

مجلة كلية التراث الجامعة

مجلة علمية محكمة
متعددة التخصصات نصف سنوية
العدد الأربعون

30 آب 2024
ISSN 2074-5621



رئيس هيئة التحرير

أ.د. جعفر جابر جواد

1988

مدير التحرير

أ.م. د. حيدر محمود سلمان

رقم الايداع في دار الكتب والوثائق 719 لسنة 2011

مجلة كلية التراث الجامعة معترف بها من قبل وزارة التعليم العالي والبحث العلمي بكتابها المرقم
(ب 3059/4) والمؤرخ في (2014/ 4/7)



About Nonparametric Regression Model

Estimation : A review

م.م. شرين علي حسين¹ أم.د. سعد كاظم حمزة²

¹ قسم ضمان الجودة والاداء الجامعي ، رئاسة جامعة بغداد ، بغداد ، العراق

² كلية الادارة والاقتصاد ، قسم الاحصاء ، جامعة بغداد ، بغداد ، العراق

المستخلص

على الرغم من ان نماذج الانحدار اللامعلمي هي اضعف وصفا من نماذج الانحدار المعلمي الا انها وبنفس الوقت تحتاج الى قيود او شروط اقل من النماذج المعلمية وهذا الامر تحديدا هو الذي جعل من نماذج الانحدار اللامعلمية اداة مرغوبة جدا لدى الباحثين كون ان البيانات الفعلية ليست دوما لديها مواصفات مثالية مما ادى الى انتعاش طريقة التفكير بنماذج الانحدار اللامعلمي مما ادى الى تطور اساليب كثيرة في العقود الاخيرة وفي مجالات عديدة كما في الهندسة والطب والجغرافيا والاقتصاد . وبالتالي، يتم تقديم تقنية لاستبدال هذه القيم المفقودة بقيم تقديرية معينة من أجل التعامل بشكل صحيح مع البيانات الجزئية (المفقودة). تم استخدام مجموعة متنوعة من التقنيات لتقدير نماذج الانحدار اللامعلمية . لذلك، يتم تقديم تحليل شامل للطرق المختلفة المستخدمة للتقدير دوال الانحدار اللامعلمي. يحدد الأنواع المختلفة من الاساليب المطروحة ويشير إلى أكثر أنواع تحليل الانحدار شيوعاً. الهدف الأساسي من ورقة المراجعة هذه هو تعريف القارئ بالعديد من الاساليب اللامعلمية المستخدمة في تحليل الانحدار مع توضيح كيفية تأثير هذه الطرق على الاستنتاجات النهائية للدراسة.

الكلمات المفتاحية: التقديرات اللبية ، التقديرات الشرائحية ، تحويل فوارير ، التحويل المويجي المتقطع ، التحويل المويجي المستمر ، الانحدار اللامعلمي.

1. المقدمة

تعد نماذج الانحدار واحدة من اهم اصناف النظرية الاحصائية وذلك لما قدمته للباحثين في شتى المجالات العلمية والانسانية من حلول علمية لمشاكلهم وبسبب تنوع مجالات عملها كان لابد من تنوع اشكالها هي الاخرى ، وان هذا التنوع يعود سببه لاختلاف طبيعة الموارد المتاحة للباحثين من طبيعة البيانات من جهة ومن جهة اخرى بسبب اختلاف ما الذي يريده الباحثون من معالجة مشكلة ما . على اساس ذلك قسمت نماذج الانحدار الى صنفين اساسيين بحسب طبيعة البيانات وهي نماذج الانحدار المعلمية والتي تقوم على ايجاد العلاقة بين السبب ونتيجته من خلال عدد من النقاط الاساسية والتي تصف العلاقة من خلال معالم الانموذج المتمثلة بنقطة تقاطع خط الانحدار مع المحور الصادي β_0



وكذلك مقدار الميل او الميول $(\beta_1, \beta_2, \dots, \beta_k)$. الا ان استعمال هكذا نوع يتطلب العديد من الشروط الاولية التي يجب توفرها لكي تكون قراءة هذه النماذج قراءة صحيحة كما تتطلب وجود بيانات من النوع الكمي ولها توزيع معروف بالإضافة الى العلاقة الخطية بين المتغيرات التوضيحية والمتغير التابع .

الا ان هذه الشروط الصارمة دفعت الباحثين الى البحث عن شروط اقل صرامة يمكن التعامل معها خاصة وان البيانات في الواقع العملي غالباً لا تحقق الشروط المفروضة لتطبيق النماذج المعلمية بسبب تطور الحياة وتداخل العوامل فيما بينها وتداخل العوامل فيما بينها مما ادى الى ظهور النماذج اللامعلمية والتي تقوم على ايجاد العلاقة من خلال منحني يصف تلك العلاقة اذ يكون الباحث في مثل هكذا نماذج مهتماً بإعطاء وصفاً عاماً للعلاقة وليست دراسة تفاصيل العلاقة الدقيقة ، وعلى الرغم ان نماذج الانحدار اللامعلمي هي اضعف وصفاً من نماذج المعلمية الا انها بنفس الوقت تحتاج الى قيود وشروط اقل وهذا الامر تحديداً هو الذي جعل منها اداة مرغوبة جداً لدى الباحثين كون ان البيانات الفعلية ليست دوماً لديها مواصفات مثالية . مما ادى الى انتعاش طريقة التفكير بنماذج الانحدار اللامعلمي وتطورت بشكل كبير في العقود الاخيرة وفي مجالات عديدة منها الهندسة والطب والجغرافية والاقتصاد بل تعدى الامر الى دراسات مثل الفلك والدراسات الجنائية .

2. تمهيد الانحدار Regression smoothing

اقترح Mahalambis عام (1921) تحليل الانحدار الحر (Distribution Free Regression) واطق عليه في حينها التحليل الرسومي المرن (Tractile Graphical Analysis) وقد اثبت انه يمتاز بالاتساق الا ان الهدف الاكبر هو ايجاد توزيع

مرن بدلاً من ان يتبع المنحنيات العشوائية باستخدام مجموعات كبيرة من البيانات التي تشترط شروطاً صارمة قد تكون غير موجودة او متوفرة في الواقع التطبيقي [6,10].

3. طرائق اختيار التمهيد Choice of Smoothing Method

من الاحداث المتكررة هي النقاش الدائم بين الباحثين الاحصائيين بشكل عام او ان يدور نقاش بين باحثين تخصصوا بتمهيد الانحدار مع اخرين لم يجربوا هذا الاسلوب بخصوص اي من هذه الاساليب اكثر فعالية وفضلها ، يمكن تقديم العديد من الاجابات حول ذلك اعتماداً الى حد كبير على الشخص الذي يتم سؤاله بأي مجال من مجالات التمهيد اكثر تخصصاً ، الا انه بشكل عام لكل اسلوب من اساليب التمهيد له اتباعه الراسخون ، الذين يتبنون نقاط القوة فيه ونقاط القوة المفضلة لديهم من ذلك الاسلوب ، الا ان جميع نقاشات الباحثين باختلاف اختصاصاتهم ذهبت الى استنتاج عام مفاده ان الجميع على حق في خياراتهم المختلفة .

وان ما يميز تفضيل اسلوب على اخر هو عوامل غير قابلة للمقارنة مثل :

1- الوفرة (Availability) اي هل ان الاسلوب موجود في حزم البرامج المتاحة ؟

2- قابلية التفسير (ماذا يخبرنا التمهيد عن البيانات ؟)



3- الكفاءة الاحصائية (ما مدى قرب الممهد من المنحنى الحقيقي؟)

4- قابلية الحساب السريع .

وبناءً على اعلاه فان جميع اساليب التمهيد التي سيتم طرحها والتي تم افتراضها خلال حقبات زمنية متفاوتة او انها ظهرت لتعالج القصور في سابقتها من الاساليب لها نقاط قوة وضعف مختلفة ولا يمكن ان تهيمن اي من هذه الاساليب على اسلوب اخر في جميع الخواص اذ يمكن ان يكون اي اسلوب او طريقة هو الافضل وذلك ببساطة عن طريق الوزن الشخصي المناسب للعوامل المختلفة المشار اليها سابقاً [11,13].

