تقدير دوال الانحدار اللامعلمي باستعمال بعض مقدرات كيرنل $(kernel)$ اللامعلمية لدالة الوزن (Gaussian)

Using Some of Nonparametric Kernel Estimators to Estimate the Nonparametric Regression Functions with Gaussian Weight Function

أ.م. د. جاسم ناصر حسين[2] — الباحث. محمد عبدالرضا داخل[1]

Jassim N. Hussain — Mohammed A. Dakhil

jasim.nasir@uokerbala.edu.iq — muhammed.a@s.uokerbala.edu.iq

كلية الادارة والاقتصاد - جامعة كربلاء / قسم الاحصاء

بحث مستل من رسالة ماجستير (للباحث)

المستخلص:

حينما لا يمكن تطبيق الطرائق المعلمية بسبب بعض القيود المفروضة عليها ولفقدانها للمرونة في التقدير وتحليل البيانات لذا تم اللجوء الى الطرائق اللامعلمية التي اثبتت كفائتها في تحليل البيانات دون الحاجة الى افتراضات مسبقة على الباحث واصبحت البيانات وماتحمله من معلومات هي التي تحدد شكلها الدالي للمجتمع المدروس وليس ثمة معلمات تنوب عن المشاهدات. ان الهدف من تقدير دالة الانحدار اللامعلمي هو تقريب دالة الانحدار الى دالة الانحدار الحقيقية وجاء بحثنا هذا ليلقي الضوء على بعض مقدرات كيرنل اللامعلمية لدالة الوزن $(Gaussian)$ وهي كل من (مقدر الانحدار الثابت الموضعي"$N.W$" و مقدر الانحدار الخطي الموضعي"$L.L$" و مقدر بريستلي تشاو**P.Ch** وقد اعتمد الجانب التطبيقي على اجراء التحليل الاحصائي واستخلاص النتائج والرسوم التوضيحية للمقارنة بين المقدرات باستعمال البرنامج الاحصائي (R) وقد توصلنا في الجانب التجريبي من خلال المقارنة بين هذه المقدرات إلى عدة استنتاجات أهمها: إن مقدر بريستلي تشاو ("P.Ch") أظهر أفضلية واضحة على بقية المقدرات من خلال النتائج والأشكال ولكل حالة من حجوم العينات الأربع وثلاث مستويات للانحراف المعياري وكذلك للنماذج المعتمدة في المحاكاة المتضمنة نتائج مقدرات كيرنل اللامعلمية.

الكلمات المفتاحية: الانحدار اللامعلمي ، أنموذج انحدار كيرنل اللامعلمي، طريقة كيرنل لمتعدد الحدود الموضعي ، مقدر "بريستلي تشاو"

Abstract

Parametric methods no longer meet the researcher's need due to the restrictions imposed on them because they lost flexibility in parameter estimation and data analysis. Therefore, non-parametric methods were used that because of their efficient in analyzing data without request from the researcher to make Pre-assumptions. The data and the information have the main rule to determining the function form of the studied population, and there are no parameters that represent the observations. Consequently, the purpose of estimating the nonparametric regression function is to approximate the regression function to the true regression function. Our research aims to Study and apply some nonparametric Kernel estimators for the Gaussian weight function, which are both (the localized constant regression estimator, the local linear regression estimator, and the Priestley Chow estimator. The experimental side relied on experiments Simulation on consistent data that simulates the real data that was used in the application side in representing community data, representing random errors, conducting statistical analysis and extracting results and illustrations for comparison between estimators and showing the best among them, using three criteria of comparison, average mean square error, average absolute mean error, and mean integrated square error, five different functions were assumed to generate data in the experimental side, four sample sizes, and three standard deviation values.

The important results of the experimental side are the Priestly Chow estimator show outperform of the other estimators for each of the four sample sizes and three levels of standard deviation, as well as for the five models adopted in the simulation that included the results of Kernel's estimators Nonparametric.

Key Word: Nonparametric methods, Nonparametric Kernel models، Local Polynomial kernel Method، Priestley-Chao model.

١. المقدمة:

أن تحليل الانحدار هو طريقة احصائية لبناء نموذج رياضي سببي للتنبؤ بمتوسط متغير عشوائي اعتمادا على قيم متغير عشوائي واحد او أكثر، فهو يبحث في ايجاد العلاقة بين المتغيرات التوضيحية ومتغير الاستجابة ومن هذه النماذج (نماذج الانحدار المعلمي والانحدار اللامعلمي) ولكل منهما مميزاته وان استعمال احدى الأنموذجين لا يمنع من استعمال الأنموذج الأخر.

كما انهما يكملان بعضهما البعض من خلال صحة كل منهما للأخر فأن نماذج الانحدار المعلمي تفترض أن العينة تأتي من مجتمع معين له عائلة معروفة من التوزيعات الاحتمالية ثم العمل على تقدير المعالم المجهولة لتلك العائلة باستعمال الطرائق الكلاسيكية لكن تلك الافتراضات الخاصة بالنماذج المعلمية قد تكون غير متوفرة لأن التوزيع المعلمي المفترض لا يكون بالضرورة التوزيع الفعلي للمسألة المراد حلها وقد يؤدي بالطرائق الاحصائية المستخدمة الى استنتاجات غير صحيحه لذا نلجأ الى نماذج الانحدار اللامعلمي في حالة عدم توفر هذه المعلومات ومع اختلاف حجوم العينات ، وزيادة تعقيد البيانات، فإن ذلك يعيق الحصول على مقدر الدالة بالطرائق التحليلية التقليدية.

٢. هدف البحث:

يهدف الحث إلى دراسة بعض الطرائق اللامعلمية التي تستعمل لتقدير دالة أنموذج الانحدار اللامعلمي باستعمال مقدرات كيرنل $(Kernel\ Estimator)$ اللامعلمية لدالة الوزن $(Gaussian)$، والمقارنة بين مقدراته المختلفة من خلال دراسة نظرية تشتمل على تسليط الضوء على أفضل ما قدمته الدارسات السابقة والبحوث الحديثة المنشورة في مجال الانحدار اللامعلمي عموما مقدرات كيرنل خصوصا، وأن هذه الأساليب قد تم اعتمادها في تقدير أنموذج الانحدار اللامعلمي وتم تطبيقها في الجانب التجريبي

٣. مشكلة البحث:

مشاكل عدة قد تواجه الباحث في تقدير نماذج الانحدار اللامعلمي، ومن هذه المشاكل إيجاد طريقة كفوءة تلائم نماذج متنوعة من دوال الانحدار، لذا اقتضت الحاجة إلى استعمال مقدرات كيرنل اللامعلمية لدالة الوزن $(Gaussian)$ لتقدير دوال الانحدار اللامعلمي وتم استعمال ثلاثة مقدرات كل من [مقدر الانحدار الثابت الموضعي⟨Local Constant Regression Estimator⟩ اوما يسمى نادريا واتسون "N.W" ⟨Nadaraya-Watson estimator⟩ومقدر الانحدار الخطي الموضعي"L.L" ⟨Local linear smoother⟩ ومقدر بريستلي تشاو"P.Ch"] ⟨priestley − chao estimator⟩ والمفاضلة بينهما من خلال معايير المفاضلة للأخطاء كل من [متوسط مربعات الخطأ (AMSE) ومتوسط الخطأ المطلق(AMAE) وتكامل مربعات الخطأ](MISE) و التوصل الى افضل مقدر في ظل ما تعانيه النماذج اللامعلمية من مشاكل عديدة كأسلوب اختيار المعلمة التمهيدية(h) و حجم البيانات المستخدمة .

٤. الجانب النظري

٤-١ الانحدار اللامعلمي **Non parametric regression**

ويستخدم الانحدار اللامعلمي [1] لتقدير دالة الانحدار $\mu(x_i)$ بشكل مباشر وبدون وجود أي صيغة محددة لها وبعيداً عن تقدير معلمات الأنموذج كما هو الحال في الانحدار المعلمي (وان الدالة) $\mu(x_i)$ دالة مستمرة $(Continuous\ Function)$ وممهده (smoothing) فإذا جمعنا العينة العشوائية المكونة من (n) من المشاهدات $[(x_i, y_i)]_{i=1}^{n}$ ، يمكن أن تصف العلاقة بين متغير الاستجابة [Response Variable] والمتغيرات التوضيحية [Explanatory Variables] بأنموذج الانحدار اللامعلمي الاتي[2]:

-

$$y_i = \mu(x_i) + \varepsilon_i \qquad (i = 1,2, \quad , \dots , n) \qquad \dots (1)$$

إذ أن: -

- y_i: تمثل متغير الاستجابة في i .
- x_i: قيم المشاهدة للمتغير التوضيحي في i .
- $\mu(x_i)$: هي دالة انحدار (Y) على (X) $\mu_{(xi)} = E(Y/X = x)$ أو منحنى الانحدار المراد تقديره عند (x_i) وبطبيعة الحال لا تحتوي على معالم، [3] إذ يتم الافتراض بأن دالة الانحدار $\mu_{(xi)}$ هنا على انها دالة متزايدة $\{Increasing\}$ في (x_i) إذا كانت $\mu(x_i) \le \mu(x_i')$ عندما $x_i \le x_i'$.
- ε_i: يمثل الخطأ العشوائي $\varepsilon_i \sim N(O, \sigma^2)$ وهي ان قيمة الأخطاء تتوزع طبيعيا بتوقع مقداره (صفر) و تباين ثابت (σ^2).

