Journal of Economics and Administrative Sciences Vol.26 (NO. 119) 2020, pp. 428-443

Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

تقدير الانموذج الخطي الجزئي بإستعمال المهدات المويجية واللبية

أ.د. مناف يوسف حمود
كليـة الادارة والاقتصاد / جامعة بغداد / قسم الاحصاء
munaf_yousif@yahoo.com

الباحث/ يونس عامر حسن
كلية الادارة والاقتصاد / جامعة بغداد / قسم الاحصاء
younis.amer91@gmail.com

Received :14/12/2019 Accepted :15/1/2020 Published :April / 2020

مستخلص البحث:

يهدف هذا البحث الى تقدير الانموذج الخطي الجزئي (Partial Linear Regression Model) بإستعمال طريقتين من طرائق التمهيد وهما طريقتي التمهيد المويجي (Wavelet Smoother) والتمهيد اللبي (Kernel Smoother). تم استعمال تجارب المحاكات لبيان افضل تلك الطرائق في تقدير هذا النوع من النماذج بإختلاف الحالات والدوال وحجوم العينات والتباينات المستعملة. تم الاعتماد على معيار معدل متوسط مربعات الخطأ (Mean Average Squares Error) كأحد معايير المقارنة بين الطرائق. واثبتت النتائج افضلية طريقة التمهيد المويجي ولجميع الحالات المستعملة.

المصطلحات الرئيسة للبحث/ إنموذج الانحدار الخطي الجزئي, التمهيد المويجي, التمهيد الخطي الموضعي.

* البحث مستل من رسالة ماجستير

1- **المقدمة** Introduction:

الانموذج الخطي الجزئي يمثل دالة استجابة ويجمع بين النماذج المعلميه والنماذج اللامعلميه. وغالباً ما يتم استعماله في الظروف التي قد لا يعمل فيها الانموذج اللامعلمي بشكل جيد او عندما يرغب الباحث بإستعمال انموذج معلمي ولكن عندما يكون فيه الشكل الدالي لمجموعة فرعية من المتغيرات التوضيحية غير معروفا أو لا يمكن افتراض أن توزيع الخطأ محدد مسبقًا[11]. اول من قام بتطبيق عملي على الانموذج الخطي الجزئي الباحثين[5] Engle, Granger, Rice and Weiss عام 1986 لدراسة العلاقة ما بين درجة الحرارة واستهلاك الكهرباء اذ وضحوا انه يمكن التعبير عن العلاقة بإنموذج خطي جزئي لأن هناك العديد من العوامل مثل متوسط الدخل وسعر البضائع وقدرة المستهلك على الشراء وبعض الأنشطة الاقتصادية الأخرى. بعض هذه العوامل ذات صلة ببعضها البعض وقد تؤثر على النتيجة المطلوبة لذا قدّموا انموذجًا خطيًا جزئيًا يعمل على تمكين وتبسيط عملية التحول الخطي للبيانات. كما قدم الباحث [28] Speckman عام 1988 طريقة المربعات الصغرى بإستعمال شريحة التمهيد لتقدير انموذج الانحدار الخطي الجزئي, كما استعمل الباحثان[30] Zeger and Diggle عام 1994 الانموذج الخطي الجزئي في تحديد دورة فترة تطور كميات خلايا CD4 في فيروس نقص المناعة البشرية. وفي عام 1997 قدم الباحثان[9] Hamilton and Truong بحثاً عن التمهيد الخطي الموضعي (LLS) لتقدير الانموذج الخطي الجزئي بإستعمال طريقة Speckman في عملية التقدير. وفي عام 1999 استعمل الباحثان[27] Schmalensee and Stoker الانموذج الخطي الجزئي لدراسة العلاقة بين استهلاك البنزين ومرونة الدخل على المدى الطويل في الولايات المتحدة. وفي مجال علم البيئة استعمل الباحث[21] Prada-Sanchez لعام 2000 الانموذج الخطي الجزئي للتنبؤ بتلوث ثاني أكسيد الكبريت. في عام 2004 اقترح الباحثان[2] (Xiao-Wen Changa & Leming Qub) تقدير الانموذج الخطي الجزئي باستعمال التمهيد المويجي. الهدف من الانموذج الخطي الجزئي (PLM) هو تقدير الجزء المعلمي β غير المعلوم والدالة اللامعلميه $g(t)$ و تقليل الافتراضات على واحد او اكثر من المتغيرات التوضيحية والبقاء على خطية العلاقة بين المتغيرات[10]. اذ ان متغير الاستجابة Y يعتمد على المتغير X بطريقة خطية وعلى المتغير T بطريقة لاخطية. يعرف الانموذج الخطي الجزئي (PLM) حسب الصيغة الاتية: [24][6]

$$y_i = X_i^T \beta + g(t_i) + \varepsilon_i \quad \dots (1)$$

ويمكن كتابته بصيغة المصفوفات كالاتي:

$$\underline{Y} = X^T \underline{\beta} + \underline{g} + \underline{\varepsilon} \quad \dots (2)$$

