تحديد أفضل أسلوب نواة تمهيدي حصين لتقدير انموذج انحدار لا معلمي

أ.م.د. فارس طاهر حسن*   م.م. غياث حميد مجيد**

المستخلص:

تمثل تقنية الانحدار اللامعلمي طريقة مختلفة في تحليل الانحدار عن الانحدار المعلمي، لكنه بدوره لا يعني أن استخدام طريقة محل طريقة يمنع من استخدام الأخرى. فأساليب وتقنيات الانحدار اللامعلمي يمكن أن تستخدم في تقييم شرعية الإنموذج المعلمي المفترض والعكس صحيح، وربما شكل مطابقة منحنى الانحدار والحاصل عليه من تقنيات الانحدار اللامعلمي يقترح الإنموذج المعلمي المناسب ليستخدم في الدراسات المستقبلية.

تناول البحث استخدام أسلوب تمهيد حصين للنواة لغرض تقدير انموذج الانحدار اللامعلمي، وبهدف الحصول على ذلك تم الاعتماد على نتائج تجربة المحاكاة والموضحة في الجانب التجريبي، وتمت المقارنة بين الدوال التي تم استخدامها في التمهيد باستخدام معياري (MAE) و(MSE) للوصول إلى أفضل مقدر لتمثيل البيانات التي تم توليدها باستخدامها المحاكاة.

Abstract

Nonparametric regression technique represents a different way in regression analysis of parametric regression technique, but it does not mean that the use of one method prevent the use of the other. The methods of nonparametric regression can be used to assess the legality of supposed nonparametric model and vice versa, and matching the form of regression curve that we have it from nonparametric regression techniques may suggests the appropriate parametric regression model to be used in future studies.

The research contains the use of Robust Kernel Smoothing to estimate a nonparametric regression model, in order to get that we depends on the results of a simulation study that described in experimental side. The comparison was made between the functions that were used in smoothing by using (MAE) and (MSE) criteria to reach the best estimator that represents the data that generated by simulation.

الجانب النظري

الانحدار اللامعلمي Nonparametric Regression

يمثل اسلوب الانحدار اللامعلمي المرحلة النهائية في عملية تحليل البيانات، أو خطوة استكشافية في عمليات النمذجة. يمكن تمثيل انموذج الانحدار بالشكل التالي:[2]

$$Y_i = m(X_i) + \varepsilon_i\ , \qquad i = 1,\dots,n \qquad \text{...................... (1)}$$

* جامعة بغداد / كلية الادارة والاقتصاد .

** الجامعة المستنصرية / كلية الادارة والاقتصاد .

مقبول للنشر بتأريخ 2015/8/30

مستل من أطروحة دكتوراه

حيث أن m تمثل دالة الانحدار المجهولة و ε_i أخطاء المشاهدة العشوائية والتي بدورها تكون غير مرتبطة وتتوزع توزيعاً طبيعياً بمتوسط حسابي صفر وتباين قدره σ^2. والهدف هنا هو تقدير m التي يشار إليها بدالة الانحدار أو منحنى الانحدار عند النقاط $x_1, \ldots, x_n$.

الهدف من تحليل الانحدار هو عمل تحليل سببي معقول لدالة الانحدار المجهولة m . وعن طريق الحد من أخطاء المشاهدة فإنه يتيح لنا التركيز على التفاصيل المهمة للمتوسط بإعتماد Y على X. أسلوب منحنى التقريب هذا يدعى بالتمهيد (Smoothing).

تمهيد النواة (التمهيد اللبّي) Kernel Smoother [4] [5]

زودنا ممهد النواة Kernel بطريقة تمكننا من ايجاد هيكل مجموعة من البيانات دون الحاجة إلى ضرورة وجود إنموذج معلمي، بالتالي ونظراً لوجود المعوقات في استخدام انموذجاً معلمياً لغرض تقدير الدالة m الموضحة في الصيغة (1) فإن الدالة m تكون مبهمة، وتكوّن قيوداً قد تعيق أن يكون التقدير لدالة الانحدار صحيحاً، وبالتالي فإن عملية استخدام انموذجاً معلمياً هنا قد يؤدي بنا إلى استنتاجات غير صحيحة في تحليل انموذج الانحدار. [4]

1. ممهد النواة المتعدد الحدود الموضعي [3] [4] [5] Local Polynomial Kernel Smoother [LPK]

إن فكرة تمهيد النواة المتعدد الحدود الموضعي الرئيسية هي التقريب الموضعي للدالة m باستخدام متعدد حدود من درجة معينة. أساس هذه الفكرة هو توسيع تايلر Taylor Expansion ، والذي ينص بدوره على أن أي دالة تمهيدية يمكن أن تتقارب موضعياً من متعدد الحدود من درجة معينة.