ان الانموذج اللامعلمي يزودنا بعدد من النتائج الرئيسية وكالاتي:

i. وصف العلاقة العامة بين متغيرين.
ii. تسمح هذه النماذج بالتعويض عن القيم المفقودة لهذا يمكن ان تعد هذه النماذج مرنه.
iii. التنبؤ بالمشاهدات (Prediction of observations)

٤-٢ أنموذج انحدار كيرنل اللامعلمي **kernel Nonparametric regression model**

يعود التحليل الاحصائي لتقنيات الانحدار اللامعلمي لخمسينيات القرن الماضي حيث تم اقتراحها لأول مرة من قبل الباحثين[4] $(Rosen\ \ blatt)$ في عام (1956) و(Parzen) [5]في عام (1962) اشارا الى نوع عام من أساليب التقدير اللامعلمية للدوال اي تقدير دالة الكثافة الاحتمالية. ان الغرض من هذا التقدير هو تعديل البيانات بالشكل الذي يجعل الحصول على مقدرات ذات صفات تتقارب مع خواص المعلمات الحقيقية وان هذا التقريب لدالة الانحدار يمكن عمله بأسلوبين:

- الاسلوب الاول: الطريقة المعلمية المتمثلة بأحد هذه الطرق: (طريقة المربعات الصغرى OLS ، وطريقة الامكان الاعظم MLE ، و, ...الخ)

- الاسلوب الثاني: ويتمثل بتقدير دالة الانحدار عندما لا تتوزع البيانات توزيعا معينا لأي سبب معين كقلة البيانات وهذا ما يسمى بالتقدير اللامعلمي.

[6] وبذلك تكون هذه المقدرات هي مقدّرات لقيم الدالة إذ ان هذا المقدر يشير إلى نوع عام من أساليب التقدير اللامعلمي للدوال، إذ لو افترضنا وجود مجموعة بيانات أحادية المتغير ويراد عرضها بيانيا، لذا فان مقدر $(Kernel)$ سوف يزودنا بوسيلة فعالة لتحقيق هذا الهدف وإيجاد تركيب لمجموعة البيانات بدون افتراض الأنموذج المعلمي.

٤-٣ اختيار دالة كيرنل $\mathrm{K}(u)$ Kernel function Selection

ان اختيار دالة[7] (Kernel) يعَد ضروريا ومهما للحصول على مقدرات تقترب من الخواص للبيانات الإحصائية، تستخدم دوال تمهيد كيرنل في تقدير دوال الكثافة الاحتمالية ودوال الانحدار ودوال الطيف و أن تقدير دوال الكثافة الاحتمالية يمكن اعتبارها ابسط حالة لتمهيد البيانات $(Data\ Smoothing)$ إذ ان أداء تقديرات الكثافة الاحتمالية تعتمد جوهرياً على اختيار معلمة عرض الحزمة[8]، لذا فأنها تعد العامل الاكثر اهمية في تحديد مقدار التمهيد (التنعيم) لمقدرات الكثافة ، إذ يتم التحكم في مقدار التمهيد من قبل معلمة عرض الحزمة .

أي أن تقدير الكثافة الاحتمالية يمكن اعتباره إعداد أساسي لدراسة تمهيد البيانات، إذ المقصود من تقدير الكثافة الاحتمالية (تمهيد كيرنل) اي امكانية تطبيقها بشكل مباشر على بعض المشاكل اللامعلمية، على سبيل المثال (الانحدار، التصنيف، التحليل التمييزي، الاقتصاد القياسي، التمويل، اختبار حسن المطابقة) كما تم ذكره في الدراسات من قبل الباحثين[9][7][10]: فهناك نوعان من دوال $(Kernel)$ يمكن تمييزهما وهما كل من:

- دوال كيرنل الأقل تباين والتي تعمل على تقليل التباين المحاذي.
- ودوال كيرنل المثالية والتي تعمل على تقليل متوسط مربعات الخطأ المتكامل او (الخطأ المحاذي) $[(MISE)\ Mean\ Integrated\ Squared\ Error]$ أي اشتقاق (MSE) وهي الدالة والتي تم اشتقاقها من قبل $(EpanchniKov)$ عام (١٩٦٩) [11] وكالاتي:

$$MISE = E\int[\hat{\mu}_{(x)} - \mu_{(x)}]^2 w_{(x)} d_x \qquad \dots (2)$$

تعرف دالة كيرنل بأنها دوال حقيقة ويرمز لها بالرمز $\mathrm{K}(u)$ ولها عدة تسميات منها دوال وزن ودوال نافذة ودوال شكل ودوال أساسية والتي تحقق الخصائص التالية [12] [13][14] :

١. قيمة دالة كيرنل أكبر من او تساوي صفر أي انها دالة موجبة وتمثل دالة كثافة احتمالية.

$$\mathrm{K}(u) \geq \mathbf{0}$$

٢. التكامل عند دوال كيرنل دائما يساوي واحد والمشتقة الثانية معلومة.

$$\int_{-\infty}^{\infty} \mathrm{K}(u)du = 1\ ) \qquad \dots (3)$$

٣. دوال متماثلة $(symmetric)$ وتحقق $\mathrm{K}(-u) = \mathrm{K}(u)$ لجميع قيم $(\boldsymbol{u})$ وفي هذه الحالة فان جميع العزوم الفردية حول المتوسط تساوي صفر.

٤. دوال كيرنل مستمرة $\quad -\mathbf{1} < \mathrm{K}(u) < \mathbf{1}$

٥. رتبة دالة كيرنل (order) يرمز لها بالحرف $(\boldsymbol{r})$ وتعرف بانها رتبة اول عزم غير صفري فمثلا دالة كيرنل$\mathrm{K}(u)$ تكون من الرتبة الثانية $(\boldsymbol{r} = \mathbf{2})$ إذا حققت الشرط التالي:

$$m_1\,\mathrm{K}(u) = 0 \quad , m_2\,\mathrm{K}(u) = 0$$

٤-٤ *طرائق التقدير:* ***Estimation methods***

أن الطرائق اللامعلمية في التقدير اكتسبت الكثير من الاهتمام في السنوات الأخيرة وذلك لأنها تتمتع بمرونة كافية في مطابقة الدوال بغض النظر عن كونها خطية أو لا خطية وهذه الطرائق لا تحتاج الى افتراضات شديدة وصارمة حول شكل الدالة المجهولة ، أن العديد من الباحثين استعملوا الطرائق اللامعلمية لتقدير دالة الانحدار اللامعلمي ومن هذه الطرائق طريقة المربعات الصغرى الموزونة$\langle Weighted\ Least\ Square\rangle$ أوطريقة الإمكان الأعظم $\langle Maximum\ Likelihood\rangle$ [3] وبالاستفادة من الأساليب المتقدمة في التحليل الرياضي وتطبيقا للمقولة الشهيرة "دع البيانات تتحدث عن نفسها" فوجدوا فيها بيئة خصبة للبحث والتقصي وتطبيقها في مجالات مختلفة من الظواهر الحقيقة ، ولكن من مساوئ هذه الطرائق امتلاكها تحيزاً كبير نسبياً وعند محاولة تقليل التحيز عن طريق تكبير حجم العينة حسب نظرية الغاية المركزية والحصول على خواص جيدة للمقدرات مثل الاتساق والمحاذاة سوف نواجه مشكلة أخرى وهي أن التباين سيكون كبير ، لذلك علينا مراعاة هاتين المسألتين (التحيز و التباين) وبالتالي محاولة تقليل مجموع مربعات البواقي.

وهنالك العديد من الطرائق [15] التي يمكن استعمالها في تقدير دالة الانحدار اللامعلمي ومن هذه الطرائق طريقة كيرنل لمتعدد الحدود الموضعي لتقدير دالة الانحدار اللامعلمي وتعتبر طريقة كيرنل لمتعدد الحدود الموضعي $(Local\ Polynomial\ kernel\ Method\ )$من الطرائق اللامعلمية المهمة جدا في تقدير دالة الانحدار.

٥-٤ طريقة كيرنل لمتعدد الحدود الموضعي (LPK) Local Polynomial kernel Method

يعد متعدد الحدود الموضعي أحد الأساليب الإحصائية المستعملة لتقدير أنموذج دالة الانحدار اللامعلمي[16] وهو من أفضل طرائق التمهيد حتى انه يفضل على بقية طرائق $(Kernel)$ [17]الشائعة الأخرى فضلا عن ذلك يتمتع هذا الممهد بقابلية التكيف بطبيعة الانموذج، إي انه يستعمل مع الانموذج الثابت والعشوائي (الانموذج الثابت يستعمل معلمة ممهدة ثابته والانموذج العشوائي يستعمل معلمة ممهدة متغيرة)، ولممهدات الانحدار الخطي الموضعي كفاءة عالية مقارنة مع الممهدات الأخرى المختلفة.

عليه فان الفكرة الأساسية لهذه الطريقة هي تقدير الدالة $\mu_{(xi)}$ بصورة موضعية بدلا من استعمال متعددة حدود كلية $(Global\ Polynomial)$ [18]والتي تستعمل البيانات جميعها (n) لتقدير الدالة $\mu(x_i)$ من خلال تقدير $(p+1)$ من المعلمات كما في الانحدار المعلمي، ولتقدير الدالة$\mu(x_i)$ بصورة موضعية $(Locally)$ عند النقطة (x) يتم تحديد جوار $(Neighborhood)$ الذي يحتوي على النقطة (x) بالشكل $(x-h, x+h)$ حيث (h) عرض الحزمة $(Bandwidth)$ او معلمة التمهيد$(smoothing)$ و الذي يحدد عرض الجوار حول (x) وسوف تستخدم فقط المشاهدات (yi) التي تقطع نقاط بياناتها (xi) ضمن الفترة السابقة في تقدير الدالة $\mu(x_i)$ يلاحظ أن تقدير الدالة باستعمال متعددة الحدود الموضعية يعتمد بشكل أساسي على عناصر أساسية هي: (اختيار درجة متعددة الحدود، وعرض الحزمة (h) .