4. الممهيات اللبية Kernel Smooth

تعد الممهيات اللبية الانطلاقة الحقيقية لتقدير دالة الانحدار التي لا تحقق بياناته الشروط الاساسية لتطبيق الانحدار المعلمي ، ولفهم عمل هذه الممهيات ، لتكن x_i حيث ان $i = 1, \dots, n$ يشير الى عينة عشوائية من مجتمع معين له دالة كثافة $f(x)$ ، ولتكن y_i تشير الى قيم متغير الاستجابة ، وبافتراض ان $f(x, y)$ تشير الى دالة الكثافة الاحتمالية المشتركة للمتغيرين فان دالة الانحدار يشار اليها $m(x) = E(y/X = x)$ ، ولدالة التباين الشرطية $\sigma^2(x) = var(y/X = x)$. هنالك العديد من طرائق التمهيد اللبية المستعملة لتقدير دالة الانحدار ومنها [14,21] :

1.4 ممهد ناداريا- واتسون Nadarya-Watson

قدم هذا الممهد من قبل الباحثان (Watson & Nadarya) عام 1964 ، ويعد من المقدرات اللبية الاكثر شيوعاً بالاعتماد على طريقة متسلسلة الاوزان ، يستخدم هذا المقدر في حالة كلاً من التصميم العشوائي والثابت لكن غالباً ما يستخدم في حالة التصميم العشوائي ويمتاز بانه يقدم تقديراً مستمراً للانحدار عندما تكون الدالة اللبية المستخدمة مستمرة ويعرف هذا الممهد رياضياً وفق الاتي [6,21] :

$$\hat{m}(x) = \frac{\sum_{i=1}^n k\left(\frac{X_i - x}{h}\right) y_i}{\sum_{i=1}^n k\left(\frac{X_i - x}{h}\right)} \quad (1)$$

اذ ان (X_i) هي متجه افقي عشوائي من المتغيرات التوضيحية (X_i) ، h هي عرض الحزمة او ما تسمى بمعلمة التمهيد (Smoothing Parameter) ، $k\left(\frac{X_i - x}{h}\right)$ هي الدالة اللبية (Kernel Function) .

ويمكن اعادة صياغة المعادلة رقم (1) بصيغة المصفوفات وكالاتي

$$\hat{m}(x) = WY \quad (2)$$

اذ ان $Y = (Y_1, Y_2, \dots, Y_n)$ ، $W = (W_1', W_2', \dots, W_n')$ اذن

$$w_i = \frac{k\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n k\left(\frac{X_i - x}{h}\right)} \quad (3)$$

2-4 اختيار الدوال اللبية Selection of Kernel Functions

وهو الجزء الاول من دالة التقدير اللبية (NW) ويكون اقل اهمية في كفاءة التقدير ، اذ ان هنالك نوعان من الدوال اللبية يمكن تمييزهما للتقدير اللامعلمي لدوال الانحدار وهما : الدوال ذات اقل تباين والتي تعمل على التباين المحاذي ، والدوال المثلى والتي تعمل على تقليل تكامل مجموع مربعات الخطأ (MISE) المحاذي اي اشتقاقه بالنسبة للدالة اللبية .

ان الدالة اللبية ذات اقل تباين تمثل حل لمسألة التصغير الاتية [16,19].

$$V(k) = C_k = \int_{-1}^1 k^2(u) du \quad (4)$$

اما التحيز المحاذي و (MISE) المحاذي فيتم تعريفهما كالآتي :

$$B = d_k = \int_{-1}^1 u^2 k(u) du \quad (5)$$

$$MISE = E \int (\hat{m}(x) - m(x))^2 dx \quad (6)$$

وفيما يلي جدول يوضح بعض الدوال اللبية الاكثر استخداماً

جدول رقم (1) يبين بعض الدوال اللبية

Name	Explicit Form
Gaussian	$k(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right), u \in [-\infty, \infty]$
Uniform	$k(u) = \frac{1}{2}, u \in [-1, 1]$
Triangular	$k(u) = (1 - u), u \in [-1, 1]$
Epanechnikov	$k(u) = \frac{3}{4}(1 - u^2), u \in [-1, 1]$
Quartic	$k(u) = \frac{15}{16}(1 - u^2)^2, u \in [-1, 1]$
Triweight	$k(u) = \frac{35}{32}(1 - u^2)^3, u \in [-1, 1]$



3-4 عرض الحزمة (المعلمة التمهيدية) Smoothing Parameter

وتدعى في بعض الاحيان بالمعلمة التمهيدية او سعة القيد ويرمز لها بالرمز h وتسمى احياناً بمعلمة الانتشار وتعد الجزء الاهم في دقة التقدير وكفاءته ، لذلك اهتم العديد من الباحثين بطرائق اختيار عرض الحزمة للحصول على تقريب ملائم لغرض الموازنة بين التحيز والتباين وبسبب وجود علاقة طردية بين عرض الحزمة ومقدار التحيز وعلاقة عكسية بينها وبين التباين (اي بزيادة عرض الحزمة يزداد التحيز ويقل التباين والعكس بالعكس) لذلك يتوجب على الباحث اختيار عرض الحزمة بطرائق معينة للوصول الى مرحلة التوازن بين التحيز والتباين فضلاً عن تآثر شكل المنحني بقيمة معلمة التمهيد وذلك عن طريق صغر قيمتها والباعدة من الصفر وفي ادناه نوضح بعض طرائق اختيار عرض الحزمة الاكثر شيوعاً [22,36].

4-4 طريقة التقاطع الشرعية Cross- Validation

اقترحت من قبل Rudemo عام (1982) وعدل عليها Bowman عام (1984) وهي من اوائل اختيار معلمة التمهيد المميزة والجذابة وان فكرتها انطلقت من توسيع (MISE) لدالة $\hat{f}(\cdot; h)$ والحصول عليها وفق الاتي:

$$MISE = E \int \hat{f}(x; h)^2 dx - 2E \int \hat{f}(x; h) f(x) dx + \int f(x)^2 dx \quad (7)$$

وان المقدر غير المتحز هو :

وتطور هذا الاسلوب يعتمد على حذف مشاهدة واحدة في كل مرة واصبح يسمى مقدر ترك مرة واحدة (Leave-one-out) من قبل كلاً من Hau & Marron عام (1991) ثم طور هذا الاسلوب ليكون الحذف نصف المشاهدات في كل مرة من قبل الباحث Nason عام (1996) [15,28].

5-4 طريقة الملئ Plug - in

وهي من طرائق اختيار عرض الحزمة المتغيرة اقترحت من قبل Hall & Marron عام (1987) من خلال ايجاد عرض حزمة امثل تقاربي من خلال الصيغة الاتية [15]:

$$\hat{h}_{opt} = (0.25\hat{V}_2 / \hat{B}_3)^{1/3} n^{-1/3} \quad (10)$$

حيث ان :

$$\hat{V}_2 = 2A_1(k)\hat{D}_2 F \quad (11)$$

6-4 طريقة قاعدة الابهام Rule of Thumb



تعود هذه الطريقة الى الباحث Deheuvels والتي تعتمد على استبدال الدالة المجهولة f بدالة التوزيع وقد تم تعميمها من قبل الباحث Silverman عام (1986) حيث اشار بان قاعدة الابهام تعطي اختيار سريع وتكتب الصيغة العامة لها :

$$h_{Rule} = 1.06\hat{\sigma}n^{\frac{1}{5}} \quad (12)$$

يشير المقدار $\hat{\sigma}$ الى قيمة الانحراف المعياري للعينة .
اذ تدل ν على درجة Kernel والتي تمثل اول عزم غير صفري وذلك كون العزوم من الدرجة الفردية تساوي صفر لذلك فان ν تكون عدداً زوجياً دائماً ، بينما يشير المقدار $\hat{\sigma}$ الى قيمة الانحراف المعياري للعينة .

وتعتبر $CV(k)$ ثابتة وكما في الجدول رقم (2) ادناه والمعتمدة على ν (درجة Kernel) [39]

جدول رقم (2) يبين قيم $CV(k)$ لكل دالة وحسب الدرجة اللبية

درجة Kernel	V=2	V=4	V=6
Biweight	2.78	3.39	3.84
Gaussian	1.06	1.08	1.08

7-4 ممهد الانحدار الخطي المحدد Local Linear Regression Smoother

لمعرفة فكرة هذا الممهد فيجب ان نفترض ان المشتقة الثانية لـ $m(x)$ تكون موجودة واننا سنقوم بايجاد قيمتي a, b وذلك بتصغير المقدار التالي [32] :

$$\sum_{i=1}^n (Y_i - a - b(x - X_i))^2 K\left(\frac{x - X_i}{h}\right) \quad (13)$$

اذ ان : $K(0)$: يمثل دالة Kernel ، h : يمثل عرض الحزمة .
 a, b : يمثلان حل لمسألة المربعات الصغرى الموزونة وبحساب بسيط ينتج ان :

$$\hat{a} = \frac{\sum_{i=1}^n W_i Y_i}{\sum_{i=1}^n W_i} \quad (14)$$

لذلك فان ممهد الانحدار الخطي المحدد يساوي :

$$\hat{m}(x) = \frac{\sum_{i=1}^n W_i Y_i}{\sum_{i=1}^n W_i} \quad (15)$$

اما W_i فيعرف كالآتي :

$$W_i = K\left(\frac{x - X_i}{h}\right)[S_{n,2} - (x - X_i)S_{n,1}] \quad (16)$$

$$S_{n,i} = \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)(x - X_i)^i \quad i=1,2 \quad (17)$$