فإذا كانت x_o تمثل نقطة اعتباطية ثابتة زمنية بحيث نقوم بتقدير دالة m الموضحة في الصيغة (1). وعلى افتراض أن يكون لدينا (p+1) مشتقة مستمرة لبعض قيم $p \geq 0$ عند النقطة x_o. وباستخدام توسيع تايلر، يمكننا تقريب الدالة $m(x)$ موضعياً باستخدام متعدد الحدود من الدرجة p، وكما موضح بالصيغة التالية:

$$m(x) \approx m(x_o) + (x - x_o)m^{(1)}(x_o) + \ldots + (x - x_o)^p m^{(p)}(x_o)/p!$$

إذ أن $m^{(r)}(x_o)$ تمثل المشتقة r للدالة $m(x)$ عند النقطة x_o.

لنفترض هنا أن تكون $\beta_r = m^{(r)}(x_o)/r!$ حيث أن r = 0, 1, ..., p ، وأن $\hat{\beta}_r , r = 0, 1, \ldots, p$ تمثل القيم التي تقلل معيار المربعات الصغرى الموزونة (WLS) التالي:

$$\sum_{i=1}^{n} \{y_i - \beta_o - \beta_1(x - x_i) - \ldots - \beta_p(x - x_i)^p\}^2 K_h(x - x_i) \qquad \ldots\ldots\ldots (2)$$

حيث أن $K_h(.) = K(./h) / h$ ، والتي يتم الحصول عليها بإعادة احتساب دالة النواة $K(.)$ مع الثابت $h > 0$ ، والذي يدعى بدوره بعرض الحزمة أو معلمة التمهيد. أن عرض الحزمة يستخدم بشكل أساسي في تحديد حجم الجوار الموضعي local neighborhood، والموضح بالصيغة الأتية:

$$I_h(x_o) = [x_o - h + x_o + h] \qquad \ldots\ldots\ldots (3)$$

تقوم دالة النواة $K(.)$ بتحديد كيف أن المشاهدات ضمن الفترة $I_h(x_o)$ قد تساهم في الملاءمة عند النقطة x_o. ويمكن الاشارة إلى مقدار المشتقة r ، $m^{(r)}(x_o)$ ، على أنه $\hat{m}_h^{(r)}(x_o)$ وكما موضح بالصيغة الأتية:

$$\hat{m}_h^{(r)}(x_o) = r!\, \hat{\beta}_r \quad , \quad r = 0, 1, \ldots, p$$

وبشكل خاص، تم الحصول على مقدار LPK من الدرجة p للدالة $m(x_o)$ وهو $\hat{m}_h(x_o) = \beta_o$. كما ويمكن تمثيل $\hat{m}_h(x_o)$ بشكل مصفوفة كما موضح في أدناه:

$$X = \begin{bmatrix} 1 & (x_1 - x_o) & \ldots & (x_1 - x_o)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x_o) & \ldots & (x_n - x_o)^p \end{bmatrix}$$

وأن

$$W = diag\,(K_h(x_1 - x_o), \dots, K_h(x_n - x_o))$$

i.e.

$$W = \begin{bmatrix} K_h(x_1 - x_o) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K_h(x_n - x_o) \end{bmatrix}$$

تمثل مصفوفة التصميم ومصفوفة الوزن لملاءمة مقدر النواة الموضعي المحلي حول x_o. وبالتالي يمكن إعادة كتابة معيار WLS كما موضح في أدناه:

$$(y - X\beta)^T W (y - X\beta) \qquad \dots\dots\dots\dots\dots\dots\dots\dots \textbf{(4)}$$

حيث أن $y = (y_1, \dots, y_n)^T$ **، وأن** $\beta = (\beta_1, \beta_2, \dots, \beta_n)^T$ **. وهذا بدوره يؤدي إلى:**

$$\hat{m}_h^{(r)}(x_o) = r!\, e_{r+1}^T\, S_n^{-1}\, T_n Y \quad , \quad r = 0, 1, \dots, p$$

إذ أن e_{r+1} تشير إلى متجه وحدة من الدرجة (p+1) والتي تكون عناصره في المجال (r+1) مساوية للواحد وأما باقي عناصره الأخرى فتكون أصفاراً، وأن:

$$S_n = X^T W X \quad , \quad T_n = X^T W$$

وهنا سيتم التركيز هنا على ممهد المنحنى، وكما موضح في أدناه:

$$\hat{m}_h(x_o) = e_1^T\, S_n^{-1}\, T_n\, y \qquad \dots\dots\dots\dots\dots\dots\dots\dots \textbf{(5)}$$