١-٥-٤ مقدر الانحدار الثابت الموضعي Estimator Local Constant Regression

مقدر الانحدار الثابت الموضعي $(Local\ Constant\ Regression\ Estimator)$ [19]وهو عبارة عن مقدر(Nadaraya − Watson estimator) يعد هذا المقدر أحد أقدم المقدرات اللامعلمية وأكثرها شيوعا واستعمالا وتعود تسميته الى اسماء الباحثين اللذين اقترحا هذا المقدر كل من الباحثين⟨Nadaraya and Watson⟩ [20][21] عام (١٩٦٤) علماً إن الطريقة تعود إلى تقديم سابق للباحث TuKey في عام (١٩٦١) لما يسمى $(Regressogram)$ الذي يعتمد على استعمال فكرة الرسم البياني نفسها لتقدير دالة الكثافة. وهو من المقدرات الشائعة الاستعمال في التطبيق لنماذج الانحدار اللامعلمي ويمتاز بأنه دالة محددة ومستمرة وذات قيم موجبة وغير سالبة وتكاملها مساوي للواحد وبالاعتماد على طريقة متسلسلة الاوزان ويمكن إثبات ذلك باستعمال طريقة المربعات الصغرى الموزونة (WLS) وبذلك يكون مقدر[18] Nadaraya − Watson estiμato (N.W) للدالة $\hat{\mu}$ كالآتي:

$$\hat{\mu}_{(N.W)} = \frac{\Sigma_{i=1}^{n} K_n\left(\frac{x-x_i}{h}\right) y_i}{\Sigma_{i=1}^{n} K_n\left(\frac{x-x_i}{h}\right)} \qquad \dots(4)$$

حيث $K(u)$ دالة $(Kernel)$ ، و(h) عرض الحزمة [16]الذي يتحكم بعرض الجوار في دالة $(Kernel)$ والذي بدوره يتحكم بكمية التمهيد للمقدر الناتج. ويلاحظ مما تقدم عند اختيار عرض حزمة صغير جدا ويقترب من الصفر فإن قيمة عرض الحزمة تكون قريبة من الصفر وعندها سوف نستعمل النقطة (x_i) فقط في ايجاد التقدير وأن:

$$K\left(\frac{x_i - x}{h}\right) = K\left(\frac{0}{h}\right) = K(\mathbf{0}) \qquad \dots(5)$$

أي أن المنحنى المقدر سوف يمر بجميع نقاط البيانات وعندها سيكون المنحنى المقدر منحنى متذبذباً ويتصف بكون تباينه عالٍ وتحيزه أقل ويسمى في هذه الحالة منحنى تحت التمهيد $\langle Under\ Smooth\rangle$.

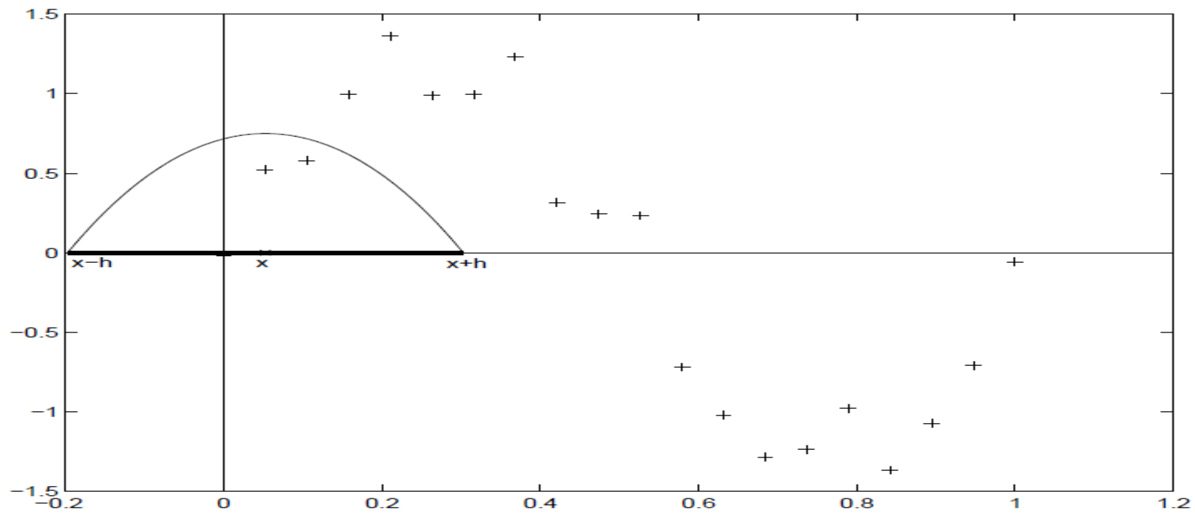
[22] أما عندما يتم اختيار عرض الحزمة كبيراً جداً او يقترب من $\langle\infty\rangle$ فإن مقدر $\boldsymbol{NW}$ سيصبح كالآتي:

$$\hat{\mu}_{NW}(xi) = lim_{h\to\infty} \sum_i \frac{K\left(\frac{x_i - x}{h}\right) y_i}{\sum_i K\left(\frac{x_i - x}{h}\right)} \qquad \dots(6)$$

$$\boldsymbol{\hat{\mu}_{NW}(xi)} = \sum_i \frac{K(\mathbf{0})\boldsymbol{y_i}}{\sum_i K(\mathbf{0})} = \frac{K(\mathbf{0})\sum_i \boldsymbol{y_i}}{\boldsymbol{n}K(\mathbf{0})}\boldsymbol{n}^{-1}\sum_i \boldsymbol{y_i} = \boldsymbol{\bar{y}} \qquad \dots(7)$$

وهذا يعني أن المنحنى الناتج سيكون قريباً من الخط المستقيم وفي هذه الحالة يسمى المنحنى المقدر بمنحنى فوق التمهيد $\langle Over\ smooth\rangle$ [23] والذي يتصف بكونه ذي تباين قليل وتحيز عالٍ. ولذلك يجب اختيار عرض الحزمة ليوازن بين هاتين الغايتين تحت التمهيد "Under smoothing" وفوق التمهيد "Over smoothing" .

ويمكن القول أنه مما يؤخذ على مقدر Nadaraya − watson estimato (N.W) أنه يعاني من التحيز العالي حيث أن التحيز فيه يعتمد على مواقع نقاط البيانات $\boldsymbol{x_i}$ ، كما إنه يعاني من تأثير التحيز عند الحد⟨Boundary Biase⟩ والذي ينتج من كون نصف الأوزان تكون غير معروفة وتقع خارج الحد [24] كما مبين في الشكل (1)



الشكل رقم (1) يمثل تأثير الحد كما يظهر في الجهة اليسرى [25]

٤-٥-٢ مقدر الانحدار الخطي الموضعي Local linear Regression Estimator

يعد ممهد الانحدار الخطي الموضعي الذي تم اقتراحه من قبل كل من Stone (1984) و Fan (1992) [26] من أفضل طرائق التمهيد حتى أنه يفضل على بقية طرائق $(Kernel)$ الشائعة الأخرى فضلاً عن ذلك يتمتع هكذا ممهد بقابلية التكيف الانموذج: حيث انه يستعمل مع الأنموذجين العشوائي والثابت. ولممهدات الانحدار الخطي الموضعي كفاءة مقاربة وعالية (أي يمكن ان تقترب 100%) مع خيار ملائم لدالة كيرنل وعرض الحزمة من بين كل الممهدات الخطية الممكنة أو المتاحة. اما في حالة التعويض عن $\boxed{p = 1}$ فإن المقدر الناتج هو مقدر الانحدار الخطي الموضعي

ولتوضيح عمل هذا المقدر وصيغته فإننا نفترض أن المشتقة الثانية لدالة الانحدار $(\mu'')_{(xi)}$ تكون موجودة. في المنطقة المجاورة الصغيرة للنقطة x ومتجه المعلمات المقدرة يعبر عنه بالآتي $\hat{\beta} = \begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{pmatrix}$ والناتج من تصغير المربعات الصغرى الموزونة أي عندما $(p = 1)$ فإن المقدر الناتج هو مقدر الانحدار الخطي الموضعي (Local linear smoother " L.L")

$$\hat{\mu}_{(X,l,h)} = n^{-}\frac{\Sigma_{i=1}^{n} K_n(x - x_i)y_i[\hat{s}_2(x,h) - \hat{s}_1(x,h)(x - x_i)]}{\hat{s}_0(x,h)\hat{s}_2(x,h) - \hat{s}_1(x,h)^2} \qquad \dots(8)$$

[24]احياناً يتم اختيار (p) درجة ممهد متعدد الحدود الموضعي $(L.P.K)$ المناسبة، لكن لا تكون مهمة بقدر أهمية اختيار عرض الحزمة (h). إذ أن كل من ممهد لانحدار الثابت الموضعي $(Local\ Constant\ Regression\ Estimator)$ (عندما $p = 0$) وممهد الانحدار الخطي الموضعي (Local linear smoother " L.L") (عندما $p = 1$)غالباً ما تكون جيدة بما فيه الكفاية لمعظم مشاكل التطبيق إذا تم تحديد دالة كيرنل $K(u)$ وعرض الحزمة (h) بشكل كاف.