نلاحظ من الصيغة المذكورة انفاً $\hat{m}(x)$ يمثل معدل موزون الى الاستجابات ويدعى بالممهد الخطي ايضاً وايضاً ان دالة الوزن W_i تحقق الشرط الاتي :

$$\sum_{i=1}^n (x - X_i)W_i = 0 \quad (18)$$

وهذه الخاصية تضمن التحيز للممهد لا يعتمد على المشتقات لدالة الكثافة الاحتمالية لـ x يمتلك هذا الممهد خصائص معاينة مهمة تتمثل بالاتي :

- ان هذا الممهد يتكيف مع كلاً من التصاميم العشوائية والثابتة ، اي انه يستعمل مع هذه التصاميم وكذلك يستعمل مع دوال تصميم مختلفة .
- ان هذا الممهد سيكون افضل ممهد خطي اي انه يكون اصغر – اكبر ممهد خطي محاذي عند دالة الانحدار تكون من نوع الدوال التي تمتلك مشتقة ثانية .
- اما الخصائص المحاذية لهذا المقدر فتكون :
- ان دالة الانحدار $m(0)$ تمتلك مشتقة ثانية محددة .
- ان دالة الكثافة الاحتمالية $f(0)$ تحقق شرط Lipschitz او Holder .

$$|m(x) - m(y)| \leq c|x - y|^a, \quad 0 < a < 1, \quad m(x) > \quad (19)$$

اذ ان c يمثل ثابتاً موجباً ، والتباين الشرطي $\sigma^2(x) = \text{var}(y/X = x)$ يكون محدداً ومستمرأ .
وان دالة Kernel تكون دالة كثافة مستمرة وتحقق :

$$\int_{-1}^1 k(y)dy = b \quad (20)$$

$$\int_{-1}^1 yk(y)dy = 0 \quad \int_{-1}^1 y^2k(y)dy \neq 0 \quad (21)$$

$$\int_{-1}^1 y^{2r}k(y)dy < \infty, \quad r=1,2,\dots \quad (22)$$

ومن الجدير بالذكر ان ممهد LLS يكتب في بعض الاحيان بالصيغة :

$$\hat{m}^*(x) = \frac{\sum_{i=1}^n W_i y_i}{\left(\sum_{i=1}^n W_i + n^{-2}\right)} \quad (23)$$

تجنباً لوجود الصفر في المقام ومن ثم زواله .

8-4 ممهد الشرائح Spline Smoothing



هذا الممهد يعتبر ايضاً من الطرائق المهمة لتقدير دالة الانحدار وبالتالي امكانية رسم شكل المنحنى للدالة f ، يعتمد هذا الممهد على مجموع مربعات البواقي فضلاً عن ضرورة وجود المشتقة الثانية للدالة f وتكتب الصيغة العامة كالآتي [1,2]:

$$S(f) = \sum_{i=1}^n (Y_i - f(x_i)) - h \int_{\min x_i}^{\max x_i} (f(x_i))^2 dx \quad (24)$$

اذ يمثل الحد الاول من المعادلة اعلاه مجموع مربعات البواقي بينما يمثل الحد الاخر المشتقة الثانية للدالة f ويسمى حد الجزاء ، ويشير الرمز h الى معلمة التمهيد الذي بدوره يتحكم بكمية الجزاء اذ كلما تزداد قيمة معلمة التمهيد فان منحنى الدالة يصبح ثابتاً ، اذ يفوق المقدار الاخير مجموع مربعات البواقي بينما يخفتي حد الجزاء عندما تقترب معلمة التمهيد من الصفر .

1-8-4 دالة الشريحة Spline Function

تسمى دالة الشريحة S من الدرجة r عند العقد $\{x_i\}_{i=1, \dots, k}$ حيث ان $\gamma < x_0 < x_1 < x_k < x_{k+1} < \gamma_k$ اذا كانت :

1- S متعدد حدود من درجة لا تزيد عن r ضمن الفترة $[x_i, x_{i+1}]$.

2- $S \in C^r$ ، $S \in C^{r-1}, \dots, S \in C^1$ هي دوال مستمرة عند العقد S_i .

ومن اكثر الشرائح شيوعاً هي الشرائح التكعيبية بسبب خصائص الاستمرارية التي تملكها والتي تكتب وفق الصيغة التالية :

$$S(x) = b_0 + b_1 X + b_2 X^2 + b_3 X^3 + \sum_{i=1}^k g_i (x_i - x_j)^3 \quad (25)$$

ويطلق عليها ايضاً (دالة متعددة الحدود المبتورة) .

حيث ان $(x_i - x_j)_+^3 = (x_i - x_j)^3$ عندما $x > x_j$ وتساوي صفر للقيم الاخرى .

اما الشريحة التكعيبية الطبيعية (Natural C.S) فنتحقق باضافة الشرط التالي للشرطين السابقين .

3- الدالة S خطية في الفترتين $(-\gamma, x_1), (x_k, \gamma)$ [33] .

2-8-4 الشرائح من نوع B (Splines of Type B)

يتكون هذا النوع من الشرائح من قطع من متعددات حدود تصل بين نقاط ربط تصل بينها نقاط ربط تسمى (عقد) ولقد وضع (De Boor) صيغة تكرارية لحساب الشرائح من اي درجة مع عقد بمسافات غير متساوية بعد تعريف الشريحة من الدرجة صفر .

كما وضع صيغة تستخدم المسافات المتساوية بين العقد وكما يلي :

إذا كانت $B_j(x; q)$ تمثل قيمة الشريحة B_j من الدرجة q عند النقطة x لمتابعة من العقد بمسافات متساوية فإن الشريحة من الدرجة صفر تعرف كما يلي :

$$B_j(x; q) = 1 \quad (j-1)dx \in x - x_{\min} < jdx \quad (26)$$

$0 \qquad \qquad \qquad o.w$

حيث ان $dx = (x_{\max} - x_{\min})/n$ وان $n \in \mathbb{Z}$ تمثل عدد الفترات بين العقد وبعد حساب الشريحة من الدرجة صفر تحسب بقية الشرائح من اي درجة بالصيغة التكرارية التالية :

$$B_j(x, q) = \frac{q+p-j}{q} B_{j-1}(x; q-1) + \frac{j-p}{q} B_j(x; q-1) \quad (27)$$

حيث ان $p = x - x_{\min}, B_0(x, q) = 0$ كما توجد صيغة اخرى لحساب الشرائح B كفروقات للشرائح (JBF) وهي :

$$B_j(x; q) = -I^{q+1} D^{q+1} f_j(x; q) / (h^p p)$$

حيث ان $f_j(x; q) = (x - x_j)_+^q$.

ان الشرائح B اكثر استقراراً من الناحية العددية من الشرائح (TPF) ولها العديد من الخصائص التي تجعلها اكثر ملائمة منها :

$$1 - B_j(x; q) = 0 \text{ لاي درجة } q .$$

$$2 - B_j(x; q) > 0 \text{ حيث ان } [2,33] \cdot t_j < x < t_{j+q}$$

9-4 انحدار الشرائح الممهدة Smoothing Spline Regression

ان القياس الشائع لمدى مطابقة الدالة g لمجموعة من البيانات هو مجموع مربعات البواقي $\sum_{i=1}^n (Y_i - g(X_i))^2$ ولكن استخدام هذا المعيار وحده فقط امر غير مقبول من قبل الاحصائيين حيث ينتج تقدير متذبذب لاستكمال البيانات (Interpolate Data) لذلك يتم اضافة ما يسمى الجزاء غير ممهد (Roughness Penalty) الى المعيار السابق لينتج معيار المربعات الصغرى الجزائية (PRSS) وكالاتي :

$$PRSS = \sum_{i=1}^n (Y_i - g(X_i))^2 + I_a^b (g(m(x)))^2 dx \quad (29)$$

وبتصغير المعيار السابق امام صف من الدوال g حيث ان $g \in w_2[a, b]$ في فضاء (Sobolv) فان المقدر الناتج يسمى الشريحة الممهدة من الدرجة $(2m-1)$ وكحالة خاصة عندما $(2m-1)$ فان الشريحة الممهدة تسمى شريحة ممهدة تكعيبية (Cubic Smoothing Spline) وعلى فرض $a=0, b=1$



وتحقيق الدالة g (لشروط الحد الطبيعية) فان المقدر الناتج يسمى شريحة تكعيبية طبيعية بعقد عند X_i وللحصول على عمليات حسابية $o(n)$ قام (Reinch) بوضع صيغة تربيعية للجزء غير الممهّد وكما في الصيغة الآتية :

(30)

$$(g(m(x)))^2 dx = g^T kg$$

حيث k مصفوفة الدرجة $n \times n$ وان $k = QR^{-1}Q^J$ حيث المصفوفة Q هي مصفوفة (Tridagonal) من الدرجة $n'(n-2)$ والمصفوفة R هي مصفوفة (Tridagonal) متماثلة من الدرجة $(n-2)'(n-2)$ ، اما I تسمى المعلمة التمهيدية والتي تلعب دوراً كبيراً في التوازن بين جودة التوافق المقاسة بواسطة مجموع مربعات البواقي وبين كمية التمهيد المقاسة بواسطة الجزء غير الممهّد .