هذا في حالة ما لم يتم مناقشة تقدير المشتقة. فإذا افترضنا أن $\hat{y}_i = \hat{m}_h(x_i)$ **تمثل القيمة الملائمة fitted value للدالة** $m(x_i)$ **. وباستخدام الصيغة (5) يتضح أن:**

$$\hat{m}_h(x_i) = a(x_i)^T y$$

حيث أن $a(x_i)$ **تمثل** $e_1^T S^{-1} T_n$ **بعد أن يتم استبدال x_o محل x_i . وبافتراض أن تمثل** $\hat{y}_h = [\hat{y}_1, \dots, \hat{y}_n]$ **القيم الملائمة عند جميع نقاط وقت التصميم design time points. وبالتالي يمكننا توضيح قيمة** $\hat{y}_h$ **كما في الشكل الآتي:**

$$\hat{y}_h = A_h y \qquad \dots\dots\dots\dots\dots\dots\dots \textbf{(6)}$$

حيث أن:

$$A_h = (a(x_1), \dots, a(x_n))^T \qquad \dots\dots\dots\dots \textbf{(7)}$$

وهي ما تعرف بمصفوفة التمهيد لممهد النواة المتعدد الحدود الموضعي LPK. وحيث أن A_h لا تعتمد على متجه الاستجابة y، فأن ممهد النواة المتعدد الحدود الموضعي $\hat{m}_h$ **يعرف على أنه ممهداً خطياً.**

2. الممهدات الثابتة الموضعية والممهدات الخطية الموضعية [4] [5]

Local Constant and Linear Smoothers

تعد الممهدات الثابتة الموضعية والخطية الموضعية من أبسط أنواع ممهدات LPK وأكثرها شيوعاً واستخداماً. إذ يُعرف الممهد الثابت الموضعي بإسم ممهد Nadaraya-Watson (N-W) والذي تم اقتراحه من قبل كل من (Nadaraya-Watson) في العام (1964). إذ تم استحصال هذا الممهد من ممهد النواة الموضعي LPK وذلك بجعل p=0:

$$\hat{m}_h(x_o) = \frac{\sum_{i=1}^{n} K_h(x_i - x_o)\, y_i}{\sum_{i=1}^{n} K_h(x_i - x_o)} \qquad \dots \textbf{(8)}$$

ومع وجود جوار موضعي **local neighborhood** بالشكل $I_h(x_o)=[x_o-h\,,\,x_o+h]$ والتي تناسب البيانات مع وجود ثابت. وهذا هو مقلل $\hat{\beta}_o$ لمعيار **WLS** التالي:

$$\sum_{i=1}^{n}(y_i-\beta_o)^2\,K_h(x_i-x_o)$$

لذا فإن مقدّر **N-W** يعد مقدّر سهل الفهم وبسيط الحساب في الوقت نفسه. ولنفترض أن $I_A(t)$ يشير إلى مؤشر الدالة لمجموعة معينة **A**. فعندما تكون دالة النواة **K** من النوع المنتظم **Uniform Kernel** الموضحة في أدناه:

$$K(t)=I_{[-1,1]}(x)/2=\begin{cases}\dfrac{1}{2} & , \quad x\in[-1,1]\\[2mm] 0 & , \quad \text{otherwise}\end{cases} \qquad \text{........ (9)}$$

فإن مقدّر **N-W** الموضح بالصيغة (8) يمثل المتوسط الموضعي لقيم y_i التي تكون داخل الجوار الموضعي $I_h(x_o)$ الموضح في الصيغة (3):

$$\hat{m}_h(x_o)=\frac{\sum_{i=1}^{n}I_{[x_o-h+x_o+h]}(x_i)y_i}{\sum_{i=1}^{n}I_{[x_o-h+x_o+h]}(x_i)}=\left\{\sum_{x_i\in I_h(x_o)}y_i\right\}\Big/ m_h(x_o)$$

حيث أن تمثل $m_h(x_o)$ عدد المشاهدات الواقعة ضمن الجوار الموضعي $I_h(x_o)$.