٤-٥-٣ مقدر "بريستلي تشاو" *priestley- chao estimator*

هو المقدر الذي تم اقتراحه من قبل الباحثين [27] $(priestley - chao)$ عام (١٩٧٢) حيث يعرف الصيغة الاتية:

$$\hat{\mu}_{P.Ch(xi)} = h^{-1}\sum_{i=2}^{n}(x_i - x_{i-1})\,K(u)\left(\frac{x - x_i}{h}\right)y_i \qquad \dots(9)$$

- حيث $\hat{\mu}_{P.Ch(xi)}$: تمثل دالة الانحدار المقدر بطريقة $(priestley - chao)$
- $K(u)$: تمثل دالة كيرنل التي يفترض انها متماثلة حول نقطة الصفر
- h : تمثل عرض الحزمة

ولوحظ ان مقدر بريستلي تشاو $"P.Ch"$ $(priestley - chao\ estimator)$ متكافئ تقريبا لمقدر الانحدار الثابت الموضعي $(Local\ Constant\ Regression)$ له نفس صيغة التحيز والتباين التقاربين[28].

$$E[\hat{\mu}(x_i)] = \frac{1}{nh}\sum_{i=1}^{n} K(u)(\frac{x-x_i}{h})\mu(x_i) \qquad \dots(10)$$

$$V[\hat{\mu}(x_i)] = \frac{\sigma^2}{n^2h^2}\sum_{i=1}^{n} K(u)\left(\frac{x-x_i}{h}\right)^2 \qquad \dots(11)$$

٤-٦ معايير المفاضلة

إن قيم معايير المفاضلة بكافة انواعها تمثل مدى اقتراب دالة التخمين $\hat{\mu}(x_i)$ من دالة الانحدار الاصلية $\mu(x_i)$حيث كلما اقتربت قيمة المعيار من الصفر فإن دالة التخمين تكون قد اقتربت من الدالة الاصلية وهناك عدد من معايير المفاضلة المستخدمة[29]

١. معدل متوسط مربعات الخطأ $\langle Average\ Mean\ Square\ Error\rangle$ وصيغته:

$$AMSE = \frac{\sum_{i=1}^{N} MSE(i)}{N} \qquad \dots(12)$$

حيث N تمثل عدد التكرارات

٢. معدل متوسط الخطأ المطلق ⟨Average Mean Absolute Error⟩ وصيغته:

$$AMAE = \frac{\sum_{i=1}^{N} MAE(i)}{N} \qquad \dots(13)$$

حيث N تمثل عدد التكرارات

٣. تكامل مربعات الخطأ

$$\text{MISE} = \frac{1}{100}\sum_{i=1}^{100}\left(\mu(x_i) - (\hat{\mu})_{(xi)}\right)^2 \qquad \dots(14)$$

عادة ما تصادف صعوبات كثيرة في حل هذا التكامل لذا يتم اللجوء الى الطريقة التقريبية من خلال تقسيم فترة التغير للمتغير (x) الى ⟨١٠٠⟩ فترة (هذا الرقم ممكن ان يزيد او يقل حسب قناعة الاحصائي بالنتائج المعيارية)

٥. المحاكاة (Simulation)

٥-١ المقدمة

من أجل تطبيق نماذج (كيرنل) اللامعليمة المعتمدة التي تم التطرق إليها في (الجانب النظري)، فقد اعتمدنا الأسلوب التجريبي باستعمال المحاكاة (Simulation) لما فيه من مميزات جيدة في محاكاة الواقع العملي للنماذج المستخدمة، إذْ تـُعرف المحاكاة بأنها أسلوب يتم من خلاله إيجاد أفضل أنْمُوذَج بديل مماثل للأنموذج الحقيقي من دون المحاولة للحصول على الأنموذج الحقيقي نفسه. وباستطاعتنا السيطرة على تجربة المحاكاة من خلال تنفيذ التجربة لعدة مرات وتغيير المعلمات ولاختبار سلوك الأنموذج تحت مختلف الشروط.

٥-٢ توليد المتغيرات

تمتاز تجارب المحاكاة باختصارها للوقت لتنفيذ العملية بدقائق قليلة على الحاسبة الإلكترونية فضلاً عن تمتع أسلوب المحاكاة بالمرونة العالية والحرية في اختيار حجوم العينات المختلفة وافتراض تباينات مختلفة للأخطاء العشوائية. فمثلا يعتمد اسْتِعْمال المحاكاة $(Box-Muller)$ (والتي تمثل خوارزمية حسابية تتضمن تكرار التجربة لمئات أو آلاف المرات) كليا على توليد المتغيرات العشوائية، إذْ تعتبر طريقة$(Box-Muller)$ [30]عن أسلوب المحاكاة بواسطة العينة،

(أولا) -توليد المتغيرات التوضيحية: *Explanatory variables*

وفي هذه الخطوة يتم توليد متغيرات توضيحية (Xi) ؛إذاْ يتم الاعتماد على أجهزة الحاسوب لتوليد ارقام عشوائية بأعداد كبيرة جداً وهناك شرطان رئيسيان لتوليد المتغيرات التوضيحية: الأول أنْ تتبع التوزيع المنتظم المستمر ضمن الفترة (1,0)، والثاني أنْ تكون مستقلة فيما بينها. وهذان الشرطان يكونان اساسيان في عملية التوليد. ويجب أنْ تتوافر صفة إعادة التوليد في الأرقام العشوائية المتولدة وذلك من أجل اختبار دقة البرنامج إضافة إلى امتياز عملية التوليد بالسرعة.

(ثانيا)- توليد الأخطاء العشوائية: Random Errors

وفي هذه الخطوة يتم توليد الأخطاء العشوائية (εi) التي تتوزع توزيعاً طبيعياً بتوقع قدره صفر وتباين ثابت قدره $(\widehat{\sigma^2})$

$$[\varepsilon i \sim N(0,\sigma^\wedge 2) \qquad i = 1,2, \quad \dots, n] \qquad \dots(15)$$

وباستعمال أشهر الطرائق واكثرها شيوعاً هي طريقة[31] $(Box - Muller)$ التي تعتبر إحدى طرائق توليد الأخطاء العشوائية التي تم اقترحها من قبل الباحثين: $(Box - Muller)$ في عام (١٩٥٨) والتي تعتمد على المحاكاة في توليد الأرقام العشوائية وباستخدام الحاسوب[7].

(ثالثا)- المتغير المعتمد Dependent variable

تم توليد المتغير المعتمد (yi) مباشرة من خلال النماذج المستعملة في تجارب المحاكاة وذلك باستعمال دالة الانحدار من خلال جمع بين المتغيرات التوضيحية التي تم توليدهما في الفقرة (اولا) أعلاه، مع الأخطاء العشوائية التي تم توليدها في الفقرة (ثانيا) أعلاه ولكل أُنْمُوذَج من النماذج قيد الدراسة وكما يأتي:

$$yi = \mu(xi\ ) + \varepsilon i \qquad i = 1, \cdots, n \qquad \dots (16)$$

إذْ:

(xi):قيم المتغير التوضيحي المتولدة في الفقرة (اولا)، εi: قيم الأخطاء العشوائية المتولدة في الفقرة (ثانيا)

μ :أُنْمُوذَج الدالة المراد تقديرها.

ولتوليد المتغير المعتمد من خلال المحاكاة وباستعمال عدد من الدوال ومنها (الدالة اللاخطية غير المتجانسة مكانيا والدالة متعددة الحدود من الدرجة الخامسة) تم الاعتماد على بعض الدراسات السابقة في تحديد هذه الدوال وكالاتي:

١. دالة متعددة حدود من الدرجة الخامسة وصيغتها: -

⟨Polynomials of the fifth degree⟩

$$y_i = 1 - 5x + 36x^2 - 53x^3 + 0.15x^4 + 22x^5 \qquad \dots (17)$$

٢. الدالة اللاخطية غير المتجانسة مكانياً وصيغتها:

⟨Spatially Heterogeneous Function⟩

$$y_i = \sqrt{x(1-x)}\sin\left(\frac{2\pi(1 + (2.5)^{\frac{9-4j}{6}}}{x + (1.5)^{\frac{9-4j}{6}}}\right) \qquad j = 6 \qquad \dots (18)$$

٥-٣ تحديد حجم العينة

تم اختيار حجوم مختلفة للعينة بشكل يتناسب مع معرفة مدى تأثير حجم العينة على دقة وكفاءة النتائج المستحصلة من طرائق التقدير المستعملة في الدراسة، وكما هو معلوم بانه كلما زاد حجم العينة كانت النتائج أفضل. وقد اعتمدت دراستنا على أربعة حجوم للعينات من اختيار الباحث ومن خلال الاطلاع على عدد من الدارسات السابقة والبحوث المنشورة في مجال الانحدار اللامعلمي[2]:حجوم العينات عند (٣٠ مشاهدة) حجوم العينات عند (٦٠ مشاهدة)حجوم العينات عند (١٢٠ مشاهدة) حجوم العينات عند (٢٤٠ مشاهدة)

٥-٣-١ تكرار حجوم العينات:

تم اختيار التكرار لأحجام العينات والمساوي الى (L= 100) لكل تجربة من أجل الحصول على دقة وتجانس للمقدرات.

