mean(Smae03)];
mKoutput03=[mean(mKmse03) mean(mKmae03)];
mSoutput03=[mean(mSmse03) mean(mSmae03)];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
xlswrite('D:\Sim_Hala.xlsx',Koutput00,'output_fx4_30_3_20','A1:B1');
xlswrite('D:\Sim_Hala.xlsx',Soutput00,'output_fx4_30_3_20','C1:D1');
xlswrite('D:\Sim_Hala.xlsx',mKoutput00,'output_fx4_30_3_20','E1:F1');
xlswrite('D:\Sim_Hala.xlsx',mSoutput00,'output_fx4_30_3_20','G1:H1');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
xlswrite('D:\Sim_Hala.xlsx',Koutput01,'output_fx4_30_3_20','A4:B4');
xlswrite('D:\Sim_Hala.xlsx',Soutput01,'output_fx4_30_3_20','C4:D4');
xlswrite('D:\Sim_Hala.xlsx',mKoutput01,'output_fx4_30_3_20','E4:F4');
xlswrite('D:\Sim_Hala.xlsx',mSoutput01,'output_fx4_30_3_20','G4:H4');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
xlswrite('D:\Sim_Hala.xlsx',Koutput02,'output_fx4_30_3_20','A7:B7');
xlswrite('D:\Sim_Hala.xlsx',Soutput02,'output_fx4_30_3_20','C7:D7');
xlswrite('D:\Sim_Hala.xlsx',mKoutput02,'output_fx4_30_3_20','E7:F7');
xlswrite('D:\Sim_Hala.xlsx',mSoutput02,'output_fx4_30_3_20','G7:H7');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
xlswrite('D:\Sim_Hala.xlsx',Koutput03,'output_fx4_30_3_20','A10:B10');
xlswrite('D:\Sim_Hala.xlsx',Soutput03,'output_fx4_30_3_20','C10:D10');
xlswrite('D:\Sim_Hala.xlsx',mKoutput03,'output_fx4_30_3_20','E10:F10');
xlswrite('D:\Sim_Hala.xlsx',mSoutput03,'output_fx4_30_3_20','G10:H10');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
xlswrite('D:\Sim_Hala.xlsx',Koutput04,'output_fx4_30_3_20','A13:B13');
xlswrite('D:\Sim_Hala.xlsx',Soutput04,'output_fx4_30_3_20','C13:D13');
xlswrite('D:\Sim_Hala.xlsx',mKoutput04,'output_fx4_30_3_20','E13:F13');
xlswrite('D:\Sim_Hala.xlsx',mSoutput04,'output_fx4_30_3_20','G13:H13');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
xlswrite('D:\Sim_Hala.xlsx',Koutput05,'output_fx4_30_3_20','A16:B16');
xlswrite('D:\Sim_Hala.xlsx',Soutput05,'output_fx4_30_3_20','C16:D16');
xlswrite('D:\Sim_Hala.xlsx',mKoutput05,'output_fx4_30_3_20','E16:F16');
xlswrite('D:\Sim_Hala.xlsx',mSoutput05,'output_fx4_30_3_20','G16:H16');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
xlswrite('D:\Sim_Hala.xlsx',tElapsed,'output_fx4_30_3_20','A20')
xlswrite('D:\Sim_Hala.xlsx',itn,'output_fx4_30_3_20','A21')
```

برنامج لحساب الدالة الحصينة:

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

```
function [yhat] = Ro(B,y)
n=length(y);
cv=1;
c=1.345;
mp=B*y;
r=y-mp;
Mr=median(r);
r1=abs(r-Mr);
MAD=median(r1);
s=1.4836*MAD;
while cv==1,
    % mp=B*y;
    r=y-mp;
    r2=r./s;
    for i=1:n
        psi(i)=max(-c, min(c,r (i)));
    end
    z=mp+n(psi'*MAD)./2;
    mpnew=B*z;
    epsilon= sqrt(((sum(mpnew-mp)).^2)) ;
    if epsilon< sqrt((sum(mp)^2)) ;
        cv=2;
        yhat=mpnew;
    else
        cv=1;
        y=mpnew;
    end
end
clearvars -except yhat
end
```

برنامج لحساب دالة kernel الحصينة

```
function [yhat,B]=glpfitsim(data,V,Vd1,Vd2,params,kstr)
n=size(data,1);
subj=data(:,1);
usubj=unique(subj);
nsubj=length(usubj);
x=data(:,2);
y=data(:,3);
[ux,~,flag2]=unique(x);
nux=length(ux);
xrange=max(ux)-min(ux);
```

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

```
if nargin<2||isempty(V),
    V=sdy*cdsdy*sdy;
end
V=V(flag2,flag2);
if nargin<3||isempty(Vd1),