عندما $I \in \mathbb{R}_0$ سوف يقل تأثير الجزء الممهّد حتى يتلاشى ($I=0$) وعندها لا توجد منطقة جزء ونعود الى مجموع مربعات البواقي لينتج منحنى استكمال البيانات .

وعندما $I \in \mathbb{R}_\gamma$ فان $[g(m(x))]^2 \in \mathbb{R}^0$ ويكون المقدر عبارة عن خط مستقيم وهو تقدير المربعات الصغرى ($I=\gamma$) وبكتابة المعيار السابق بصيغة المصفوفات

$$PRSS = (Y - g)^T (Y - g) + Ig^T kg \quad (31)$$

وبايجاد المشتقة ومساواة الناتج للصفر فان المقدر الناتج يكون

$$\hat{g} = (I - \Gamma K)^{-1} Y \quad (32)$$

$$\hat{g} = A_1 Y$$

الصيغة 32 تسمى صيغة (Reinsch) والمصفوفة A_1 تسمى (Hat Matrix) [7]

10-4 انحدار الشرائح الجزائية Penalty Spline Regression

قدم هذا الانموذج الباحثان (Eilers & Marks) عام 1986 وسميا (B- Spline) ، والتي تعني التمهيد باستخدام الشرائح (B) مع عقد بمسافات متساوية ولكن عددها على الاغلب اقل من عدد البيانات وللسيطرة على مشكلة تحت التمهيد (Under Smooth) استخدم الباحثان جزء متقطع (Discrete Penalty) من فروقات من رتب عليا لمعلمات الشرائح (B) . تحتوي شرائح الجزء على عدد من الخصائص المقيدة هذه الخصائص موروثه جزئياً من شرائح الاساس (B- Spline) ، اذ نلاحظ ان شرائح الجزء لا تطهر اي تأثيرات حدودية كما تفعل العديد من دوال التمهيد اللبية (Kernel Smoothing) وهذا يعني ان شكل انتشار منحنى الملائمة او دالة الكثافة للبيانات يكون مصحوباً عموماً بالانحناء نحو الصفر .

ويمكن لشرائح الجزء ملائمة بيانات متعدد الحدود ، ولنفرض ان (x_i, y_i) معطى ، فانه سوف تتلائم تماماً مع البيانات .



فاذا كانت (y_i) متعدد الحدود في x من الدرجة k فان شرائح الاساس (B- Spline) من الدرجة k او اعلاه سوف تتلائم تماماً مع البيانات .
وبتطبيق الامر نفسه على شرائح الجزاء ، اذا كانت رتبة الجزاء من الدرجة $(k+I)$ او اعلاه مهما كانت قيمة معلمة الضبط λ ، اذ يمكن ان تحافظ شرائح الجزاء من الدرجة $(k+I)$ ودوال الجزاء من الرتبة $(k+I)$ او اعلاه فان :

$$\sum_{i=1}^m X^k Y_i = \sum_{i=1}^m X^k \hat{Y}_i \quad (33)$$

هذه الخاصية مفيدة بشكل خاص في سياق تمهيد دوال الكثافة (pdf) حيث ستكون مقدرات دالة الكثافة الاحتمالية (المتوسط والتباين) مساوية لمتوسط وتباين البيانات لاي مقدار من التسوية .
وهذه الميزة بالمقارنة مع تمهيد الدوال اللبية تضخم التباين بشكل متزايد وتعطي تمهيد اقوى . ان حدود شرائح الجزاء التي تتلائم مع البيانات مع تمهيد قوي هي متعدد الحدود اما بالنسبة للقيم الكبيرة من معلمة الجزاء (λ) عند جزاء من الرتبة k فان سلسلة الملائمة ستقترب من متعدد الحدود من الدرجة $(k-1)$ اذا كانت درجة الشرائح الاساس (B- Spline) تساوي او اكبر من k وبالتالي فان مقدر انحدار شرائح الجزاء (\hat{B}) يمكن التعبير عنه بالصيغة التالية [1,7]:

$$\hat{B} = \arg \min \sum_{i=1}^n (Y_i - X(x_i)'B)^2 + \lambda \sum_{i=1}^m B_{p+j}^2 \quad (34)$$

لكل معلمة جزاء $\lambda > 0$ ، ولنرمز الى مصفوفة تصميم الشريحة بالرمز F حيث ان $F = \{X(x_1)' \dots X(x_n)'\}'$ وان قيمة متغير الاستجابة $Y = (Y_1, \dots, Y_n)'$ ، اما مصفوفة معاملات الشرائح فيرمز لها بالرمز D حيث ان $D = \text{diag}(o_{p+1}, I_k)$ والتي تشير الى جزاء معاملات الشرائح فقط ، وعليه فان (\hat{B}) يمكن التعبير عنه بالصيغة [31]:

$$\hat{B} = (F'F + \lambda D)^{-1} F'Y$$

11-4 اختيار المعلمة التمهيدية

لاختيار المعلمة التمهيدية لمقدر الشرائح نستخدم معيار العبور الشرعي المعمم (Generalized Cross Validation) ومختصره (GCV) والذي يكتب وفق الصيغة التالية [37]:

$$GCV = n^{-1} \sum_{i=1}^n (Y_i - \hat{g}(X_i; \Gamma)) / (1 - n^{-1} \text{tr}(A_1))^2 \quad (36)$$

حيث ان المقدار $\text{tr}(A_1)$ يساوي مجموع عناصر القطر الرئيسي للمصفوفة (Hat) ويسمى هذا المجموع (Effective Degree of Freedom) ، حيث ان : $df = \text{trace}(A_1)$.
ولايجاد المعلمة التمهيدية نختار من بين مجموعة كبيرة من القيم المرشحة من قيم Γ التي تجعل معيار (GCV) اقل ما يمكن .

1- تحويل فوريير Fourier Transform (FT)

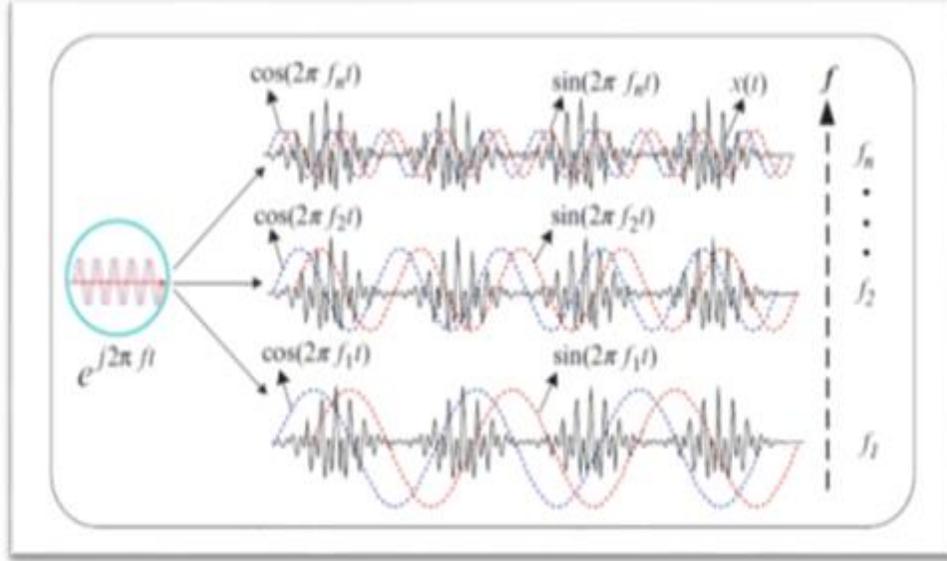
من اكثر التحويلات الرياضية انتشاراً ولا سيما في مجال معالجة الاشارة ويعبر عنه رياضياً
بانه تحويل شكل الاشارة من المجال الزمني الى المجال الترددي [34]



الشكل رقم (1) يمثل تحويل فوارير

قدمه العالم الفرنسي جوزيف فوارير عام 1922 وهو طريقة من اجل تمثيل الاشارات الدورية باستخدام
سلسلة من دوال الجيب والجيب تمام (الدوال الاسية والعقدية) وبترددات مختلفة .
ثم تم تطويرها من اجل اشارة حتى لو لم تكن دورية بانها دورها الى اللانهاية ، عندها ينتج ما يعرف
بتحويل فوارير الذي يعرف رياضياً للاشارة $X(t)$ [17,38] :

$$X(f) = \int_{-\infty}^{\infty} X(t)e^{-j\omega t} dt \quad , \quad \omega = 2\pi f \quad (37)$$



الشكل (2) يمثل الإشارة كمجموعة سلسلة توابم sine وبترددات مختلفة
 اما تحويل فوريير العكسي $X(f)$ للإشارة فيعبر عنه رياضياً بأنه تحويل شكل الإشارة من المجال
 الترددي الى المجال الزمني ويعطى بالعلاقة التالية :