٥-٤ نتائج تجارب المحاكاة

بعد إجراء تجارب المحاكاة والحصول على النتائج تم تحليل النتائج بالاعتماد على مقدرات $(kernel)$ اللامعلمية ومعايير المفاضلة كل من متوسط مربعات الخطأ $(AMSE)$ ومتوسط الخطأ المطلق$(AMAE)$ وتكامل مربعات الخطأ $(MISE)$ ثم حساب معدل كل منهما لعدد (100) من التكرارات وإجراء المقارنة واختيار أفضل قيمة لمعايير المفاضلة ووضع النتائج لكل أُنْمُوذَج على حده ويمكن تلخيص نتائج تجارب المحاكاة المطبقة كما موضح في أدناه:

ولاً. دالة متعددة حدود من الدرجة الخامسة وصيغتها: ⟨Polynomials of the fifth degree⟩

$$y_i = 1 - 5x + 36x^2 - 53x^3 + 0.15x^4 + 22x^5$$

لقد تم تطبيق مقدرات كيرنل(kernel) اللامعلمية على دالة متعددة حدود من الدرجة الخامسة ⟨$Polynomials\ of\ the\ fifth\ degree$⟩ والمبينة صيغتها أعلاه ومن خلال تحليل النتائج وبالاعتماد على معايير المفاضلة ($AMSE, AMAE\ , MISE$) من خلال المقارنة واختيار أفضل قيمة لمعايير المفاضلة يمكن أنْ نستنتج من النتائج ادناه أفضل طريقة لمقدرات كيرنل ($kernel$). إذْ تمّ الاعتماد على نوع من دوال كيرنل ($kernel$) كدالة وزن وهي دالة($Gaussian$) وكانت النتائج كما في الجدول (1) والأشكال: [(3), (2)]

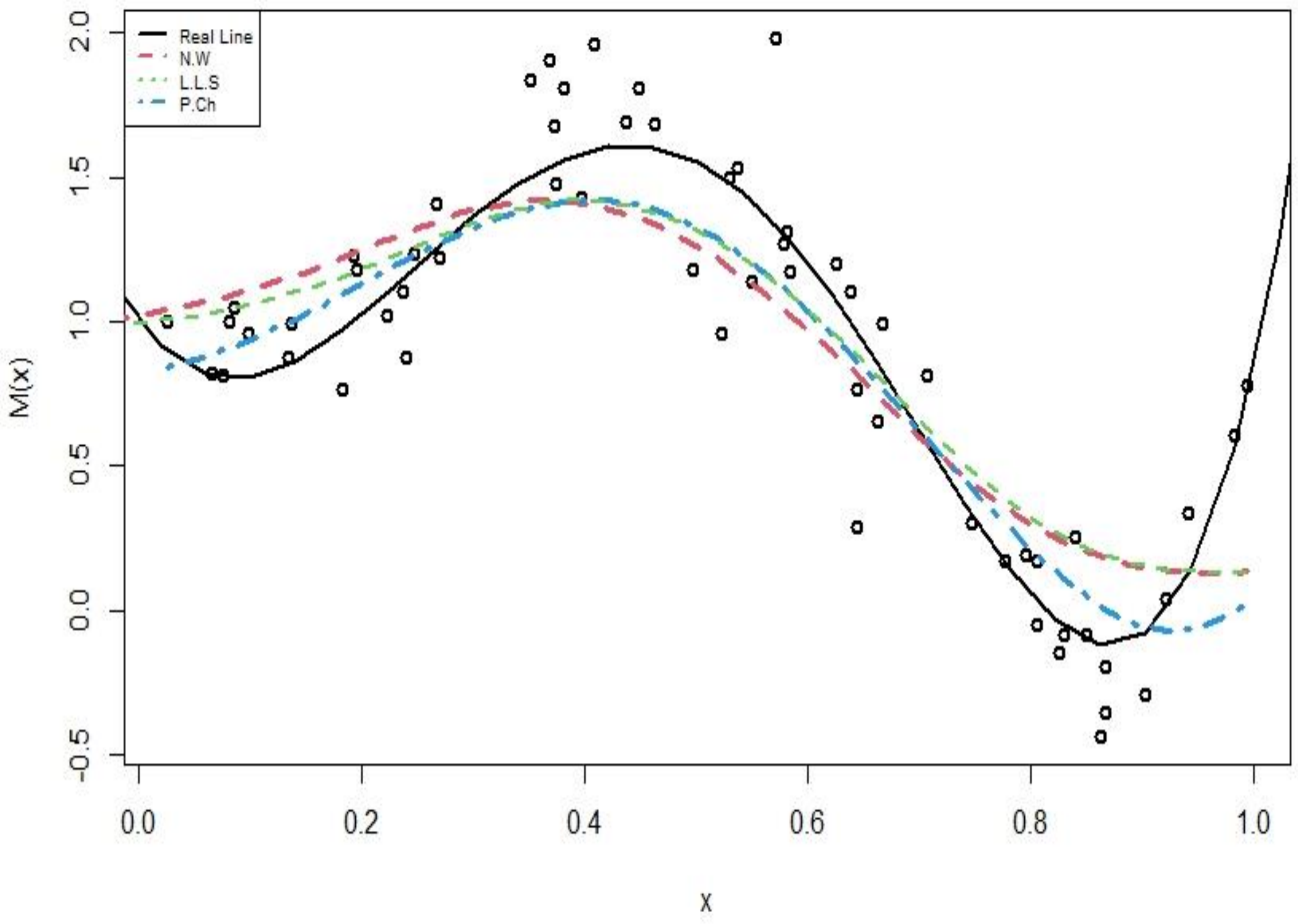
الجدول (1) يمثل نتائج معايير المفاضلة لمقدرات كيرنل اللامعلمية لدالة الوزن ($Gaussian$) وبالتعويض في دالة متعددة حدود

التسلسل	الطريقة	حجم العينة	AMSE	AMAE	MISE	AMSE	AMAE	MISE	AMSE	AMAE
		الانحرافات معيارية	σ=0.125			σ=0.25				
1.	N. W	30	0.6045618	0.6664444	0.006045	0.5552621	0.6511589	0.0055526	0.5702924	0.6503152
		60	0.5891503	0.6263402	0.005891	0.6493587	0.6515436	0.0064935	0.8461217	0.7497828
		120	0.3985405	0.5158953	0.003985	0.4369229	0.5437187	0.0043692	0.5838700	0.6295040
		240	0.4095028	0.5185011	0.004095	0.4667882	0.5503165	0.0046678	0.6670726	0.6502838
2.	L. L	30	0.6118123	0.6656583	0.006118	0.5639996	0.6521674	0.0056399	0.5788831	0.6512884
		60	0.6508021	0.6608934	0.006508	0.7104018	0.6823295	0.0071040	0.9069427	0.7721431
		120	0.5211032	0.5881524	0.005211	0.5633729	0.6179913	0.0056337	0.7179873	0.6979448
		240	0.4822941	0.5553794	0.004822	0.5418945	0.5869571	0.0054189	0.7473532	0.6910717
3.	P. Ch	30	0.6119262	0.6733836	0.006119	0.5776932	0.6579733	0.0057769	0.6240678	0.6815173
		60	0.5744065	0.6198247	0.005744	0.6338182	0.6428466	0.0063381	0.8290908	0.7414866
		120	0.3982285	0.5155036	0.003982	0.4366536	0.5435892	0.0043665	0.5836768	0.6294166
		240	0.4087704	0.5183261	0.004087	0.4659688	0.5501111	0.0046596	.6660773	0.6500575

من الدرجة الخامسة⟨Polynomials of the fifth degree⟩ عند أحجام عينات وانحرافات معيارية مختلفة:

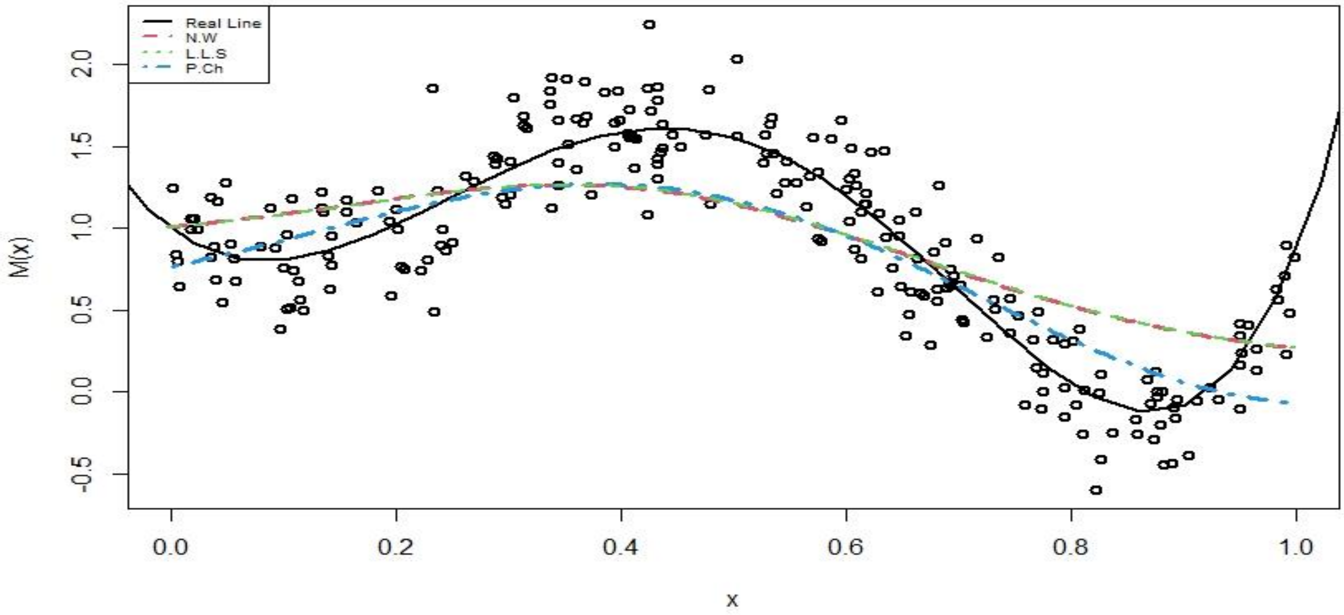
ISSN:2618-0278 Vol. 3 No. 6 June 2021

وفيما يأتي الرسوم لنتائج معايير المفاضلة التوضيحية لمقدرات كيرنل (kernel) لدالة الوزن (Gaussian) وبالتعويض في دالة متعددة حدود من الدرجة الخامسة ⟨Polynomials of the fifth degree⟩إذْ تم استِعْمال قيمة الانحراف المعياري تساوي ($\sigma = 0.25$) ولقيمتين لحجم العينة هي أصغر حجم عينة ($n = 30$) وأكبر حجم عينة وهي ($n = 240$):



الشكل رقم (٢)

الشكل أعلاه رقم (2) يمثل نتائج تجارب المحاكاة لمعايير المفاضلة المستخدمة في دالة متعددة حدود من الدرجة الخامسة ⟨Polynomials of the fifth degree⟩فعندما يكون حجم العينة يساوي ($n = 30$) وقيمة الانحراف المعياري تساوي ($\sigma = 0.25$)ويتضمن الشكل أعلاه كل من (مقدر الانحدار الثابت الموضعي ومقدر الانحدار الخطي الموضعي ومقدر بريستلي تشاو) لدالة الوزن (Gaussian).