أما بالنسبة للممهد الخطي الموضعي الذي أقترح من قبل كل من **Stone (1984)** و**Fan (1993)** فإنه يتم استحصاله عن طريق ملاءمة مجموعة بيانات موضعياً مع دالة خطية. هنا نفترض أن يكون لدينا $(\hat{\beta}_o,\hat{\beta}_1)$ التي تقلل معيار **WLS** التالي:

$$\sum_{i=1}^{n}\left[y_i-\beta_o-(x_i-x_o)\beta_1\right]^2\,K_h(x_i-x_o)$$

وبالتالي فإن الممهد الخطي الموضعي هو $\hat{m}_h(x_o)=\hat{\beta}_o$. إذ يمكن الحصول عليه بسهولة وذلك باستخدام ممهد **LPK** والموضح في الصيغة (5) وذلك بجعل **p=1**. لذا يمكن التعبير بشكل مبسط عن الممهد الخطي الموضعي كما موضح بالصيغة في أدناه:

$$\hat{m}_h(x_o)=\frac{\sum_{i=1}^{n}\left[s_2(x_o)-s_1(x_o)(x_i-x_o)\right]K_h(x_i-x_o)y_i}{s_2(x_o)\,s_o(x_o)-s_1^2(x_o)} \qquad \text{..... (10)}$$

حيث أن:

$$s_r(x_o)=\sum_{i=1}^{n}K_h(x_i-x_o)(x_i-x_o)^r \quad , \qquad r=0,1,2$$

عادة فإن اختيار درجة ممهد **LPK** المناسبة، **p**، لا تكون مهمة بقدر أهمية اختيار عرض الحزمة، **h**. إذ أن كل من ممهد الثابت الموضعي (عندما **p=0**) والممهد الخطي الموضعي (عندما **p=1**) غالباً ما تكون جيدة بما فيه الكفاية لمعظم مشاكل التطبيق إذا تم تحديد دالة النواة **K** وعرض الحزمة **h** بشكل كاف.

لذا فإن دالة النواة **K(.)** المستخدمة في ممهد **LPK** المتمثل بالصيغة (5) هي عادة ما تكون دالة كثافة احتمالية متماثلة. وبما أن عرض الحزمة **h** يحدد حجم الجوار الموضعي $I_h(t_o)$ ، فإن دالة النواة **K(.)** تحدد الكيفية التي تساهم بها المشاهدات لملائمة ممهد **LPK** عند النقطة x_o.

ان عملية اختيار النواة عادة ما تكون غير حاسمة بشكل نهائي كونها لا تقوم بتحديد معدل التقارب لممهد **LPK** من المنحنى الأساسي. ومع ذلك، فإنها تقوم بتحديد الكفاءة النسبية لممهد **LPK**.

3. تحديد عرض الحزمة Bandwidth Selection [1] [4] [5]

يُعد الممهد جيداً اذا نتج عنه خطأ تنبوئياً صغيراً، وهذا عادة ما يُقاس عن طريق متوسط الخطأ المطلق Mean Absolute Error (MAE) أو متوسط مربع الخطأ Mean Square Error (MSE) الخاص بالممهد. فبالنسبة لممهد النواة المتعدد الحدود الموضعي LPK $\hat{m}_h(x_o)$ ، فإنه يمكن تعريف كل من MAE و MSE كما موضح بالشكل التالي:

$$\begin{aligned} MSE\left(\hat{m}_h(x_o)\right) &= E\left(\hat{m}_h(x_o) - m(x_o)\right)^2 \\ &= Bias^2\left(\hat{m}_h(x_o)\right) + Var\left(\hat{m}_h(x_o)\right) \end{aligned} \quad \text{........ (11)}$$

$$MAE(\hat{m}_h) = \frac{1}{n}\sum_{i=1}^{n}\left|m_{(i)h} - \hat{m}_{(i)h}\right|$$

حيث أن:

$$Bias\left(\hat{m}_h(x_o)\right) = E\,\hat{m}_h(x_o) - m(x_o)\,,$$

$$Var\left(\hat{m}_h(x_o)\right) = E\left(\hat{m}_h(x_o) - E\hat{m}_h(x_o)\right)^2$$