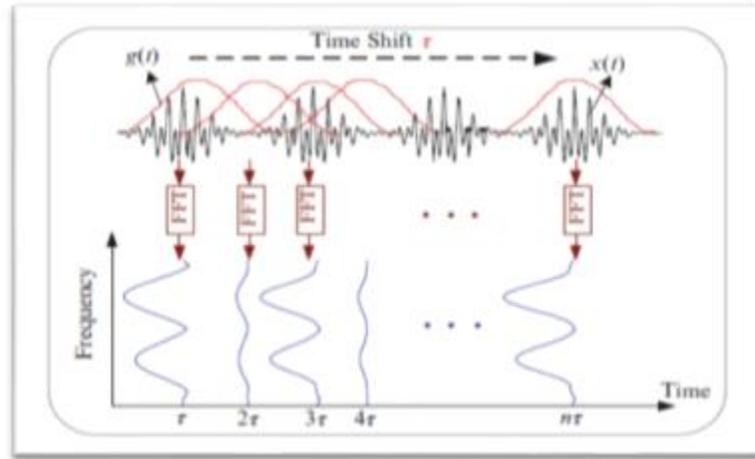
$$X(f) = \int_{-\infty}^{\infty} X(t)e^{j\omega t} dt \quad , \quad \omega = 2\pi f \quad (38)$$

يعطي تحويل فوريير معلومات عن التمثيل الطيفي للإشارة لكن لا يزود بمعلومات عن ازمنة تلك
 المركبات الترددية المكونة للإشارة ، ويعبر عن هذه المعلومات في المجال الزمني غير مهمة في حال
 كون الإشارة ثابتة .

لكن المشكلة هي ان تحويل فارير يصبح غير فعال بالنسبة للإشارات غير الثابتة لانه لا يعطي
 المعلومات عن المحتوى الترددي خلال الزمن . ومعظم الاشارات الطبية والحيوية والاقتصادية هي
 اشارات غير ثابتة لذلك عمل الباحثون على ايجاد حل لتلك المشكلة من خلال تطوير ما يعرف بتحويل
 فوريير القصير زمنياً (Short – Time Fourier Transform) ومختصره (STFT) [20,34] .

6- تحويل فوريير القصير زمنياً (STFT) Short – Time Fourier Transform
 اوجد الباحث (Dennis Gabor) تحويلاً يعد تطوراً لتحويل فوريير وبمثابة حلاً للمشكلة السابقة
 ونقطة الضعف التي كان يعانيتها التحويل السابق اذ انه يدعى بتحويل فوريير القصير زمنياً

(STFT) اذ يعمل هذا التحويل بتقنية تدعى (Windowing The Signal) حيث يتم تحليل جزء صغير من الزمن من الإشارة غير الثابتة وتعتبر الإشارة ثابتة ضمن النافذة وكما موضح في الشكل رقم (3) [18].



شكل رقم (3) تحويل فورييه القصير زمنياً

اذ يقوم هذا التحويل بتمثيل الإشارة زمنياً وترددياً على حساب دقتها الزمنية والترددية وهذا يتعلق بحجم النافذة المستخدمة والشكل الرياضي. لهذا التحويل يعطى بالعلاقة التالية :

$$STFT_x^w(t', f) = \int_{t'} (X(t) \cdot w^*(t - t') \cdot e^{-j2\pi ft}) dt \quad (39)$$

حيث ان $X(t)$ هي الإشارة قيد الدراسة ، $w(t)$ هي دالة النافذة التي تم ازاحتها بمقدار t' ، اذ يتم تحريك النافذة على طول الإشارة مع اجراء التحويل بضرب الإشارة بدالة النافذة في عدة نقاط زمنياً ، والدالة الاسية $e^{-j2\pi ft}$ تقوم بنقل الإشارة من المجال الزمني الى المجال الترددي .

لكن تكمن المشكلة في هذا التحويل في الضيعة الزمنية والترددية اذ انه عند استخدام نافذة قصيرة يتم الحصول على دقة زمنية جيدة لكن بدقة ترددية قليلة ، ويحدث العكس عند استخدام نافذة كبيرة وايضا عند النافذة الثابتة بالنسبة للزمن والتردد لذلك تم تطوير ما يعرف بالتحويل المويجي [20,26].

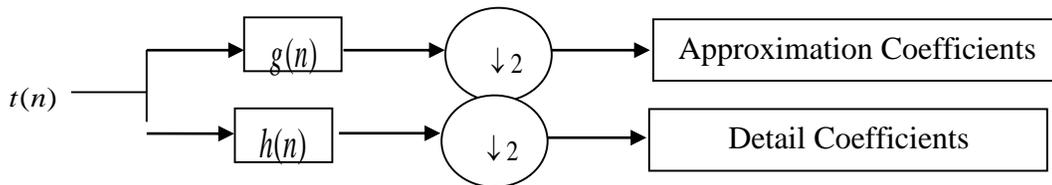
7-التحويل المويجي Wavelet Transformation

ظهرت التحويلات المويجية كبديل للتحويلات السابقة مثل تحويل فوارير وتحويل فورير القصير والتي اقتصر في تمثيلها للإشارة أما على مجال الزمن واهمال التردد أو العكس من ذلك ،

كذلك افترقت الى التغير في عرض النافذة واكتفت بنافذة ثابتة ، لذلك ظهرت هذه التحويلات والتي تميزت بتمثيل الاشارة في مجالي الزمن – تردد وعرض حزمة متغير مع طبيعة انتشار البيانات . وعليه فقد أخذت التحويلات المويجية حيزاً واسعاً من اهتمام الكثير من الباحثين وذلك لملائمتها لكثير من الظواهر ولمختلف المجالات منها معالجة الاشارة والصور وتحليل بيانات بصمات الأصابع الرقمية وتحسين الصورة وتحليل اشارات الزلازل وغيرها من التطبيقات التي يصعب حصرها . يوجد نوعان للتحويل المويجي الأول هو تحويل المويجات المستمر (Continuous Wavelet Transformation) والثاني هو تحويل المويجات المتقطع (Discrete Wavelet Transformation) والتي سيتم عرضه في هذا البحث . ان الفرق الأساسي بين التحويل المويجي المستمر والمتقطع هو أنه يمكن اختيار مجموعة فرعية من التدرجات والانتقالات اللازمة في عملية المعالجة دون الحاجة الى اجراء التحويل لكافة التدرجات الانتقالات من خلال اجراء مقاطعات زمنية في الاشارة ، يتم الحصول على كم كافي من المعلومات مع زمن حساب قليل وعدم خسارة المعلومات المهمة في الاشارة [5,27].

1-7 التحويل المويجي المتقطع Discrete Wavelet Transformation

ان ما يميز التحويل المويجي المتقطع عن بقية التحويلات كونه يستعمل ما يسمى بالمرشحات أو بنوك المرشحات (Filter Banks) والتي تم تطويرها من قبل (Mallat) عام (1988) ، حيث يتم تقسيم الاشارة $t(n)$ الى مرشح ممر منخفض $g(n)$ يتم من خلاله الحصول على المعاملات التقريبية (approximation coefficients) والتي تمثل مكونات الاشارة ذات التردد المنخفض وتدعى عادةً بمعاملات القياس (scale coefficients) وعلى مرشح ممر عالي $h(n)$ للحصول على المعاملات التفصيلية (detail coefficients) والتي تمثل مكونات الاشارة ذات التردد العالي والتدرج المنخفض والتي عادةً ما تسمى بمعاملات المويجة (wavelet coefficients) . ومن خصائص هذا التقسيم أنه يسمح بتحليل الاشارة ومن حزم ترددية مختلفة مع دقة مختلفة [4,8] .



الشكل رقم (4) يوضح تقسيم الاشارة في التحويل المويجي المتقطع

ويعبر عن هذا التقسيم رياضياً بالعلاقات التالية :

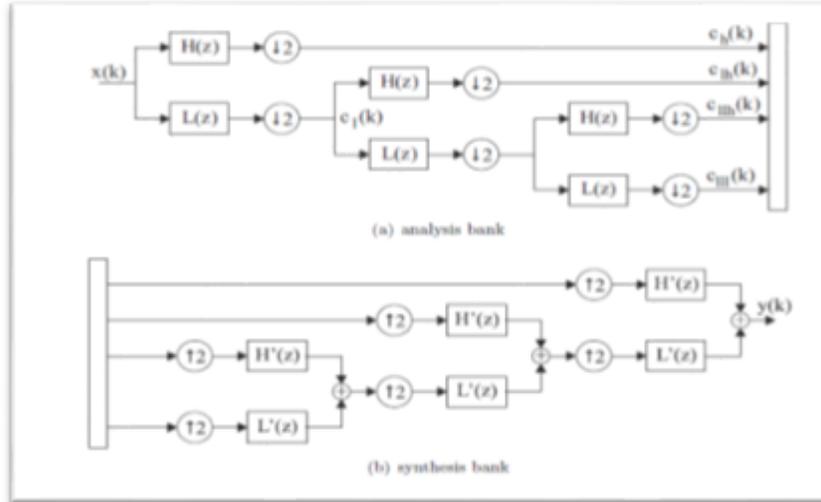
$$S_{j-1,k} = \sum_{n \in \mathbb{Z}} h_n - 2k c_{j,n} \quad (40)$$

$$d_{j-1,k} = \sum_{n \in \mathbb{Z}} g_n - 2k c_{j,n} \quad (41)$$

ان المعادلة رقم (40) تمثل معامل التعريف أو القياس عند مستوى $j-1$. والمعادلة رقم (41) تمثل معامل التفصيل عند مستوى $j-1$