اذ تشير: [2]

$\underline{Y}$: متجه متغير الاستجابة

X: مصفوفة المتغيرات التوضيحية المعلمية ذات الابعاد $(n \times p)$ اذ ان $P \le n$

t_i: قيم متغير احادي مضافة مثل الوقت الذي حصلت فيه المشاهدة اذ ان: $t_i = \frac{i}{n}, t \in [0,1]$

$\underline{\beta}$: متجه معلمات لمصفوفة المتغيرات التوضيحية X لـ P من الابعاد وتكون غير معلومة

g : دالة غير معلومة

$\underline{\varepsilon}$: تمثل الاخطاء العشوائية وتتوزع توزيعاً طبيعياً $\varepsilon \sim N(0, \sigma^2)$

2- **التمهيد المويجي** Wavelet Smoother:

تحليل المويجات هو فرع من فروع علم الرياضيات, تم تطويره سريعًا منذ الثمانينيات من القرن الماضي وهو يمثل تعميماً لتحليل فورير المويجات هي أداة رياضية يمكن استعمالها لاستخراج المعلومات من العديد من أنواع البيانات المختلفة كما في تحليل الإشارات، ومعالجة الصور والتحليل العددي وما إلى ذلك. اقترح الباحث فوريير في عام 1807 تحليل يدعى باسم تحليل فوريير. يعتمد في عمله على أنه يمكن تمثيل الدوال كمجموعة من دوال الجيب والجيب تمام اذ يحول الدالة التي تعتمد على الوقت إلى دالة جديدة تعتمد على التردد او بالعكس[9]. بعدها اكتشف الباحث Haar عام 1909 تحويل Haar المويجي وهو أبسط عائلات المويجات. ومفهوم عائلة المويجات هو ان المويجات الأب تُعد نقطة البداية, من خلال زيادة حجم المويجات الأب وتحليلها ، نحصل على المويجة الأم ومن ثم البنات والأبناء والحفيدات .. إلخ. بعد مساهمة Haar في المويجات اوجد الباحث Paul Levy أن دالة أساس المقياس المختلفة - التي أنشأها Haar كانت أساسًا

أفضل من دوال أساس فوريير[25]. اذ يمكن تقطيع دالة Haar إلى فواصل زمنية مختلفة مثل الفاصل الزمني من 0 إلى 1 أو الفاصل الزمني من 0 إلى ½ و ½ إلى 1 ، و دوال فوريير لها فاصل زمني واحد فقط, لذلك يمكن أن تكون مويجات Haar أكثر دقة في تصميم الدالة. بعد ذلك اوجدت الباحثة (Daubechies) عام 1988 القواعد المتعامدة مصنوعة من مويجات ذات دعم مرصوص, ثم صمم الباحث (Mallat) عام 1989 خوارزمية تحويل المويجات السريعة بعدها قدم كل من الباحثان (Donoho and Johnstone) عام 1994 حد العتبة (Thresholding) التي تعمل على تقليل الضوضاء (ازالة التشويش) للاشارة او الصورة وغيرها هنالك العديد من الدراسات والبحوث الخاصة بالتحليل المويجي.[13]

تحليل المويجات هو تحليل الإشارة الموقعي الى تردد - زمن مع حجم نافذة ثابت وشكل قابل للتغيير. وهذا يعني أنه في التردد المنخفض ، فإنه يظهر دقة تردد أعلى ودقة زمنية أقل. في الجزء عالي التردد ، فإنه يظهر دقة وقت أعلى ودقة تردد أقل.

اساس المويجات تكون هنالك دالة محددة تسمى المويجة الام $\psi(t)$ تولد مجموعة من الدوال تسمى مويجات الابناء التي تشكل القاعدة الاساس. ان اي دالة تكون دالة المويجة الام اذا حققت الشروط انها تتحلل الى الصفر بحيث يكون متوسط القيمة يساوي صفر.[14]

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \qquad \dots(3)$$

وهذا يعني ان المويجة الام $\psi(t)$ تكون قيمها صغيرة جداً او صفر خارج فترة زمنية مغلقة وتسمح للموجة ايضاً ان تتحلل بسرعه في مجال التردد لذلك يتم تحليل المويجات في كل من الوقت والتردد. تتمثل الطريقة العامة في التحويل المويجي بإنشاء مركب تقريبي بإستعمال دالة القياس (مرشح تردد واطيء) ودوال مويجية (مرشح تردد عالي) ينتج منهما نافذة ذات ابعاد متغيرة للحصول على الدقة العالية في الزمن والتردد.[1] [13]

$$\hat{g}(\mathrm{a},\mathrm{b}) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} g(t)\ \psi(\frac{t-b}{a})dt \qquad \dots(4)$$