والتي بدورها تمثل كل من مقدار التحيّز والتباين لـ $\hat{m}_h(x_o)$ ، وأن w(x) تمثل دالة الوزن weight function، والتي غالباً ما تستخدم في تحديد مدى معين من مجال اهتمام الباحث. إذ يتحكم عرض الحزمة h في مربع التحيز والتباين لممهد النواة الموضعي LPK، إذ يتناسب عرض الحزمة h طردياً مع مربع التحيز ولكنه يتناسب عكسياً مع التباين، أي إذا كانت h صغيرة فإن التحيز يكون صغيراً أيضاً ولكن التباين يكون كبيراً والعكس صحيح. بالتالي فإن الاختيار الصحيح لعرض الحزمة h يعتمد على المفاضلة بين قيمتي التحيز والتباين وذلك لغرض الحصول على قيم صغيرة لكل من معياري (MAE) و (MSE). كما ويقوم عرض الحزمة h بتحديد حجم الجوار الموضعي $I_h(x_o) = [x_o - h + x_o + h]$ ، فعندما تكون h صغيرة فإن الفترة $I_h(x_o)$ ستحتوي على مشاهدات قليلة، وعليه فإن قيمة مقدار LPK ($\hat{m}_h(x_o)$) المستحصل عليه من استخدام أسلوب المربعات الموزون WLS ستكون مقاربة إلى حد ما من قيمة دالة الانحدار الحقيقية $m_h(x_o)$ ، وبالتالي نحصل على تحيز صغير لمقدر $\hat{m}_h(x_o)$ ولكن مع وجود تباين كبير. أما في حالة كون h كبيرة فأن الفترة $I_h(x_o)$ ستحتوي على مشاهدات كثيرة وعليه فأن $\hat{m}_h(x_o)$ وبذلك يكون لها تحيز كبير وتباين صغير. لذا يتم اختيار عرض الحزمة h التي تؤدي بنا إلى الحصول على تصغير لقيمتي معياري (MAE) و(MSE)، إذ سيتم الاعتماد على معيار المفاضلة وهو معيار العبور الشرعي (CV) وذلك لكون العملية الحسابية لهذا المعيار تتطلب تحديد قيمة دالة الانحدار $m_h(x_o)$ والتي تكون بدورها غير معلومة.

تم اعتماد عدة صيغ لدوال النواة التي من الممكن اعتمادها في عملية تمهيد دالة الانحدار الامعلمي. وإن أفضل دالة من دوال النواة المقترحة هي التي تقلل من قيمة معياري متوسط الخطأ المطلق Mean Absolute Error (MAE) ومتوسط مربعات الخطأ Mean Squared Error (MSE). الجدول التالي يوضح بعض أهم دوال النواة والمستخدمة في تقدير نماذج الانحدار اللامعلمي مع ثلاث دوال مقترحة:

جدول (1)

يوضح دوال النواة المعتمدة والمقترحة

Function Name	Mathematical Formula	Interval
Gaussian	$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$	$I\{\lvert x\rvert \le \infty\}$
Epanechnikov	$K(x) = \frac{3}{4}(1 - x^2)$	$I\{\lvert x\rvert \le 1\}$
Quartic	$K(x) = \frac{15}{16}(1 - x^2)^2$	$I\{\lvert x\rvert \le 1\}$
Triweight	$K(x) = \frac{35}{32}(1 - x^2)^3$	$I\{\lvert x\rvert \le 1\}$

SKernelG1	$K(x)=\frac{15}{28}(1-\frac{x^4}{3})$	$I\{\lvert x\rvert\le 1\}$
SKernelG2	$K(x)=\frac{405}{712}(1-\frac{x^4}{3})^2$	$I\{\lvert x\rvert\le 1\}$
SKernelG3	$K(x)=\frac{3}{4}(1-x^2)$	$I\{\lvert x\rvert\le 1\}$

الجانب التجريبي

وصف تجربة المحاكاة

تم تنفيذ تجارب المحاكاة بأعتماد ثلاث حجوم للعينات هي (n_3=150, n_2=100, n_1=50) وذلك لتوليد البيانات الخاصة بالمتغيرات العشوائية لغرض محاكاة البيانات الحقيقية بالنسبة لأسلوب تمهيد النواة الموضعي الحصين وذلك بكتابة برنامج في لغة MATLAB، إذ تم تكرار كل تجربة على حدة 1000 مرة لغرض الحصول على نتائج متسقة. تم اعتماد ثلاث دوال رياضية مختلفة لتمثيل المتغير التوضيحي مع الأخذ بنظر الاعتبار الصيغة الخطية واللاخطية في تلك الدوال لتتلاءم وطبيعة البيانات المتعلقة بالمجال الاقتصادي، والدوال التي تم استخدامها هي:

دالة خطية [3] $f(x)=2x-1$

دالة لاخطية [2,3] $f(x)=x+2\exp(-16x^2)$

دالة من الدرجة الثانية [2] $m(x)=8(x-0.5)^2$

كما تم تلويث البيانات لكل حجم من حجوم العينات أعلاه بنسب تلويث (10% , 20% , 30%). وكانت نتائج تجربة المحاكاة كما يأتي :

جدول (2)

يبين المعدل لقيم المعيارين (MAE) و (MSE) * ولجميع حجوم العينات ومستويات التلويث لمقارنة دوال النواة المستخدمة في طريقة النواة الحصينة (RLPK) بالنسبة لأنموذج المقارنة الأول