    V=sdy*cdsdy*sdy;
    Vd1=pinv(diag(diag(V)));
end
V=V(flag2,flag2);
Vd1=Vd1(flag2,flag2);
if nargin<4||isempty(Vd2),
    V=sdy*cdsdy*sdy;
    Vd2=diag(diag(pinv(V)));
end
V=V(flag2,flag2);
Vd2=Vd2(flag2,flag2);
if nargin<5||isempty(params),
    h=0;
    dpoly=1;
elseif length(params)==1,
    h=params;
    dpoly=1;
elseif length(params)==2,
    h=params(1);
    dpoly=params(2);
end
if nargin<7||isempty(kstr),
    kstr='(1/sqrt(2*pi))*exp(-.5*t.^2)';
end
if h~=0,
    nh=1;
else
    nh=10;
    hmin=.5*(dpoly+1)*xrange/nux;
    hmax=xrange/8;
    vh = logspace(log10(hmin),log10(hmax),nh);
end
if nh>1,
    for ii=1:nh,
        h=vh(ii);
        for jj=1:nux,
```

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

```

temp=x-ux(jj);
t=(temp/h)*h^(-1) ;
W=diag(eval(kstr));
X=ones(n,1);
for r=1:dpoly,
    X=[X,temp.^r];
end
AA=0;
BB=[];
for kk=1:nsubj,
    flagi=(subj==usubj(kk));
    Wi=W(flagi,flagi);
    Xi=X(flagi,:);
    yi=y(flagi);
    ni=length(yi);
    Vi=V(flagi,flagi);
    Vd1i=Vd1(flagi,flagi);
    temp=Wi.*diag(diag(pinv(Vi)));
    AA=AA+Xi'*temp*Xi;
    BB=[BB,Xi'*Wi];
end
temp=pinv(AA)*BB;
iii=eye(size(temp));
temp1=pinv(iii+temp*(Vd1-Vd2))*temp*Vd1;
A(jj,:)=temp1(1,:);
end
B=A(flag2,:);
yhat=B*y;
df(ii)=trace(B);
gcv(ii)=mean((y-yhat).^2)/(1-df(ii)/n)^2;
end
vhgcv=[vh(:),gcv(:)];
[gcv,temp]=min(gcv);
h=vh(temp);
hgcgv=vhgcv(temp,:);
end
for jj=1:nux,
    temp=x-ux(jj);
    t=(temp/h)*h^(-1);
    W=diag(eval(kstr));
    X=ones(n,1);

```

دراسة تجريبية لمقارنة طرق حصينة مقترحة لتقدير الدالة اللامعلمية للبيانات العنقودية مع طرق أخرى

```

for r=1:dpoly,
    X=[X,temp.^r];
end
AA=0;
BB=[];
for kk=1:nsubj,
    flagi=(subj==usubj(kk));
    Wi=W(flagi,flagi);
    Xi=X(flagi,:);
    yi=y(flagi);
    ni=length(yi);
    Vi=V(flagi,flagi);
    Vd1i=Vd1(flagi,flagi);
    temp=Wi.*diag(diag(pinv(Vi)));
    AA=AA+Xi'*temp*Xi;
    BB=[BB,Xi'*Wi];
end
temp=pinv(AA)*BB;
iii=eye(size(temp));
temp1=pinv(iii+temp*(Vd1-Vd2))*temp*Vd1;
A(jj,:)=temp1(1,:);
end
efit=A*y;
B=A(flag2,:);
yhat=B*y;
df=trace(B);
dfa=xrange/h/sqrt(2*pi);
resid=y-yhat;
gcv=mean(resid.^2)/(1-df/n)^2;
hsig2=sum(resid.^2)/(n-df);
hgcv=[h,gcv];
uyfit=efit;
uysig=sqrt(hsig2*diag(A*A'));
if nh==1,
    vhgcv=hgcv(1:3);
end
clearvars -except yhat B
end

```

برنامج لحساب دالة الـ spline الحصينة

```

function [yhat,A]=ssfitsim(data,spar,V)
subj=data(:,1);
x=data(:,2);

```



```

y=data(:,3);
n=length(x);
[knots,~,flag2]=unique(x);
nknot=length(knots);
E=zeros(n,nknot);
for i=1:n,
    E(i,flag2(i))=1;
end
S=E'*E;
S=diag(diag(S).^(-1/2));
Gmat=splmat(knots);
[U,D]=eig(S*Gmat*S);
d=diag(real(D));
if nargin<2||isempty(spar),
    nspar=15;
    d0=sort(d);
    dmin=d0(3);
    dmax=max(d);
    sparmax=((nknot-2)/(nknot-4)-1)/dmin;
    sparmin=((nknot-2)/(4-2)-1)/dmax;
    vspar = logspace(log10(sparmin),log10(sparmax),nspar)';
else
    vspar=spar;
    nspar=1;
end
if nargin<3||isempty(V),
    V=eye(n);
end
for k=1:nspar,
    A=E*pinv(E'*V*E+(n)*vspar(k)*Gmat)*E'*V;
    yhat=A*y;
    gcv(k)=mean((y-yhat).^2)/(1-trace(A)/n)^2;
end
vspargcv=[vspar(:),gcv(:)];
[gcv,temp]=min(gcv);
spar=vspar(temp(1));
spargcv=[spar,gcv];
A=E*pinv(E'*V*E+(n)*spar*Gmat)*E'*V;
yhat=A*y;
resid=y-yhat;
SSE=sum((resid).^2);
clearvars -except yhat A
end

```