الشكل رقم (٣)

الشكل أعلاه رقم (3) يمثل نتائج تجارب المحاكاة لمعايير المفاضلة المستخدمة في دالة متعددة حدود من الدرجة الخامسة ⟨Polynomials of the fifth degree⟩ فعندما يكون حجم العينة يساوي ($n = 240$) وقيمة الانحراف المعياري تساوي ($\sigma = 0.25$) ويتضمن الشكل أعلاه كل من (مقدر الانحدار الثابت الموضعي ومقدر الانحدار الخطي الموضعي ومقدر بريستلي تشاو) لدالة الوزن ($Gaussian$).

حيث أظهرت نتائج الجانب التجريبي في الجدول (١) والاشكال [(٢)،(٣)] وعلى اختلاف احجام العينات وقيم تباين الخطأ (الانحراف المعياري):

١. اظهرت نتائج وكما هو موضح في الجدول (١) ومن خلال المقارنة بين قيم معايير المفاضلة كل من معدل متوسط مربعات الخطأ (AMSE)ومعدل متوسط الخطأ المطلق (AMAE) وتكامل مربعات الخطأ (MISE) والنتائج المبينة اعلاه نلاحظ ان أفضل مقدر من بين المقدرات الأخرى والذي يمتلك أكثر عدد مرات أفضلية على بقية مقدرات كيرنل اللامعلمية هو مقدر بريستلي تشاو("P.Ch") . فيما عدا نتائج المقارنة التي كانت فيها افضلية معاير المفاضلة الثلاث الى مقدر الانحدار لثابت الموضعي ("N.W")عندما تكون قيم الانحراف المعياري تساوي (σ=0.25,)(σ=0.5,) وحجم العينة يساوي (n=30) أيضا عند حجم العينة (n=30) والانحراف المعياري (σ=0.125) اضهرت النتائج افضلية معياري المفاضلة كل من معدل متوسط مربعات الخطأ (AMSE) وتكامل مربعات الخطأ (MISE) لمقدر الانحدار لثابت الموضعي ("N.W") . وأظهرت النتائج أيضا افضلية معيار المفاضلة معدل متوسط الخطأ المطلق (AMAE) لمقدر الانحدار الخطي الموضعي(”L.L”) عند حجم العينة (n=30) والانحراف المعياري (σ=0.125) .

٢. كانت بعض نتائج قيم معايير المفاضلة (AMSE,AMAE ,MISE) جيدة حيث تقل قيمتها كلما زادت قيمة حجم العينة (n)، ويلاحظ أن نتائج قيم معايير المفاضلة تأخذ أكبر القيم عند حجم العينة الصغيرة (n=50) وعند مقارنتها مع نتائج قيم معايير المفاضلة عند حجم العينة الكبيرة (n=250) حيث تكون جيدة وتقل قيمتها الى ان بعض القيم تتأثر بمقدار الانحراف المعياري وحجم العينة وعرض حزمة او حجم نافذة ⟨h⟩ وكما هو موضح في الجدول (١) اعلاه.

٣. ومن خلال النتائج تم ملاحظة ان بعض قيم معايير المفاضلة (AMSE,AMAE ,MISE) تتناقص حيث تقل قيمتها كل ما قلة قيمة تباين الخطأ (الانحراف المعياري) ولأغلب الطرائق المستخدمة وكما هو موضح في الجدول (١) اعلاه.

٤. الأشكال اعلاه [(٢)،(٣)] تبين مدى افضلية مقدر بريستلي تشاو("P.Ch") على بقية مقدرات كيرنل اللامعلمية كما يظهر في الاشكل أعلاه ولأغلب احجام العينات والانحرافات المعيارية.

٥. كما يظهر في الأشكال اعلاه [(٢)،(٣)] ان مقدر بريستلي تشاو("P.Ch") كان الأقرب إلى المنحنى الحقيقي رغم تذبذب في البيانات وتقارب قيم معايير المفاضلة في وسط المنحنى.

ثانيا .الدالة اللاخطية غير المتجانسة مكانياً وصيغتها:⟨Spatially Heterogeneous Function⟩

$$y_i = \sqrt{x(1-x)}\,\sin\left(\frac{2\pi\left(1+(2.5)^{\frac{9-4j}{6}}\right)}{x+(1.5)^{\frac{9-4j}{6}}}\right) \qquad j = 6$$

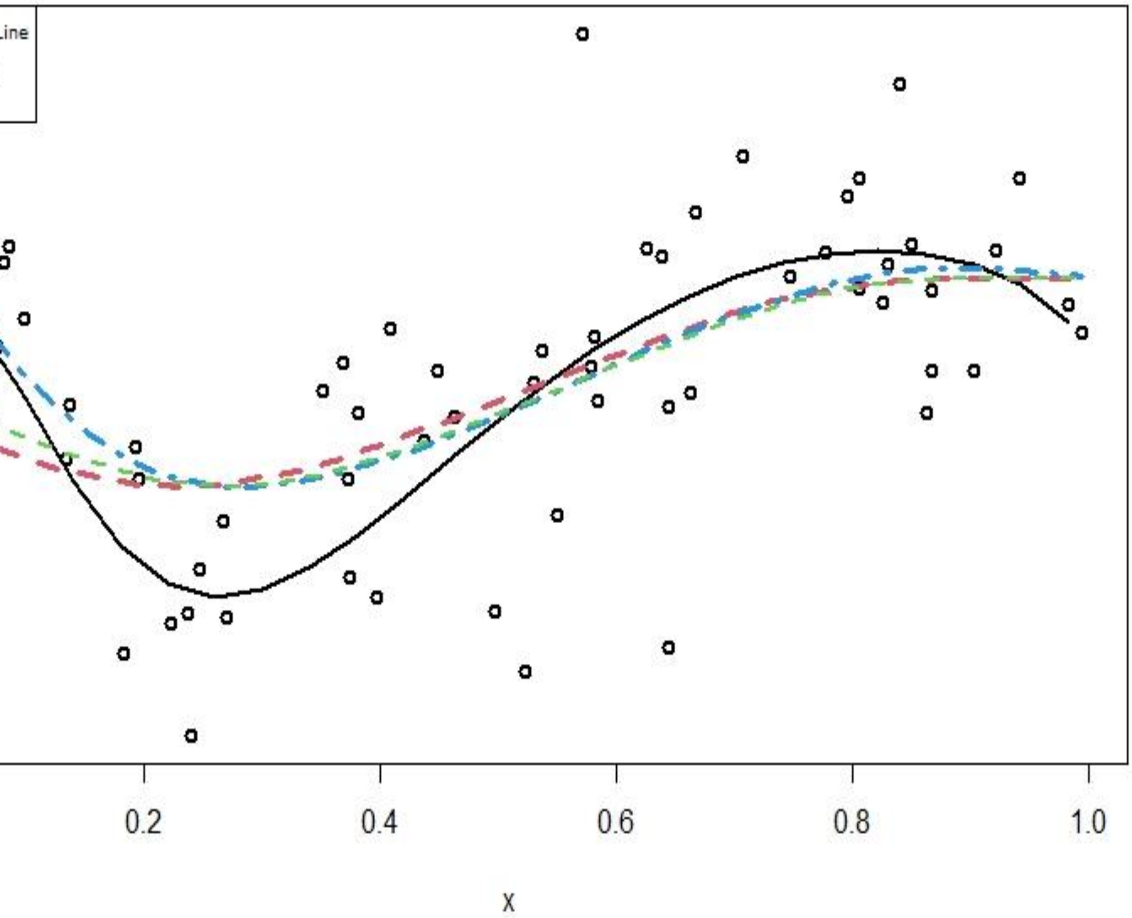
لقـــد تـــم تطبيـــق مقـــدرات كيرنـــل$(kernel)$ اللامعلميـــة علـــى دالـــة الدالـــة اللاخطيـــة غيـــر المتجانســـة مكانيـــاً $\langle Spatially\ Heterogeneous\ Function\rangle$ والمبينـة صـيغتها أعـلاه ومـن خـلال تحليل النتـائج وبالاعتمـاد علـى معـايير المفاضـلة $(AMSE, AMAE\ , MISE\ )$ من خلال المقارنـة واختيـار أفضـل قيمـة لمعـايير المفاضـلة يمكن أنْ نستنتج من النتـائج ادنـاه أفضـل طريقـة لمقدرات كيرنل $(kernel)$.إذْ تمّ الاعتماد على نوع من دوال كيرنـل $(kernel)$ كدالـة وزن وهـي دالـة $(Gaussian)$ وكانـت النتـائج كما في الجدول (2) والأشكال :$[(4), (5)]$

الجدول (2) يمثل نتائج معايير المفاضلة لمقدرات كيرنل اللامعلمية لدالة الوزن (Gaussian) وبالتعويض في الدالة اللاخطية غير المتجانسة مكانياً
⟨Spatially Heterogeneous Function⟩عند أحجام عينات وانحرافات معيارية مختلفة:

الانحرافات معيارية			σ=0.125			σ=0.25					
التسلسل	الطريقة	حجم العينة	AMSE	AMAE	MISE	AMSE	AMAE	MISE	AMSE	AMAE	
1.	**N. W**	30	0.1506806	0.325192	0.001506	0.245027	0.407457	0.002450	0.508485	0.548964	0.0
		60	0.0879926	0.241768	0.000879	0.112598	0.267983	0.001125	0.238156	0.366059	0.0
		120	0.0979552	0.262057	0.000979	0.136528	0.304277	0.001365	0.283858	0.429667	0.0
		240	0.0846551	0.234611	0.000846	0.121276	0.277363	0.001212	0.280232	0.421001	0.0
2.	**L. L**	30	0.1531047	0.330323	0.001531	0.245280	0.407798	0.002452	0.503850	0.554576	0.0
		60	0.0936844	0.248963	0.000936	0.120269	0.275527	0.001202	0.250781	0.377079	0.0
		120	0.1071837	0.273364	0.001071	0.144697	0.313800	0.001446	0.289801	0.437615	0.0
		240	0.0870208	0.236254	0.000870	0.124546	0.280962	0.001245	0.285856	0.423689	0.0
3.	**P. Ch**	30	0.1300051	0.300049	0.001300	0.243943	0.400477	0.002439	0.459254	0.532166	0.0
		60	0.0875509	0.239282	0.000875	0.112198	0.265265	0.001121	0.237942	0.367897	0.0
		120	0.0979136	0.262178	0.000979	0.136448	0.304237	0.001364	0.283692	0.429717	0.0
		240	0.0846176	0.234631	0.000846	0.121385	0.277278	0.001222	0.280332	0.421140	0.0

ISSN:2618-0278 Vol. 3 No. 6 June 2021

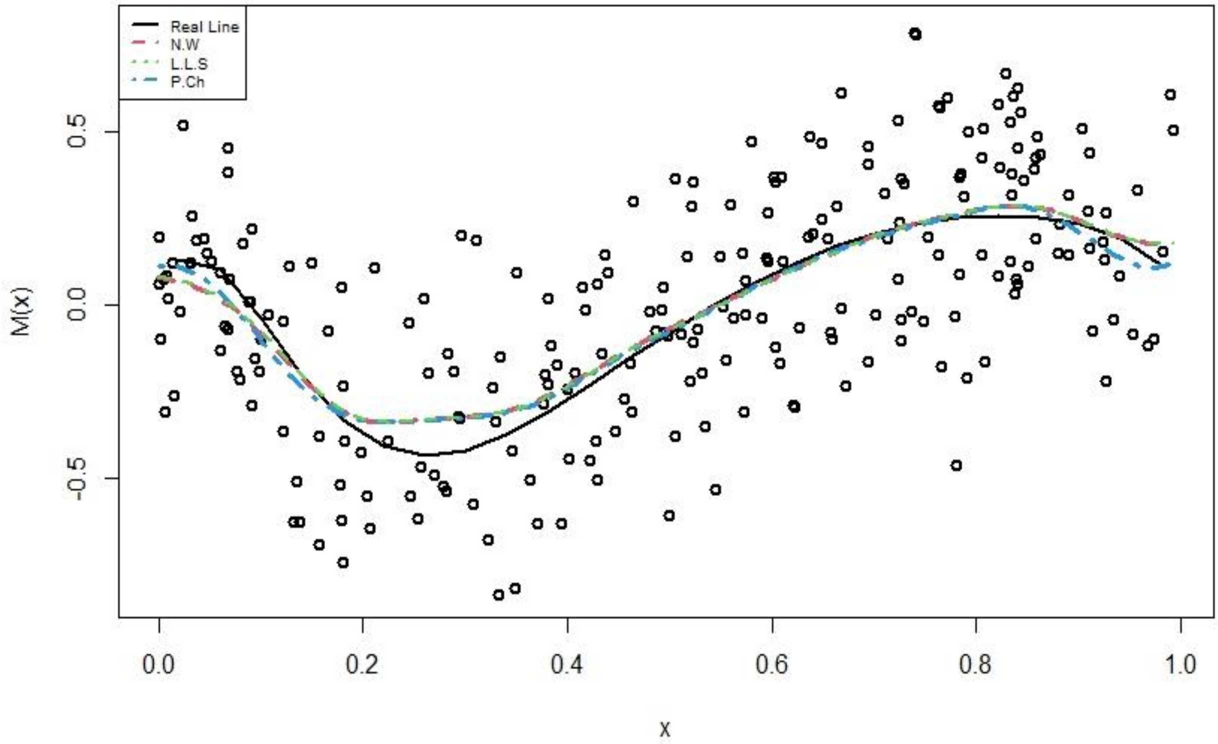
وفيما يأتي الرسوم لنتائج معايير المفاضلة التوضيحية لمقدرات كيرنل ($kernel$) اللامعلمية لدالة الوزن ($Gaussian$) لدالة اللاخطية غير المتجانسة مكانياً⟨Spatially Heterogeneous Function⟩إذْ تم اسْتِعْمال قيمة الانحراف المعياري تساوي ($\sigma = 0.25$) ولقيمتين لحجم العينة هي أصغر حجم عينة ($n = 30$) وأكبر حجم عينة وهي($n = 240$) :



الشكل رقم (٤)

الشكل أعلاه رقم (4) يمثل نتائج تجارب المحاكاة لمعايير المفاضلة المستخدمة في الدالة اللاخطية غير المتجانسة مكانياً ⟨Spatially Heterogeneous Function⟩فعندما يكون حجم العينة يساوي ($n = 30$) وقيمة الانحراف المعياري تساوي ($\sigma = 0.25$)ويتضمن الشكل أعلاه كل من (مقدر الانحدار الثابت الموضعي ومقدر الانحدار الخطي الموضعي ومقدر بريستلي تشاو) لدالة الوزن ($Gaussian$).

ISSN:2618-0278 Vol. 3 No. 6 June 2021



الشكل رقم (٥)

الشكل أعلاه رقم (5) يمثل نتائج تجارب المحاكاة لمعايير المفاضلة المستخدمة في الدالة اللاخطية غير المتجانسة مكانياً(Spatially Heterogeneous Function)فعندما يكون حجم العينة يساوي ($n = 2\ 50$) وقيمة الانحراف المعياري تساوي ($\sigma = 0.25$) ويتضمن الشكل أعلاه كل من (مقدر الانحدار الثابت الموضعي ومقدر الانحدار الخطي الموضعي ومقدر بريستلي تشاو) لدالة الوزن ($Gaussian$).

حيث أظهرت نتائج الجانب التجريبي في الجدول (٢) والاشكال [(٤)،(٥)] وعلى اختلاف احجام العينات وقيم تباين الخطأ (الانحراف المعياري):

١. اظهرت نتائج وكما هو موضح في الجدول (٢) ومن خلال المقارنة بين قيم معايير المفاضلة كل من معدل متوسط مربعات الخطأ (AMSE)ومعدل متوسط الخطأ المطلق (AMAE) وتكامل مربعات الخطأ (MISE) والنتائج المبينة اعلاه نلاحظ ان أفضل مقدر من بين المقدرات الأخرى والذي يمتلك أكثر عدد مرات أفضلية على بقية مقدرات كيرنل اللامعلمية هو مقدر بريستلي تشاو("P.Ch") . فيما عدا نتائج المقارنة التي كانت فيها افضلية معاير المفاضلة الثلاث الى مقدر الانحدار لثابت الموضعي ("N.W")عندما تكون قيم الانحراف المعياري تساوي (σ=0.5)، وحجم العينة يساوي (n=120) وأيضا عند حجم العينة (n=60) وقيمة الانحراف المعياري تساوي (σ=0.125) وأظهرت النتائج أيضا افضلية معيار المفاضلة معدل متوسط الخطأ المطلق (AMAE) لمقدر الانحدار لثابت الموضعي ("N.W") ولنفس المقدر أظهرت النتائج أيضا افضلية معيار المفاضلة معدل متوسط الخطأ المطلق (AMAE) عند حجم العينة (n=30) والانحراف المعياري (σ=0.125)، .

٢. كانت بعض نتائج قيم معايير المفاضلة (MISE, AMSE,AMAE) جيدة حيث تقل قيمتها كلما زادت قيمة حجم العينة (n)، ويلاحظ أن نتائج قيم معايير المفاضلة تأخذ أكبر القيم عند حجم العينة الصغيرة (n=50) وعند مقارنتها مع نتائج قيم معايير المفاضلة عند حجم العينة الكبيرة (n=250) حيث تكون جيدة وتقل قيمتها الى ان بعض القيم تتأثر بمقدار الانحراف المعياري وحجم العينة وعرض حزمة او حجم نافذة ⟨h⟩ وكما هو موضح في الجدول (٢) اعلاه.
٣. ومن خلال النتائج تم ملاحظة ان بعض قيم معايير المفاضلة (MISE, AMSE,AMAE) تتناقص حيث تقل قيمتها كل ما قلة قيمة تباين الخطأ (الانحراف المعياري) ولأغلب الطرائق المستخدمة وكما هو موضح في الجدول (٢) اعلاه.
٤. الأشكال اعلاه[(٤)،(٥)] تبين مدى افضلية مقدر بريستلي تشاو("P.Ch") على بقية مقدرات كيرنل اللامعلمية كما يظهر في الاشكل أعلاه ولأغلب احجام العينات والانحرافات المعيارية.
٥. كما يظهر في الأشكال اعلاه[(٤)،(٥)] ان مقدر بريستلي تشاو("P.Ch") كان الأقرب إلى المنحنى الحقيقي رغم تذبذب في البيانات وتقارب قيم معايير المفاضلة في وسط المنحنى.