Sample Size حجوم العينات	Kernel Fu. دوال النواة	مستويات التلويث الثلاث		
		10%	20%	30%
n_1=50	Epanechnikov	0.550606321592160 (0.495908033637790)	0.514424301311390 (0.451066156828222)	0.476091562556496 (0.405022215724994)
	Triweight	0.550605296901387 (0.495905708759783)	0.514422928582408 (0.451063302677452)	0.476090227303251 (0.405020287211766)
	SKernelG1	0.550605296848528 (0.495905708866464)	0.514422928779634 (0.451063303142091)	0.476090227530885 (0.405020287343311)
	SKernelG2	0.550605296809004 (0.495905708932258)	0.514422928910512 (0.451063303450915)	0.476090227683186 (0.405020287427701)
	SKernelG3	0.550605296710188 (0.495905709096743)	0.514422929237701 (0.451063304222974)	0.476090228063935 (0.405020287638677)
	SKernelG4	0.550605980226647 (0.495907258309311)	0.514423842974052 (0.451065204831362)	0.476091117370399 (0.405021572510934)
	SKernelG5	0.550605638421178 (0.495906483407060)	0.514423385015681 (0.451064253704875)	0.476090672495203 (0.405020929744979)
n_2=100	Epanechnikov	0.553927246686775 (0.502854757424173)	0.517165922793710 (0.457288487181174)	0.474983364410770 (0.403104607955904)
	Triweight	0.553926529847818 (0.502853522055012)	0.517165333129738 (0.457287546153615)	0.474982842204403 (0.403103902873135)
	SKernelG1	0.553926529792949 (0.502853521967102)	0.517165333044438 (0.457287545979069)	0.474982842128314 (0.403103902642130)
	SKernelG2	0.553926529753343 (0.502853521902545)	0.517165332983717 (0.457287545856689)	0.474982842074445 (0.403103902481666)
	SKernelG3	0.553926529654330 (0.502853521741159)	0.517165332831918 (0.457287545550743)	0.474982841939775 (0.403103902080507)
	SKernelG4	0.553927007792511 (0.502854345470264)	0.517165726442919 (0.457288173492657)	0.474983190238358 (0.403104372960334)
	SKernelG5	0.553926768908955 (0.502853933629807)	0.517165529859340 (0.457287859717663)	0.474983016031761 (0.403104137802051)
n_3=150	Epanechnikov	0.556793936451551 (0.508033148062723)	0.517509418443547 (0.459878384451508)	0.477831112032046 (0.408009825643928)
	Triweight	0.556793575304523 (0.508032757818414)	0.517509059792099 (0.459877833265172)	0.477830729567047 (0.408009497420028)

* القيم الموضحة في الجدول والمحصورة بين قوسين تمثل قيمة المعيار MSE وباقي القيم هي للمعيار MAE.

	SKernelG1	0.556793575152542 (0.508032757400334)	0.517509059659196 (0.459877833057498)	0.477830729459045 (0.408009497083550)
	SKernelG2	0.556793575046734 (0.508032757111973)	0.517509059567280 (0.459877832913728)	0.477830729383135 (0.408009496851294)
	SKernelG3	0.556793574782214 (0.508032756391078)	0.517509059337490 (0.459877832554309)	0.477830729193356 (0.408009496270658)
	SKernelG4	0.556793815927109 (0.508033018161349)	0.517509298753405 (0.459878200746310)	0.477830984630372 (0.408009716383576)
	SKernelG5	0.556793695521890 (0.508032887844676)	0.517509179162448 (0.459878016900823)	0.477830856923972 (0.408009606784607)

جدول (3)

يبين المعدل لقيم المعيارين (MAE) و (MSE) ولجميع حجوم العينات ومستويات التلويث لمقارنة دوال النواة المستخدمة في طريقة النواة الحصينة (RLPK) بالنسبة لأنموذج المقارنة الثاني