الإشارة المتقطعة $t(n)$ تدخل الى بنك التحليل ويتم ترشيحها باستعمال المرشحات $l(z)$ و $h(z)$ التي تجزأ المحتوى الترددي للإشارة المدخلة الى حزم ترددية متساوية ، اذ يحتوي كل مرشح على نصف المحتوى الترددي المدخل . ولكن نفس عدد العينات ، وبالتالي مخرجات المرشحين معاً تحتوي على المحتوى الترددي المدخل نفسه ولكن عدد العينات سيتضاعف في المخرجات عما عليه في الادخال لذلك يتم اللجوء الى ما يسمى بتخفيض عدد العينات (down sampling) بمقدار 2 ويشار الى هذه العملية بالرمز $(\downarrow 2)$ ، أي أنه يتم تخفيض عدد العينات الى النصف ، مثلاً اذا كانت الإشارة الأصلية تحوي (256) مشاهدة من المعلومات سيكون لكل من المعاملات التقريبية (256) والتفصيلية كذلك أي يصبح العدد (512) ولمعالجة هذه المشكلة يتم استخدام (down sampling) أي مرشح نصف حزمة ويمكن العودة والحصول على الإشارة الأصلية من المعاملات التقريبية والتفصيلية والقيام بعملية زيادة العينات الى الضعف وهذا ما يدعى بالـ (up sampling) .

ويتم الحصول على الإشارة الأصلية بالتجميع المتسلسل لكل المعاملات الناتجة سابقاً (المعاملات التفصيلية والمعاملات التقريبية) بدءاً من آخر عملية التحليل ، وتستمر هذه العملية بنفس عدد مراحل التحليل حتى يتم الحصول على الإشارة الأصلية [9,23].



الشكل رقم (5) يوضح ترشيح بثلاث مراحل مرشحات التحليل في الأعلى ومرشحات التركيب في الأسفل

2-7 التحويل المويجي المستمر (CWT) Continues Wavelet Transformation

يعرف هذا التحويل كتكامل على الزمن للإشارة مضروباً بنسخ مختلفة من القياس (s) والانتقال (τ) لدالة الموجة الام ، ويعبر عنه رياضياً بالعلاقة التالية :

$$CWT(s, \tau) = \int_{-\infty}^{\infty} x(t) \psi_{(s, \tau)}^*(t) dt = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t-\tau}{s}\right) dt \quad (42)$$

حيث ان $x(t)$ تمثل الإشارة التي يجب تحليلها ، $\psi(t)$ تمثل الموجة الام ، (*) تشير الى معامل المرافق العقدي .

ويعطى التحويل المويجي المستمر العكسي (ICWT) بالعلاقة التالية :

$$x(t) = \frac{1}{c} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{|s|^2} . CWT(s, \tau) \frac{1}{\sqrt{|s|}} \psi\left(\frac{t-\tau}{s}\right) ds . dt \quad (43)$$

حيث ان $c = \int_{-\infty}^{\infty} \frac{|\psi(\omega)|^2}{|\omega|} . d\omega$ ، تشير الى تحويل فوارير للدالة $\psi(t)$.

يمكن اجراء التحليل بطريقة مشابهة لتحويل فوارير القصير زمنياً ، اذ يتم ازاحة النافذة على طول الإشارة من اجل عرض نافذة معين ثم تكرار العملية لمختلف الترددات .

لكن المشكلة تكمن في العدد الهائل من الموجات الناتجة بسبب حساب المعاملات المويجية عند مختلف الترددات والازاحات الممكنة ، وبالتالي ينتج كمية كبيرة من المعلومات ، وهذا ما يجعل عملية المعالجة تتطلب زمناً كبيراً جداً . ولحل هذه المشكلة تم ايجاد التحويل المويجي المنقطع [25,41] .

8- الإندجار المويجي Wavelet Regression

ربما يكون الإندجار هو مجال الاحصاء الذي حظى بأكبر قدر من الاهتمام من الباحثين في طرق المويجات ، و غالباً ما يتم استعمال طرائق المويجة كشكل من أشكال الإندجار اللامعلمي ، وتتخذ التقنيات العديد من الأسماء مثل تقليص المويجة أو تقدير منحنى الإندجار اللامعلمي أو اندجار المويجات ، وبشكل عام فان الإندجار اللامعلمي نفسه يشكل مجالاً هاماً وحيوياً للاحصاء الحديث والفكرة العامة لعمل الإندجار المويجي موضحة وفق التالي :

ليكن لدينا المشاهدات $y_i = (y_1, \dots, y_n)$ معطاة بالصيغة التالية :

$$y_i = g(t_i) + \varepsilon_i \quad (44)$$

و أن $t_i = i/n$ ، والهدف هو تقدير دالة $g(t_i)$ غير المعلومة $t_i \in [0,1]$ باستخدام مشاهدات مشوشة y_i .



تم تقديم مفهوم تقليص المويجة (wavelet shrinkage) أو الانحدار المويجي للأدبيات الاحصائية من قبل الباحث (Donoho) (1995) ، وتلخصت الفكرة العامة لتطبيق التحويل المويجي المتقطع على الانموذج أعلاه رقم (41) ، اذ يتم استعمال خوارزمية (Mallat) .
ليكن y_i يمثل المشاهدات ، و g يمثل الدالة غير المقدره و e_i يمثل الخطأ أو الضوضاء ، ومن خلال التحويل المويجي المتقطع يمكن كتابة انموذج التحويل وفق الآتي :

$$d^* = d + \varepsilon \quad (45)$$

و أن $d^* = W_y$ ، $d = W_g$ ، $\varepsilon = W_e$ ، W هو مصفوفة التحويل المويجي .
ثلاث خصائص أساسية لنجاح تقليص المويجة :

- 1- تتكيف التحويلات المويجية مع العديد من الدوال (دوال التمهيد المتقطعة والغير المتجانسة) .
- 2- علاوة على ذلك نظراً لعلاقة بارسفال فان الطاقة في مجال الدالة $\sum g(t)^2$ تساوي مجموع مربعات معاملات المويجات $\sum_{j,k} d_{j,k}^2$ ومع ذلك اذا تم الاخذ بنظر الاعتبار التباين فهذا يعني أن طاقة الإشارة الأصلية g تتركز في معاملات أقل ولا يتم فقد أي شيء ، وبالتالي بالنسبة الى تباين الضوضاء لن يكون المتجه d متناثراً فقط ولكن القيم نفسها غالباً ما تكون أكبر .
- 3- من خلال ان W الناتجة من التحويل المتقطع هي مصفوفة متعامدة هذا معناه أن مويجة التحويل للضوضاء والتي هي ضوضاء بيضاء تبقى بعد التحويل ضوضاء بيضاء وأنها تنتشر بالتساوي على جميع معاملات المويجة .
بناءً على الخصائص أعلاه اقترح كلاً من (Donoho & Johnstone) (1994) تعريب (تقليص) المويجة التالية والتي نحتاجها في تقدير الدالة $g(t_i)$.
كانت الفكرة الأساسية أن القيم الكبيرة لمعاملات المويجة d^* ، كانت على الأرجح حالات تتكون من إشارة وضوضاء حقيقية في حين أن المعاملات الصغيرة كانت بسبب الضوضاء فقط ، ومن ثم لتقدير d بنجاح وجدت فكرة العتبة لتقدير \hat{d} عن طريق ازالة المعاملات في d^* التي تكون أصغر من بعض عتبة وبشكل أساسي تحافظ على المعاملات الأكبر [30] .

1-8 قواعد العتبة Thresholding Rules

تعد مرحلة ازالة التشويش الموجودة في الإشارة من أهم خطوات تقدير دالة الانحدار باستخدام تقليص المويجة ، حيث أن اختيار العتبة يساهم في ازالة التشويش وبدوره يحافظ على معاملات الإشارة الأصلية كون معاملات التشويش تكون ذات تردد أقل من تردد معاملات الإشارة الأصلية .
تم اقتراحها من قبل (Donho) (1995) للتخلص من التشويش بمساعدة التحويل المويجي عن طريق ما أسماه عملية التعريب وأطلق عليها اسم التقليص (Shrinkage) وقد قدمها كأداة قوية لازالة التشويش من الاشارات الملوثة للتشويش (الضوضاء) .
يمكن أن تجرى عملية التعريب بعدة طرق أهمها وأكثرها شيوعاً هي طريقة التعريب الناعم (Soft Thresholding) والتعريب الصلب (Hard Thresholding) [30,40] .