اذ تشير:

a: معامل التقييس (scaling coefficient)

b: معامل التحليل (Translation coefficient)

t: الزمن (Time)

سيتم استعمال تحويل Daubechies المويجي المتقطع اذ يمتلك عزوم تلاشي (vanishing moments) من 1 الى *n* (تحويل هار المتقطع يمتلك عزم تلاشي واحد) لكل من دالة القياس ϕ (scaling function) و الدوال المويجية ψ (Wavelet functions).[3]

يتم ايجاد قيم المرشحات المويجية (wavelet filters) والتي يرمز لها بالرمز $(\tilde{g}_k)$ ذات التردد العالي من خلال مرشحات القياس (scaling filters) والتي يرمز لها بالرمز (h_{L-k}) ذات التردد الواطئ.[25]

$$\tilde{g}_k = (-1)^k h_{L-k} \quad , \quad k = L+1 \qquad \dots(5)$$

اذ تحتوي مصفوفة تحويل Daubechies المتقطع على n من مركبات الترشيح ولا يتم قطعها عند ادخالها في المصفوفة للحصول على مصفوفة متعامدة يمكن تمثيلها بهذا الشكل:[23]

$$\underset{N\times N}{W_N} = \begin{bmatrix} H_{N/2} \\ G_{N/2} \end{bmatrix} = \left[\begin{array}{cccccccccc} h_0 & h_1 & h_2 & h_3 & \cdots & h_{L-1} & h_L & \cdots & 0 & 0 \\ 0 & 0 & h_0 & h_1 & \cdots & h_{L-3} & h_{L-2} & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & h_{L-5} & h_{L-4} & \cdots & 0 & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & h_{L-1} & h_L \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ h_2 & h_3 & h_4 & h_5 & \cdots & 0 & 0 & \cdots & h_0 & h_1 \\ \hline g_0 & g_1 & g_2 & g_3 & \cdots & g_{L-1} & g_L & \cdots & 0 & 0 \\ 0 & 0 & g_0 & g_1 & \cdots & g_{L-3} & g_{L-2} & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & g_{L-5} & g_{L-4} & \cdots & 0 & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & g_{L-1} & g_L \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ g_2 & g_3 & g_4 & g_5 & \cdots & 0 & 0 & \cdots & g_0 & g_1 \end{array}\right]$$

يواجه الباحثين مشكلة في استرداد اشارة حقيقية من بيانات غير كاملة او غير مباشرة او مشوشه. ان تحليل المويجات يساعد في حل هذه المشكلة من خلال تقنية تسمى تقليص ضوضاء البيانات. وان عمل هذه التقنية هو عند تحليل مجموعة من البيانات بواسطة المويجات فيتم استعمال المرشحات التي تنتج مرشحات تردد عالي ومرشحات تردد واطئ كما تم توضيحها عند ايجاد المتجه (Y_j) ويحتوي هذا المتجه على المعاملات التفصيلية, اذا كانت بعض هذه التفاصيل صغيرة فقد يتم حذفها دون التأثير بشكل كبير على الميزات الرئيسة لمجموعة البيانات. فكرة العتبة ، إذاً ، هي ضبط صفر على جميع المعاملات التفصيلية التي تقل قيمها عن قيمة حد العتبة. هذه التقنية هي خطوة مهمة في التعامل مع البيانات المزعجة و التقليل من الضوضاء. والنتيجة هي تنظيف الإشارة لإظهار التفاصيل المهمة.[20]

اقترح هذه التقنية الباحثان Donoho and Johnstone عام (1994) وقدما نوعان من الدوال. - دالة العتبة الصلبة (Hard threshold function) وتكون حسب الصيغة الاتية:[4]

$$\theta_\lambda^H(t) = \begin{cases} t\,, & if \quad |t| > \lambda \\ 0, & if \quad |t| \le \lambda \end{cases} \qquad \dots(6)$$

اذ تعمل على الحفاظ على المعاملات التفصيلية التي تكون قيمها اكبر من قيمة حد العتبة ويحول التفاصيل الصغيرة (اصغر من قيمة حد العتبة) الى الصفر.