Sample Size حجوم العينات	Kernel Fu. دوال النواة	مستويات التلويث الثلاث		
		10%	20%	30%
n_1=50	Epanechnikov	0.551502479932774 (0.499352271877675)	0.512416659462867 (0.446698090647941)	0.473276147579454 (0.397882895701467)
	Triweight	0.551501022624010 (0.499349331713903)	0.512414753138089 (0.446695593360902)	0.473275079675835 (0.397880748276475)
	SKernelG1	0.551501022760987 (0.499349332138800)	0.512414753406436 (0.446695593609491)	0.473275079793959 (0.397880748724547)
	SKernelG2	0.551501022850743 (0.499349332420991)	0.512414753582971 (0.446695593770776)	0.473275079872537 (0.397880749025056)
	SKernelG3	0.551501023075127 (0.499349333126466)	0.512414754024302 (0.446695594173990)	0.473275080068977 (0.397880749776324)
	SKernelG4	0.551501993053879 (0.499351291164631)	0.512416023447624 (0.446697257735061)	0.473275791550836 (0.397882179257255)
	SKernelG5	0.551501507703879 (0.499350311345555)	0.512415387485790 (0.446696425437004)	0.473275435806884 (0.397881463698522)
n_2=100	Epanechnikov	0.553977323859071 (0.502949107995499)	0.517223960740023 (0.457419147661449)	0.474979503865473 (0.403186011770163)
	Triweight	0.553976587158453 (0.502947831740708)	0.517223355045575 (0.457418197964707)	0.474979021445756 (0.403185281149681)
	SKernelG1	0.553976587150698 (0.502947831719984)	0.517223354963371 (0.457418197806766)	0.474979021354372 (0.403185280970320)
	SKernelG2	0.553976587143498 (0.502947831701478)	0.517223354904601 (0.457418197695679)	0.474979021290227 (0.403185280845185)
	SKernelG3	0.553976587125497 (0.502947831655218)	0.517223354757675 (0.457418197417964)	0.474979021129864 (0.403185280532355)
	SKernelG4	0.553977078068155 (0.502948682352129)	0.517223758720659 (0.457418831062679)	0.474979342943126 (0.403185768213849)
	SKernelG5	0.553976832312633 (0.502948256919007)	0.517223556746300 (0.457418514407786)	0.474979182149505 (0.403185524572059)
n_3=150	Epanechnikov	0.556906954643633 (0.508208979918968)	0.517504797059077 (0.458329926245825)	0.478202951218034 (0.410574510134826)
	Triweight	0.556906566303909 (0.508208542650840)	0.517504385579340 (0.458329540004130)	0.478202582687192 (0.410573946085847)
	SKernelG1	0.556906566190528 (0.508208542288133)	0.517504385504843 (0.458329539674523)	0.478202582615163 (0.410573945969155)
	SKernelG2	0.556906566111273 (0.508208542037772)	0.517504385452119 (0.458329539446968)	0.478202582564884 (0.410573945887555)
	SKernelG3	0.556906565913134 (0.508208541411877)	0.517504385320308 (0.458329538878086)	0.478202582439185 (0.410573945683556)
	SKernelG4	0.556906825191009 (0.508208834281367)	0.517504659967735 (0.458329797622832)	0.478202828538742 (0.410574322079610)
	SKernelG5	0.556906695809801 (0.508208688322471)	0.517504522685284 (0.458329668689781)	0.478202705641694 (0.410574133997291)

جدول (4)

يبين المعدل لقيم المعيارين (MAE) و (MSE) ولجميع حجوم العينات ومستويات التلويث لمقارنة دوال النواة المستخدمة في طريقة النواة الحصينة (RLPK) بالنسبة لأنموذج المقارنة الثالث