2-8 طريقة التعتيب الخشن Hard Thresholding

هي طريقة مبسطة تتم من خلال تصفير للعناصر التي قيمتها المطلقة أقل من العتبة ويعبر عنها رياضياً [35]:

$$Thr_{\lambda}^H(y, \lambda) = \begin{cases} 0 & \text{If } |y| \leq \lambda \\ y & \text{If } |y| > \lambda \end{cases} \quad (46)$$

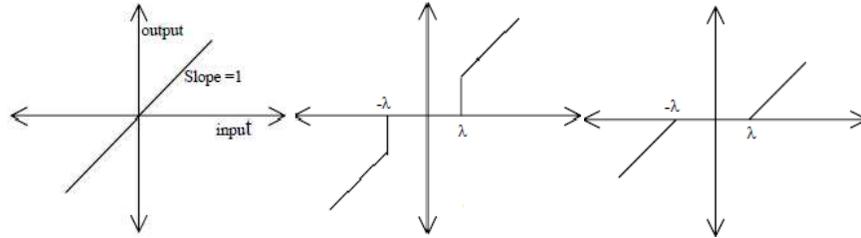
إذ أن λ هي قيمة العتبة (Thresholding value).

3-8 طريقة التعتيب الناعم Soft Thresholding

هي امتداد للطريقة السابقة وتختلف عنها أنه بعد أن يجري تصفير العناصر التي قيمتها المطلقة أقل من العتبة تراح العناصر غير الصفرية باتجاه الصفر ويعبر عنها رياضياً بالعلاقة التالية [6]:

$$Thr_{\lambda}^S(y, \lambda) = \begin{cases} 0 & \text{If } |y| \leq \lambda \\ \text{sgn}(y) (|y| - \lambda) & \text{If } |y| > \lambda \end{cases} \quad (47)$$

إذ أن λ هي قيمة العتبة (Thresholding value).



شكل (6) يمثل دوال العتبة الخطية (a) والصلب (b) والمرن (c)

4-8 قيمة العتبة Threshold Value

قيمة العتبة λ هي معلمة مهمة للغاية في خوارزمية تقليل المويجة لتقليل الضوضاء التي تعانيه الإشارة ، إذ أن هذه الضوضاء سوف تتأثر بشكل مباشر باختيار قيمة العتبة المناسب ، هنالك العديد من أنواع قيم العتبة [12].

5-8 طريقة العتبة الشاملة Universal Thresholding

قدمت طريقة العتبة الشاملة من قبل كلاً من (Donoho and Jonstone) والمعطاة وفق الصيغة التالية [3,24]:

$$\lambda_{universal} = \sigma \sqrt{2 \log(n)} \quad (48)$$

اذ ان n : طول الاشارة ، σ : الانحراف المعياري لمستوى التشويش .
ان اختيار العتبة لتكون $\sqrt{2 \log}$ من شأنه أن يزيد الضوضاء ذات الاحتمالية العالية .

6-8 طريقة العتبة Sure Thresholding

طور (Donoho and Jonstone) طريقة مهمة من طرائق اختيار قيمة العتبة أطلقوا عليها اسم (Sure Shrink) تستند هذه الطريقة على تقنية (Stein) (1981) غير المتحيزة لتقدير المخاطر (SURE) وهي اختصاراً لكلمة (Stein Unbiased Risk Estimation) لكل مستوى موجة z . ونظراً لكون التحويل المويجي متعامد وبذلك فان التحويل للضوضاء هو الآخر يكون متعامد ، أي أن المعاملات $d_{j,k}$ هو أيضاً متعامد ، وبما أن الضوضاء تتوزع توزيع (Gaussian) مما يعني أن d^* يتوزع (Gaussian) ، وبناءً على ذلك وضح (Stein) ان ذلك التقدير غير متحيز للمخاطرة .

$$SURE(\lambda_j, d_{jk}) = N - 2 \sum_{k=1}^N I(|d_{jk}| \leq \lambda_j) + \sum_{k=1}^N \min(|d_{jk}|, \lambda)^2 \quad (49)$$

وعليه فان قيمة العتبة للـ (SURE) يمكن إيجادها من الصيغة الآتية [29,30]

$$\lambda_{j,SURE} = \arg \min_{0 \leq \lambda \leq \sqrt{2 \log N}} SURE(\lambda_j, d_{jk}) \quad (50)$$

9- المناقشة

يمكن أن يكون تأثير الاسلوب المناسب على نمذجة الانحدار لمجموعات البيانات كبيراً ، كما كشفت عنه دراسات المحاكاة [3]. تشتمل طرق التقدير الموضحة في هذا العمل على الإجراءات التالية (WT ، FTST ، FT ، spline ، kernel) ، تم تلخيص إيجابيات وسلبيات كل تقنية في الجدول (3) .

جدول رقم (3) يبين ايجابيات وسلبيات الاساليب المستعملة

التسلسل	الاسلوب	الايجابيات	السلبيات
1	اللبية	مرونة التعامل مع البيانات دون اشتراط معرفة التوزيع او حجم البيانات	عرض النافذة يكون ثابتاً او محدد مما يؤدي الى خسارة بعض المعلومات التي قد تكون مهمة
2	الشرائحية	مرونة التعامل مع البيانات دون اشتراط معرفة التوزيع او حجم البيانات ودقة افضل في التقدير	حجم و عدد العقد يكون ثابتاً او محدد مما يؤدي الى خسارة بعض المعلومات التي قد تكون مهمة
3	تحويلات فوارير	سهولة الحساب مرونة التعامل مع انواع البيانات	عند تحليل الإشارة فانه اما يركز على التردد وياهمال الزمن او العكس مما يؤدي الى خسارة جزء مهم من البيانات
4	تحويلات فوارير القصير زمنياً	سهولة الحساب مرونة التعامل مع انواع البيانات بالاضافة الى حل مشكلة زمن تردد	لا يمكن التعامل مع الاشارات المتغيرة اي الغير مستقرة زمنياً
5	التحويلات الموجية	مرونة التعامل مع البيانات دون اشتراط معرفة التوزيع او حجم البيانات، تعامله بشكل كفوء مع الإشارة المدخلة من خلال زمن تردد دون اهمال اي جزء منهم ، سهولة حسابه من خلال خوارزميات معدة	يشترط شكل معين للبيانات ان تكون من النوع الثنائي وكذلك تساوي المسافات زمنياً بين المشاهدات بالإضافة الي تأثره بمشاكل الارتباط الذاتي للأخطاء او عدم توزيع الخطأ توزيعاً غاوسياً

10- معيار المقارنة comparison standard

وهي مقاييس للدقة تحسب من خلال تربيع الخطأ لكل مشاهدة في مجموعة بيانات ومن ثم ايجاد المعدل او المتوسط لمجموع متوسط مربعات الخطأ ، تكون صيغة متوسط مربعات الخطأ والتي سيتم استعمالها في الجانب العملي بالصيغة التالية :

$$MSE(\hat{g}) = \frac{1}{n} \sum_{i=1}^n [g(i/n) - \hat{g}(i/n)]^2 \quad (51)$$

اما في الجانب التجريبي فسيتم استعمال معدل متوسط مربعات الخطأ (AMSE) والتي تكون بالصيغة التالية :

$$AMSE(\hat{g}) = \frac{1}{n} E[g(i/n) - \hat{g}(i/n)]^2 \quad (52)$$

11- تقنيات إضافية لتحسين دقة النهج Additional techniques to improve approach accuracy

هنالك تقنيات عديدة لم يسعنا تناولها في هذه الدراسة جاءت لتعالج القصور الذي عانت منه الاساليب المطروحة ومن هذه الاساليب هو استعمال عرض نافذة متغير بدلاً من استعمال عرض نافذة محدد في التقديرات اللبية وكذلك تم اقتراح تحويلات موجية تتعامل مع خرق الشروط اللازمة لتطبيقاً كما في خوارزمية كوفاك -سلفر مان الذي اقترح خوارزمية لمعالجة ذلك

القصور بالإضافة الى اقتراح دوال عتبة وقيم عتبة من شأنها معالجة مشكلة عدم التوزيع الغاوسي للأخطاء التي يصعب حصرها في الدراسة الحالية .