– دالة العتبة الناعمة (Soft threshold function) تكون حسب الصيغة الاتية:

$$\theta_\lambda^S(t) = \begin{cases} t-\lambda\,, & if \quad t > \lambda \\ t+\lambda\,, & if \quad t < -\lambda \\ 0\,, & if \quad |t| \le \lambda \end{cases} \qquad \dots(7)$$

اذ تعمل على تقليص المعاملات التفصيلية التي تكون قيمها اكبر من قيمة حد العتبة ويحول التفاصيل الصغيرة (اصغر من قيمة حد العتبة) الى الصفر. هنالك عدة طرائق في إيجاد قيمة حد العتبة (threshold) ومن هذه الطرائق طريقة العتبة الشاملة (Universal threshold method) ويرمز لها بالرمز λ_{UV} قدمها كل من (Donoho and Johnstone) عام 1994

وطريقة العبور الشرعي (Cross Validation (CV والتي اقترحها الباحث Nason عام 1996 اذ يتم تقسيم البيانات الى مجموعتين متساويتين بحجم العينة المجموعة الاولى تحتوي على الاعداد الزوجية والمجموعة الثانية تحتوي على الاعداد الفردية بحيث يتم استعمال البيانات المرتبة الفردية "للتنبؤ" بالبيانات المرتبة "الزوجية" والعكس بالعكس. تسمى هذه استراتيجية "Leave-out-half".[19] [20]

اقترح الباحث Leming Qu[22] عام 2003 طريقة اخرى في تقدير الانموذج الخطي الجزئي بإستعمال دالة حد العتبة وهي طريقة مشابهة الى طريقة Speckman . اذ يمكن تطبيق التمهيد المويجي (Wavelet Smoother) على هذه الطريقة كما في التمهيد اللبي (*Kernel Smoother*) وتمهيد الشريحة (Spline Smoother).

بإستعمال طريقة Speckman في تقدير الانموذج الخطي الجزئي يكون المقدر المعلمي $\hat{\beta}$ والمقدر اللامعلمي $\hat{g}(t)$ بالصيغ الاتية:[15] [28]

$$\hat{\beta} = \left(X'(I-K)'(I-K)X\right)^{-1}X'(I-K)'(I-K)Y$$

اذ ان:

$$\tilde{Y} = (I-K)Y = Y - KY$$
$$\tilde{X} = (I-K)X = X - KX$$

لذا يكون المقدر المعلمي $(\hat{\beta}_{LS})$ والمقدر اللامعلمي $(\hat{g}(t))$ كما في الصيغة الاتية:

$$\therefore\ \hat{\beta}_{LS} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} \qquad \ldots(8)$$

$$\therefore\ \hat{g}(t) = K\left(Y - X\hat{\beta}\right) \qquad \ldots(9)$$

تمثل $S(Y)$ و $S(X)$ تقليص المويجة اللاخطي الممهد التي يتم إيجادها للحصول على المقدر المعلمي $\hat{\beta}_W$ والمقدر اللامعلمي $\hat{g}_W(t)$ كما في ادناه:

$$\underline{Y} = \underline{g} + \underline{\varepsilon} \qquad \ldots(10)$$

يتم ضرب المعادلة رقم (10) بمصفوفة التحويل المويجي المتقطع W ليصبح الانموذج بالصيغة الاتية:

$$WY = Wg + W\varepsilon \qquad \ldots(11)$$

اذ ان: $w = WY, \quad \theta = Wg$ لذا يمكن اعادة كتابة المعادلة (10) بصيغة اخرى:

$$Y = W^{-1}\theta + \varepsilon \qquad \ldots(12)$$

يتم إيجاد المقدر $\hat{\theta}$ من خلال استعمال حد العتبة الناعمة : [2]

$$\hat{\theta}_s = sgn(w) \circ (|w| - \lambda e)_+ = \arg\min_{\theta_i} \frac{1}{2}(w_i - \theta_i)^2 + \lambda|\theta_i|$$
$$i = 1, \ldots\ldots, n \qquad \ldots(13)$$

اذ تشير λ الى قيمة حد العتبة, و $(\circ)$ تعني ضرب المكونات بين متجهين.

بعدها يمكن الحصول على المقدر اللامعلمي $\hat{g}$ او $S(Y)$ لمتغير الاستجابة (Y) بضرب مقدر حد العتبة الناعمة $\hat{\theta}_s$ بمعكوس مصفوفة التحويل المويجي W.

$$\hat{g}(Y) = S(Y) = W^{-1}\hat{\theta}_s \qquad \ldots(14)$$

و لإيجاد $S(X)$ يتم تطبيق ما ذكر انفاً على كل متغير من المتغيرات التوضيحية X, بعدها يتم تطبيق التمهيد المويجي لنحصل على: [22]

$$\tilde{Y}_W = (I-S)Y = Y - S(Y)$$
$$\tilde{X}_W = (I-S)X = X - S(X)$$

بعدها يتم إيجاد مقدرات الجزء المعلمي حسب الصيغة الاتية:

$$\therefore\ \hat{\beta}_W = \left(\tilde{X}'_W\tilde{X}_W\right)^{-1}\tilde{X}'_W\tilde{Y}_W \qquad \ldots(15)$$