Sample Size حجوم العينات	Kernel Fu. دوال النواة	مستويات التلويث الثلاث		
		10%	20%	30%
n_1=50	Epanechnikov	0.553737552996597 (0.500966476271125)	0.521532025187929 (0.460305163870838)	0.477698834100755 (0.403110114879408)
	Triweight	0.553735318190301 (0.500962195073756)	0.521529259391818 (0.460300408959781)	0.477696415709360 (0.403106765222452)
	SKernelG1	0.553735317869667 (0.500962195115246)	0.521529259152775 (0.460300409282482)	0.477696415650856 (0.403106765186616)
	SKernelG2	0.553735317642689 (0.500962195129218)	0.521529258975893 (0.460300409484977)	0.477696415602838 (0.403106765150910)
	SKernelG3	0.553735317075235 (0.500962195164151)	0.521529258533685 (0.460300409991217)	0.477696415482789 (0.403106765061646)
	SKernelG4	0.553736808591550 (0.500965049341630)	0.521531104299657 (0.460303578899614)	0.477698028674341 (0.403108998560352)
	SKernelG5	0.553736062860998 (0.500963622291300)	0.521530182146826 (0.460301994096480)	0.477697222420978 (0.403107881981425)
n_2=100	Epanechnikov	0.558734001910604 (0.511196289404968)	0.519533097425109 (0.460239148141440)	0.482611647483617 (0.414576659999132)
	Triweight	0.558732576064724 (0.511193670468877)	0.519531415100579 (0.460236645572016)	0.482609817002357 (0.414574258876556)
	SKernelG1	0.558732575491715 (0.511193669799727)	0.519531414913502 (0.460236645265480)	0.482609816662900 (0.414574258339206)
	SKernelG2	0.558732575091956 (0.511193669329033)	0.519531414779954 (0.460236645046808)	0.482609816422870 (0.414574257961754)
	SKernelG3	0.558732574092563 (0.511193668152310)	0.519531414446086 (0.460236644500134)	0.482609815822800 (0.414574257018135)
	SKernelG4	0.558733527677929 (0.511195417284018)	0.519532536462698 (0.460238314393160)	0.482611037675928 (0.414575860368294)
	SKernelG5	0.558733052099899 (0.511194543922038)	0.519531975842775 (0.460237480021191)	0.482610427328972 (0.414575059686449)
n_3=150	Epanechnikov	0.560891834405105 (0.514517535382731)	0.522094176646828 (0.464603986512316)	0.483372252803853 (0.416822034489933)
	Triweight	0.560890665759419 (0.514515795673643)	0.522094179521961 (0.464602234612618)	0.483370631850987 (0.416819878541148)
	SKernelG1	0.560890665229843 (0.514515794638693)	0.522094178965989 (0.464602233685958)	0.483370631445616 (0.416819878060080)
	SKernelG2	0.560890664862053 (0.514515793922176)	0.522094178579469 (0.464602233043773)	0.483370631161519 (0.416819877721513)
	SKernelG3	0.560890663942581 (0.514515792130900)	0.522094177613171 (0.464602231438321)	0.483370630451280 (0.416819876875104)
	SKernelG4	0.560891445009014 (0.514516956650491)	0.522095044028478 (0.464603403652450)	0.483371713291395 (0.416821316535714)
	SKernelG5	0.560891055413451 (0.514516376160805)	0.522094611888135 (0.464602819159472)	0.483371172723268 (0.416820597610723)

من نتائج تجربة المحاكاة والمبينة في الجداول أعلاه يمكننا استنتاج الآتي:

1. **ومن خلال مقارنة دوال النواة المعتمدة والمقترحة منها والتي تم استخدامها في طريقة النواة لمتعدد الحدود الموضعي الحصين (RLPK) تبيّن بأن دالة النواة المقترحة (SKernelG3) أنها أفضل دالة لمعظم حجوم العينات ولمستويات التلويث المختلفة ولجميع نماذج المقارنة الستة التي تم إعتمادها في تجارب المحاكاة، وذلك بالاعتماد على معياري متوسط الخطأ المطلق (MAE) ومتوسط مربعات الخطأ (MSE) لأنموذج الانحدار اللامعلمي وفق طريقة النواة لمتعدد الحدود الموضعي الحصين (RLPK) بالاعتماد على الدالة المذكورة.**
2. **بالاعتماد على قيم معياري (MAE) و(MSE) نجد أن أفضل دالة نواة مقترحة هي الثالثة (SKernelG3) كونها حصلت على أقل قيم بالنسبة للمعيارين السابقين وذلك لأغلب نتائج تجربة المحاكاة**
3. **لوحظ من نتائج تجربة المحاكاة لتحديد أفضل دالة نواة أن قيم معياري متوسط الخطأ المطلق (MAE) ومتوسط مربعات الخطأ (MSE) لكافة تكرارات التجربة لكل حالة من حجوم العينات الثلاث ومستويات التلويث الثلاث ونماذج المقارنة الثلاث المعتمدة في التجربة، كانت صغيرة ومتقاربة مما يشير إلى تجانس قيم المعيارين لكل حالة من حالات تجربة المحاكاة.**
4. **وجد الباحثين أنه كلما زادت نسبة تلويث البيانات أدى ذلك إلى تقليل قيم معدل متوسط الخطأ المطلق ومتوسط مربعات الخطأ مما يدل على تحسين في نتائج تقدير إنموذج الانحدار اللامعلمي.**

المصادر

1. **Härdle, Wolfgang, (1994), "Applied Nonparametric Regression". Cambridge: Cambridge University Press.**
2. **Kagerer, K., (2013), "A Short Introduction to Splines in Least Squares Regression Analysis". University of Regensburg Working Papers in Business, Economics and Management Information Systems, JEL Classification: C14, C51.**
3. **Schimek, Michael G., (2000), "Smoothing and Regression: Approaches, Computation, and Application". Wiley Series in Probability and Statistics, John Wiley & Sons, Inc.**
4. **Wand, M. P., & Jones, M. C., (1995), "Kernel Smoothing". Chapman & Hall.**
5. **Wu, Hulin, & Zhang, Jin-Ting, (2006), "Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches". Wiley Series in Probability and Statistics, John Wiley & Sons, Inc.**

……………………………………………………………..
…………………………………………….
………………………