٦. الاستنتاجـــــات

على ضوء ما توصل إليه الباحث من نتائج، يمكن طرح أهم الاستنتاجات التي تم استخلاصها وهي كالاتي:

١. بالنسبة للجانب التجريبي ومن خلال المقارنة بين معايير المفاضلة لكل من متوسط مربعات الخطأ $(AMSE)$ ومتوسط الخطأ المطلق$(AMAE)$ وتكامل مربعات الخطأ $(MISE)$ لكافة تكرارات تجارب المحاكاة وكحالة عامة أنْ أفضل طريقة لمقدرات كيرنل اللامعلمية هو مقدر بريستلي تشاو "$priestley - "p.ch$ $chao\ estimator$ حيث اظهر أفضلية واضحة على باقي المقدرات من خلال النتائج والأشكال في الجانب التجريبي ولكل حالة من حجوم العينات الاربع ومستويات الانحراف المعياري الثلاث وكذلك للنماذج المعتمدة في المحاكاة المتضمنة نتائج مقدرات كيرنل اللامعلمية.
٢. نلاحظ من خلال نتائج تنفيذ تجربة المحاكاة لتحديد مقدرات كيرنل اللامعلمية أنْ قيم معيار المفاضلة لتكامل مربعات الخطأ (MISE) حققت أفضل النتائج لكافة تكرارات التجربة لكل حالة من حجوم العينات الاربع ومستويات الانحراف المعياري الثلاث وكذلك للنماذج المعتمدة في المحاكاة المتضمنة نتائج مقدرات كيرنل اللامعلمية لكل حالة من حالات تجربة المحاكاة.
٣. نلاحظ أيضاً ومن نتائج تنفيذ تجربة المحاكاة لتحديد مقدرات كيرنل اللامعلمية أنْ قيم معيار المفاضلة لمعدل متوسط مربعات الخطأ$(AMSE)$ وتكامل مربعات الخطأ $(MISE)$ لكافة تكرارات التجربة لكل حالة من حجوم العينات الاربع ومستويات الانحراف المعياري الثلاث وكذلك للنماذج المعتمدة في المحاكاة كانت صغيرة ومتقاربة مما يشير إلى تجانس بين قيم المعيارين لكل حالة من حالات تجربة المحاكاة.
٤. جميع مقدرات كيرنل$(kernel)$ اللامعلمية التي تم تقديرها تتوافق مع المبدأ الاحصائي العام المتمثل بتناقص قيم معايير المفاضلة مع زيادة حجم العينة تتناسب عكسيا وطرديا مع الانحراف المعياري رغم التشتت الحاصل في البيانات وله عدة أسباب تعود إلى حجم (h).
٥. اسْتِعْمال الدالة اللاخطية غير المتجانسة مكانياً لتوليد المتغير المعتمد قد اعطتنا أفضل نتائج لجميع معيار المفاضلة لمقدرات كيرنل$(kernel)$ اللامعلمية لكافة تكرارات التجربة لكل حالة من حجوم العينات الاربع ومستويات الانحراف المعياري الثلاث.
٦. كما يظهر في الأشكال في الجانب التجريبي أنْ معيار المفاضلة لمقدرات كيرنل اللامعلمية عندما يكون حجم العينة يساوي $(n = 2\ 40)$ وقيمة الانحراف المعياري تساوي $(\sigma = 0.25)$وعند اسْتِعْمال أسلوب المحاكاة في الدالة اللاخطية غير المتجانسة مكانياً كانت النتائج الأقرب إلى المنحنى الحقيقي.

التوصيات:

بناءاً على ما توصل إليه الباحث من استنتاجات، ندرج أهم التوصيات:

١. نوصي بإعمام اسْتِعْمال مقدر بريستلي تشاو "$priestley - chao\ estimator\ "p.ch$عند اسْتِعْمال مقدرات كيرنل$(kernel)$ اللامعلمية لكفاءته في جميع دوال الاختبار المدروسة.
٢. نوصي بأسْتِعْمال معيار تكامل مربعات الخطأ (MISE) في المقارنة بين الطرائق المختلفة بدلاً عن المعايير الشائعة الأخرى وذلك لما حققه من نتائج جيدة في الجانب التجريبي.

ISSN:2618-0278 Vol. 3 No. 6 June 2021

٣. نوصـي باسـتخدام دوال وزن أخـرى فـي مقدرات كيرنـل $(kernel)$ اللامعلميـة كدالـة (Epanchnikov) أو دالـة (Uniform) أو دالة $(Quartic\ or\ Biweight)$ في تحديد المقدر الأفضل.

المصادر

أولا: المصادر العربية: (Arabic References)

حمزة. سعد كاظم، (٢٠٠٩)،"**مقارنة بعض الطرائق اللبية في تقدير نماذج الانحدار اللامعلمية بوجود بيانات تامة وغير تامة** "،رسالة ماجستير في الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد

ثانيا: المصادر الأجنبية: (Foreign References)

W. Härdle, *Applied nonparametric regression*, no. 19. Cambridge university press, 1994..

M. Strand, "Comparison of methods for monotone nonparametric multiple regression," *Commun. Stat. Part B Simul. Comput.*, vol. 32, no. 1, pp. 165–178, 2003.

M. Rosenblatt, "Remarks on Some Nonparametric Estimates of a Density Function," *The Annals of Mathematical Statistics*, vol. 27, no. 3. pp. 832–837, 1956.

E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–1076, 1962.

J. Fan and J. Zhang, "Two- step estimation of functional linear models with applications to longitudinal data," *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, vol. 62, no. 2, pp. 303–322, 2000.

M. Wand and B. Ripley, "KernSmooth: Functions for kernel smoothing for Wand & Jones (1995)," *R Packag. version*, vol. 2, pp. 19–22, 2006.

M. Kayri and G. Zirhlioglu, "Kernel smoothing function and choosing bandwidth for non-parametric regression methods," *Ozean J. Appl. Sci.*, vol. 2, no. 1, pp. 49–54, 2009.

C. M. Hurvich, J. S. Simonoff, and C. Tsai, "Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion," *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, vol. 60, no. 2, pp. 271–293, 1998.

B. W. Silverman, *Density estimation for statistics and data analysis*, vol. 26. CRC press, 1986.

V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Theory Probab. Its Appl.*, vol. 14, no. 1, pp. 153–158, 1969.

S. Shirahata and I.-S. Chu, "Integrated squared error of kernel-type estimator of distribution function," *Ann. Inst. Stat. Math.*, vol. 44, no. 3, pp. 579–591, 1992.

F. Hoti, "Kernel regression via binned data," *Res. Reports C38*, 2001.

H. Shimazaki and S. Shinomoto, "Kernel bandwidth optimization in spike rate estimation," *J. Comput. Neurosci.*, vol. 29, no. 1–2, pp. 171–182, 2010.

K. Wen and X. Wu, "Transformation-Kernel Estimation of Copula Densities," *J. Bus. Econ. Stat.*, vol. 38, no. 1, pp. 148–164, 2018.

Fan and Gasser, "Local polynomial fitting: a standard for nonparametric regression," North Carolina State University. Dept. of Statistics, 1993.

C.-S. Chee and Y. Wang, "Minimum quadratic distance density estimation using nonparametric mixtures," *Comput. Stat. Data Anal.*, vol. 57, no. 1, pp. 1–16, 2013.

Z. Q. Lu, "Multivariate locally weighted polynomial fitting and partial derivative estimation," *J. Multivar. Anal.*, vol. 59, no. 2, pp. 187–205, 1996.

D. Aydin, "A comparison of the nonparametric regression models using smoothing spline and kernel regression," *World Acad. Sci. Eng. Technol.*, vol. 36, pp. 253–257, 2007.

E. A. Nadaraya, "Some new estimates for distribution functions," *Theory Probab. Its Appl.*, vol. 9, no. 3, pp. 497–500, 1964.

G. S. Watson, "Smooth regression analysis," *Sankhyā Indian J. Stat. Ser. A*, pp. 359–372, 1964.

K. Vopatová, "Kernel choice with respect to the bandwidth in kernel density estimates," *ACTA Univ. MATTHIAE BELII, Ser. Math.*, vol. 18, pp. 47–53, 2011.

S. Rathore, M. Hussain, and A. Khan, "GECC: Gene expression based ensemble classification of colon samples," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 11, no. 6, pp. 1131–1145, 2014.

D. Ruppert, "Local polynomial regression and its applications in environmental statistics," Cornell University Operations Research and Industrial Engineering, 1996.

Y. K. Lee, E. Mammen, and B. U. Park, "Projection-type estimation for varying coefficient regression models," *Bernoulli*, vol. 18, no. 1, pp. 177–205, 2012.

Fan and Gijbels, "Variable bandwidth and local linear regression smoothers," *Ann. Stat.*, pp. 2008–2036, 1992.

M. B. Priestley and M. T. Chao, "Non- parametric function fitting," *J. R. Stat. Soc. Ser. B*, vol. 34, no. 3, pp. 385–392, 1972.

I. Pinelis, "Monotonicity preservation properties of kernel regression estimators," pp. 1–13, 2020.

T. Mspe, "Mean squared prediction error."

W. J. Thistleton, J. A. Marsh, K. Nelson, and C. Tsallis, "Generalized Box–Müller method for generating $ q $-gaussian random deviates," *IEEE Trans. Inf. theory*, vol. 53, no. 12, pp. 4805–4810, 2007.

E. R. Golder and J. G. Settle, "The Box- Müller Method for Generating Pseudo- Random Normal Deviates," *J. R. Stat. Soc. Ser. C (Applied Stat.*, vol. 25, no. 1, pp. 12–20, 1976.