ولتقدير الجزء اللامعلمي من الانموذج الخطي الجزئي (PLM) ليكن:

$$Z = Y - X\hat{\beta}_W \qquad \ldots(16)$$

اذ ينتج متجه جديد نرمز له بالرمز (Z) ويتم تقليص المويجة اللاخطي الممهد بواسطة حد العتبة الناعمة كما تم تعريفها سابقاً على المتجه (Z) ليصبح كما في الصغية الاتية:

$$\underline{Z} = \underline{g} + \underline{\varepsilon} \qquad \ldots(17)$$

$$WZ = Wg + W\varepsilon \qquad \ldots(18)$$

اذ ان: $D = WZ$ و $\gamma = Wg$ لذا يمكن اعادة كتابة المعادلة (17) بصيغة اخرى:

$$Z = W^{-1}\gamma + \varepsilon \qquad \ldots(19)$$

$$\hat{\gamma}_s = sgn(D) \circ (|D| - \lambda e)_+ \qquad \ldots(20)$$

بعدها يتم الحصول على التقدير النهائي الخاص بالجزء اللامعلمي كما في الصيغة الاتية:

$$\hat{g}_W(t) = S(Z) = S\left(\mathrm{Y} - \mathrm{X}\hat{\beta}_{\mathrm{W}}\right) = W^{-1}\hat{\gamma}_s \qquad \dots(21)$$

3- **التمهيد اللبي الخطي الموضعي** :Local Linear Kernel Smoother

تطورت الدراسات في العقود الثلاثة الماضية لتقدير الانموذج الخطي الجزئي (PLM) ومن اشهر هذه الطرق التي تعتمد على دوال اللب Kernel ودوال الشريحة Spline والمويجات Wavelets لكل من هذه الدوال لديها نقاط قوة ونقاط ضعف وان مقدرات انحدار Kernel تتمتع بالعديد من المميزات والبساطة الرياضية في التطبيق وايضاً تستعمل في حالة التصميم الثابت والعشوائي وهنالك عدة طرق في تمهيد Kernel ومن اهمها هو تمهيد (NADARYA WATSIN) عام 1964 والتمهيد الخطي الموضعي (Local Linear Smoother): [29][17]

اقترحه كل من قبل الباحثان (Stone) عام 1977 و (Cleveland) عام 1979 اذ يمتاز بتحيز متقارب وشروط تباين لا تتاثر عند الحدودية وحجم نافذة متغير لتقليل من تاثير القيم الشاذة كما اثبت الباحث n عام 1993 بإستعمال MiniMax ان المقدرات تحقق افضل احتمالية ثبات ومعدلات تقارب ممكنة وتستعمل α و β كحل لمكشة التقليل الاتية: [18] [7]

$$\min_{(a,b)} \sum_{j=1}^{n} (y_j - \alpha - \beta(t_j - T))^2 K\left(\frac{t_j - T}{h}\right) \qquad \dots(22)$$

حيث يكون المقدرين $\hat{a} = \hat{a}(t)$ و $\hat{\beta} = \hat{\beta}(t)$ هما الحل الامثل لتقليل المشكلة. في التمهيد الخطي الموضعي تعرف S_t بالصيغة الاتية: [8] [16]

$$S_t = e_t^T (u_t^T w_t u_t)^T u_t^T w_t \qquad \dots(23)$$

اذ ان:

$$u_t = \begin{bmatrix} 1 & (T_1 - t) \\ \vdots & \vdots \\ 1 & (T_n - t) \end{bmatrix}$$

$$W_t = diag\,\{K_h(T_1 - t), \quad \dots\dots \quad , K_h(T_n - t)\}$$

$$e_t^T = (1, 0, \dots, 0)$$

لذا يمكن الحصول على مصفوفة S المستعملة لايجاد التقديرات النهائية للانموذج الخطي الجزئي (PLM):

$$S = \left(S_{t_1}^T, \dots \dots \dots, S_{t_n}^T\right)^T \qquad \dots(24)$$

بإستعمال خورازمية التراجع التكراري Backfitting يتم الحصول على تقدير المعلمة $\hat{\beta}$ للجزء المعلمي ومقدر الدالة اللامعلمية $\hat{g}(t)$ حسب الصيغ الاتية: [12]

$$\hat{\beta}_{back} = (X^T(I - S)X)^{-1} X^T (I - S) Y \qquad \dots(25)$$

$$\hat{g}_{back} = S\left(Y - x^T \hat{\beta}_{back}\right) \qquad \dots(26)$$

اذ ان I تمثل مصفوفة احادية ابعادها $n \times n$. في التمهيد الخطي الموضعي يتبين من المقدر المعلمي واللامعلمي بانها تحقق افضل معدلات احتمالية من التقارب في المشاكل الشبه معلمية المحددة. ومن شروط هذا التقدير: [8] [12]