12-الاستنتاجات Conclusions

إن تقدير دالة الانحدار اللامعلمي أمر بالغ الأهمية لتحسين الكفاءة والجودة في التحليل الإحصائي. كما أن اختيار الأسلوب الأفضل أمر ضروري لمنع التنبؤات المتحيزة والنتائج غير الصحيحة. وقد تم تطوير طرق مختلفة، بما في ذلك التقديرات اللبية والشرائحية وتحويلات فوارير للزمن القصير والتحويلات الموجية لكل من هذه الطرائق مزاياه وعيوبه فلا يوجد افضلية مطلقة على بقية الأساليب فقد يكون الأسلوب بسيطاً الا انه يكون ذات تقدير اقل كفاءة او قد يكون تقديراً كفوءاً الا انه يتطلب شرطاً معيناً بخلافها يكون التقدير غير كفوء ومع ذلك فقد ظهرت التحويلات الموجية كحل لتلك المشاكل وتمثلت بكفاءة عالية في التقدير والتعامل مع اغلب مشاكل الانحدار بمرونة عالية ولذلك اخذت حيزاً واسعاً في اغلب مجالات الحياة العملية مثل الطب والفلك والاقتصاد والعلوم بما فيها علم الجريمة والتعرف على البصمات والصورة المشوشة والاصوات بل تعدى ذلك لدراسة الزلازل والبراكين .

13-المصادر References

1. حسين رشيد, ظافر & يوسف خمور. (2005). مقارنة الطرائق الشرائحية لتقدير منحني الانحدار اللامعلمي. المجلة العراقية للعلوم الاحصائية. 62-40, (1) 5 ,
2. حسين رشيد, ظافر & يوسف خمور. (2005). مقارنة الطرائق الشرائحية لتقدير منحني الانحدار اللامعلمي. المجلة العراقية للعلوم الاحصائية. 62-40, (1) 5 ,
3. حمزة ، سعد كاظم ، (2015) ”استخدام قيم عتبة مختلفة في مقارنة بعض طرائق التقدير الموجي لدالة الانحدار اللامعلمي بوجود بيانات مفقودة“ ، مجلة كلية الرافدين الجامعة للعلوم.
4. حمزة ، سعد كاظم ، (2015) ”التحليل الموجي لانموذج الانحدار في ظل بيانات مفقودة“ ، اطروحة دكتوراه ، كلية الادارة والاقتصاد – جامعة بغداد .
5. حمزة ، سعد كاظم ، (2024) ” Comparison of some Wavelet Transformations to Estimate Nonparametric Regression Function “ Journal of Economics and Administrative Sciences , جامعة بغداد / كلية الادارة والاقتصاد.
6. ظافر حسين رشيد & مناف يوسف حمود. (2012). مقدر Nadaraya-Watson اسلوب تمهيدي لتقدير دالة الانحدار , Journal of Economics and Administrative Sciences, 18(65), 283-283.
7. Abdullah, S. A. (2023). Estimate the coefficients of penalty spline regression models using the (SOP) method. Journal of Administration and Economics, 48(141).
8. Abramovich, F., Bailey, T. C. and Sapatinas, T. (2000). Wavelet analysis and its statistical applications. The Statistician, 49 1–29.

9. Abramovich, F., Sapatinas, T. and Silverman, B. (1998).
10. Al-Sharout, M. H. & Hafez M. (2011) . A Comparison of Non-Parametric Smoothing Regression Models Using Simulation. Journal of Al-Qadisiyah for Computer Science and Mathematics, 3(2).
11. Altman, N., & Leger, C. (1995). Bandwidth selection for kernel distribution function estimation. Journal of Statistical Planning and Inference, 46(2), 195-214.
12. Antoniadis, A. , (2007), “ Wavelet Methods in Statistics : Some Resent Developments and Their Applications” , Statistics Surveys, VOL 16-55 .
13. Araveeporn, A. (2019). An estimating parameter of nonparametric regression model based on smoothing techniques. Statistical Journal of the IAOS, 35(2), 269-276.
14. Aydin, D. (2007). A comparison of the nonparametric regression models using smoothing spline and kernel regression. World Academy of Science, Engineering and Technology, 36, 253-257.
15. Aydına, D., Günerib, Ö. İ., & Fitc, A. (2016). Choice of bandwidth for nonparametric regression models using kernel smoothing: A simulation study. International Journal of Sciences: Basic and Applied Research (IJSBAR), 26(1), 47-61.
16. Basheer, Z., & Algamal, Z. (2020). Smoothing parameter selection in Nadaraya-Watson kernel nonparametric regression using nature-inspired algorithm optimization. Iraqi Journal of Statistical Sciences, 17(2), 43-50.
17. Beyene, J. (2001). Uses of the fast Fourier transform (FFT) in exact statistical inference. University of Toronto.
18. Burrus, C., Gopinath, R. and Guo, H. (1998). Introduction to wavelets and wavelet transforms: a primer. Prentice hall Upper Saddle River, New Jersey.
19. Carter, C. K., & Eagleson, G. K. (1992). A comparison of variance estimators in nonparametric regression. Journal of the Royal Statistical Society Series B: Statistical Methodology, 54(3), 773-780.
20. Chandrasekharan, K. (2012). Classical Fourier Transforms. Springer Science & Business Media.
21. Clark, R. M. (1977). Non-parametric estimation of a smooth regression function. Journal of the Royal Statistical Society: Series B (Methodological), 39(1), 107-113.



22. Dakhil, M. A., & Hussain, J. N. (2021, March). A Comparative Study of Nonparametric Kernel estimators with Gaussian Weight Function. In *Journal of Physics: Conference Series* (Vol. 1818, No. 1, p. 012058). IOP Publishing.
23. Daubechies, I. (1992). Ten lectures on wavelets (CBMS-NSF regional conference series in applied mathematics).
24. Donoho, D. and Johnstone, I. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association*, 1200-1224.
25. Donoho, L. D. , Johnstone, M. I. , (1994), “Ideal spatial adaptation by wavelet shrinkage”, *Biometrika*, 81, 3, pp. 425-55 .
26. Gray, R. M., & Goodman, J. W. (2012). *Fourier transforms: an introduction for engineers* (Vol. 322). Springer Science & Business Media.
27. Hamza, S. K., & Ali, S. (2022). Estimation of nonparametric regression function using shrinkage wavelet and different mother functions. *Periodicals of Engineering and Natural Sciences*, 10(6), 96-103.
28. Hassan, Y. A., & Hmood, M. Y. (2020). Estimation of return stock rate by using wavelet and kernel smoothers. *Periodicals of Engineering and Natural Sciences*, 8(2), 602-612.
29. He, H. (2018), “A Novel Adaptive Wavelet Thresholding with Identical Correlation Shrinkage Function for ECG Noise Removal” *Chinese Journal of Electronics* Vol 27, PP 502-513 .
30. Johnstone, I. (1999), “Wavelet Shrinkage for Correlated Data and Inverse Problems : A Daptivity Results” *Statistica Sinica* Vol 9 P 51-83 .
31. Li, N., & Xu, X. (2011). Comparison of nonparametric regression curves by spline smoothing. *Communications in Statistics-Theory and Methods*, 40(22), 3972-3987.
32. Loader, C. (2012). Smoothing: local regression techniques. *Handbook of computational statistics: concepts and methods*, 571-596.
33. Marsh, L. C., & Cormier, D. R. (2001). *Spline regression models* (No. 137). Sage.
34. Mateer, T. (2008). *Fast Fourier transform algorithms with applications* (Doctoral dissertation, Clemson University).
35. Nason, P. G. , (1995), “Choice of The Threshold Parameter in Wavelet Function Estimation”, Springer, VOL, 103 .



36. Schindler, A. (2012). Bandwidth selection in nonparametric kernel estimation (Doctoral dissertation, Göttingen, Georg-August Universität, Diss., 2011).
37. Srihera, R., & Stute, W. (2010). Nonparametric comparison of regression functions. *Journal of Multivariate Analysis*, 101(9), 2039-2059.
38. Strang, G. (1993). Wavelet transforms versus Fourier transforms. *Bulletin of the American Mathematical Society*, 28(2), 288-305.
39. Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 359-372.
40. Zeidan, O. O., & Hamza, S. K. (2022). THE COMPARISON OF DIFFERENT THRESHOLD RULES IN ESTIMATING THE WAVELET REGRESSION FUNCTION. *World Economics and Finance Bulletin*, 14, 109-120.
41. Zeidan, O. O., & Hamza, S. K. (2022). THE USE OF THE SHRINKAGE WAVELET TRANSFORMS IN ESTIMATING THE INTEREST RATE ON LONG-TERM LOANS USING DIFFERENT THRESHOLD RULES WITH APPLICATION. *World Bulletin of Management and Law*, 14, 22-28.

Abstract

Although nonparametric regression models are weaker in description than parametric regression models, they also require fewer restrictions or conditions than parametric models. This is precisely what has made nonparametric regression models a very desirable tool for researchers, as actual data do not always have ideal specifications, which has led to a revival in the way of thinking about nonparametric regression models, which has led to the development of many methods in recent decades and in many fields such as engineering, medicine, geography, and economics. Therefore, a technique is presented to replace these missing values with specific estimated values in order to properly deal with partial (missing) data. A variety of techniques have been used to estimate nonparametric regression models. Therefore, a comprehensive analysis of the different methods used to estimate nonparametric regression functions is presented. It identifies the different types of methods proposed and points out the most common types of regression analysis. The main objective of this review paper is to introduce



the reader to the many nonparametric methods used in regression analysis and to explain how these methods affect the final conclusions of the study.