- **الدوال $m_y(t), m_x(t), g(t)$ لديها مشتقة ثانية محددة ومستمرة**
- **$\sigma_x^2(t), \sigma^2(x,t)$ محددة من الاعلى وبعيدة عن الصفر اي ان:**

$$\sup_{x,t} \sigma^2(x,t) < \infty \quad and \quad \sup_t \sigma_x^2(t) < \infty$$

$$0 < \inf_{x,t} \sigma^2(x,t) \quad and \quad 0 < \inf_t \sigma_x^2(t)$$

❖ المتغير العشوائي t لديه دالة كثافة احتمالية مستمرة $f(.)$ ذات دعم مرصوص, دالة الكثافة $f(.)$ محددة بعيدة عن الصفر ومنتهية بـ C_f حيث ان $C_f \subset R^2$.

❖ عرض الحزمة للدوال المقدرة $m_y(t), m_x(t)$ هي $O(n^{-1/5})$

كما ان أختيار معلمة التمهيد او عرض الحزمة **(bandwidth)** تكون أكثر أهمية من أختيار داله **Kernel** لاداء مقدر انحدار **Kernel** لكن عندما يتم أختيار دوال **Kernel** يجب ان تكون غير سالبة ومتماثلة حول الصفر ومستمرة ولديها مشتقة ثانية, في الجدول ادناه دوال **Kernel** المستعملة في الجانب التجريبي: [18]

#	Kernel	Explict form
1	Gaussian kernel	$K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2), u \in [-\infty, \infty]$
2	Uniform kernel	$K(u) = \frac{1}{2}, \quad u \in [-1,1]$

جدول (1) يوضح دوال **Kernel**

بما ان اختيار معلمه التمهيد او عرض الحزمة **(Bandwidth)** اهم عملية في التقدير سيتم استعمال طريقة العبور الشرعي **(Cross Validation)** لإيجاد قيمة معلمة التمهيد. [26]

4- الجانب التجريبي:

الأساليب التحليلية والمحاكاة هي أساليب نمذجة تهدف إلى توفير فكرة عن أداء النظام، في ظروف مختلفة. اذ ان النموذج التحليلي عبارة عن تجريد رياضي يمكن تمديده لمعالجة مختلف ظروف العمل، وذلك بفضل بعض الافتراضات حول الطريقة التي تتم بها العملية.

هنالك العديد من المزايا لاستعمال بيانات المحاكاة اذ تستعمل أرقاماً عشوائية يتم إنشاؤها عوضاً عن البيانات التي يتم جمعها, وهي أسرع بكثير من جمع البيانات التقليدية، لذلك يمكن إجراء الاختبارات بسرعة أكبر. ويمكن توليد بيانات تكون قريبة من الواقع العملي عند صعوبة جمع البيانات او هنالك مشاكل في البيانات التي يتم جمعها.

لذا سيتم استعمال عدة تجارب لاجراء مقارنة ما بين طريقتي التمهيد الموجي والتمهيد الخطي الموضعي في تقدير الانموذج الخطي الجزئي (PLM), لتحديد اي من هاتين الطريقتين هي الافضل في عملية التقدير. تم إستعمال برنامج لغة (R) لتوليد المتغيرات العشوائية وبناء نماذج المحاكاة لغرض اجراء المقارنة.

4-1 النماذج المستعملة في المحاكاة:

ولبناء نماذج المحاكاة علينا اولاً توليد متغيرات عشوائية بأحجام عينات وتباينات مختلفة. تتوزع المتغيرات التوضيحية المعلمية x توزيعاً طبيعياً بوسط حسابي مساو للصفر وتباين مساوٍ للواحد ويتم توليدها على هذا الاساس:.

$$x_j \sim Normal\ (0,1) \quad , \qquad j = 1, 2, 3$$

والاخطاء العشوائية تتوزع ايضاً توزيعاً طبيعياً بوسط حسابي مساو للصفر وتباين σ^2 اذ تم افتراض 3 قيم لتباين الخطأ وهي **(0.5, 1, 2)**

$$\varepsilon \sim N\ (0, \sigma^2)$$

المتغير التوضيحي اللامعلمي t يتم توليده حسب الصيغة الاتية للفترة ما بين **[0,1]**:

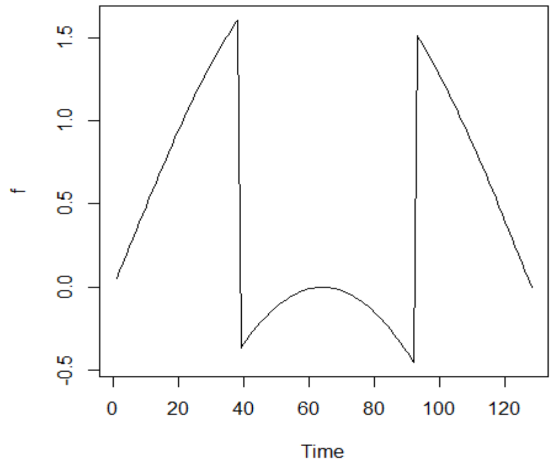
$$t_i = {}^{i}/_{n} \quad , \qquad , \qquad i = 1, \dots \dots, n$$

تم اختيار 3 دوال واسعة الانتشار ومستعملة في كثير من البحوث للحصول على قيم الدالة اللامعلميه في انموذج الانحدار الخطي الجزئي والتي تتناسب مع الطرائق المستعملة في عملية التقدير. استخلصت من عدة ابحاث:

4-1-1 دالة Heavisine:

وهي دالة متقطعة وليست مستمرة اقترحها الباحثان **Bruce, A., And H.-Y. Gao** عام (1996) وتكون بالصيغة الاتية: [22]

$$\text{Heavisine}(t) = 4\sin(4\pi t) - \text{sign}(t-0.3) - \text{sign}(0.72-t)$$

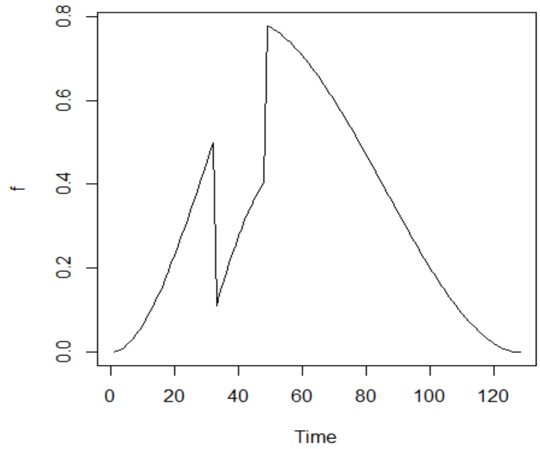


شكل (1) يوضح دالة Heavisine مع حجم عينة n = 128

4-1-2 دالة متعددة الحدود المجزئة Piecewise Polynomial:

وهي دالة متعددة حدود مجزئة غير مستمرة, اقترحها الباحث **Nason** عام **1996** وتكون بالصيغة الاتية: [2]

$$g(t) = \begin{cases} 4\,t^2\,(3-4t) & if\ 0 \le t \le 0.5 \\ \frac{3}{4}\,t(4t^2-10t+7)-1.5 & if\ 0.5 < t \le 0.75 \\ \frac{16}{3}t(t-1)^2 & if\ 0.75 < t \le 1 \end{cases}$$

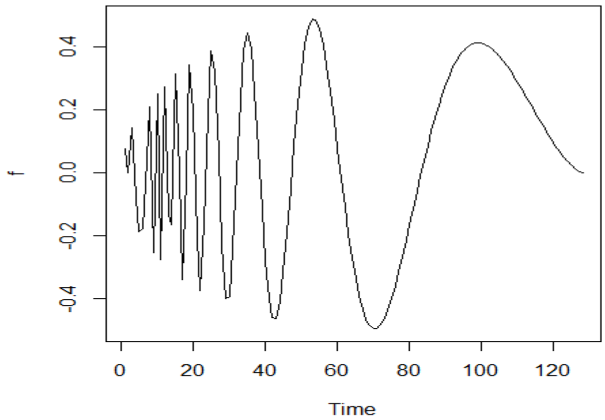


شكل (2) يوضح دالة متعددة الحدود المجزئة piecewise polynomial مع حجم عينة n=128

4-1-3 دالة Doppler:

دالة ذات ترددات متغيرة اقترحها الباحثان Donoho and Johnstone عام 1994 وتكون بالصيغة الاتية:
[22]

$$doppler\ (t) = \sqrt{t\ (1-t)}\ \sin\left(\frac{2.1\ \pi}{t+0.05}\right)$$



شكل (3) يوضح دالة Doppler مع بحجم عينة n = 128

5- النتائج والمناقشة Results and Discussion:

تم تنفيذ تجارب المحاكاة بإستعمال اربعة حجوم للعينات $(n = 64, 128, 256, 512)$ وتكرار يساوي 1000. اذ سيتم تفسير وتحليل النتائج عن طريق مقارنة المقدرات بمعيار MASE حسب كل إنموذج كما يأتي:

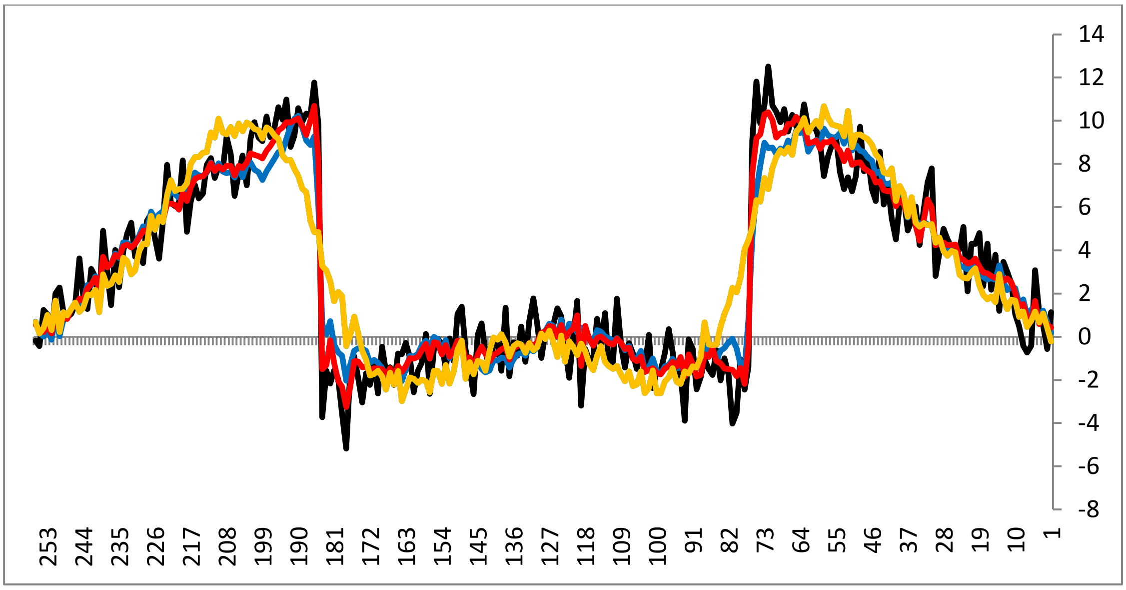
5-1 الانموذج الاول:

في ادناه نتائج محاكاة الانموذج الاول عند اختيار دالة Heavisine كدالة لامعلميه:

σ	N	Wavelet (Cross Validation)	Wavelet (Universal)	Local Linear (Uniform function)	Local Linear (Gaussian function)
0.5	64	0.6302	0.5509	4.2804	6.4206
	128	0.3313	0.2998	3.0319	4.5605
	256	0.1436	0.2091	3.0319	3.4916
	512	0.0902	0.1481	2.1692	2.5530
1	64	0.8729	0.7661	4.3371	6.4661
	128	0.5440	0.4939	3.1053	4.6201
	256	0.2083	0.3726	2.2316	3.5567
	512	0.1529	0.2726	1.6463	2.6204
2	64	1.2291	1.0974	4.4477	6.5828
	128	0.8377	0.8242	3.2079	4.7486

	256	0.3363	0.6490	2.3633	3.6898
	512	0.2794	0.5057	1.7745	2.7611

جدول (2) يوضح معدل متوسط مربعات الخطأ MASE للانموذج الاول



شكل (4) يوضح القيم الحقيقية والقيم التقديرية للمتغير المعتمد Y للانموذج الاول بإستعمال تباين (1) وحجم عينة (256)

ويكون تفسير النتائج المذكورة انفاً كالاتي:

1- طريقة التقدير المويجي (Wavelet Estimation) للانموذج الخطي الجزئي (PLM) هي افضل من طريقة التقدير الخطي الموضعي (Local Linear Smoother) بإختلاف التباينات وعند جميع احجام العينات.

2- ان التقدير المويجي (Wavelet Estimation) للانموذج الخطي الجزئي (PLM) بإستعمال طريقة العتبة الشاملة (Universal threshold method) تكون افضل من طريقة العبور الشرعي (Cross-Validation method) عند حجوم العينات (n = 64, 128) ولجميع قيم التباين ولكن تكون طريقة العبور الشرعي (Cross-Validation method) افضل من طريقة العتبة الشاملة (Universal threshold method) عند حجوم العينات (n= 256, 512) ولجميع قيم التباين عند تقدير الانموذج الخطي الجزئي (PLM).

3- ان طريقة التمهيد الخطي الموضعي (Local Linear Smoother) للانموذج الخطي الجزئي (PLM) بإستعمال دالة Uniform افضل من استعمال دالة Gaussian ولجميع قيم التباين واحجام العينات ولكن تقل نسبة الفرق عند زيادة حجم العينة. كما ان دالة Uniform تكون افضل من دالة Gaussian بسبب ان توليد البيانات يكون حسب توزيع Uniform.

4- تزداد قيمة MASE بزيادة قيمة التباين

5- تقل قيمة MASE عند زيادة حجم العينة

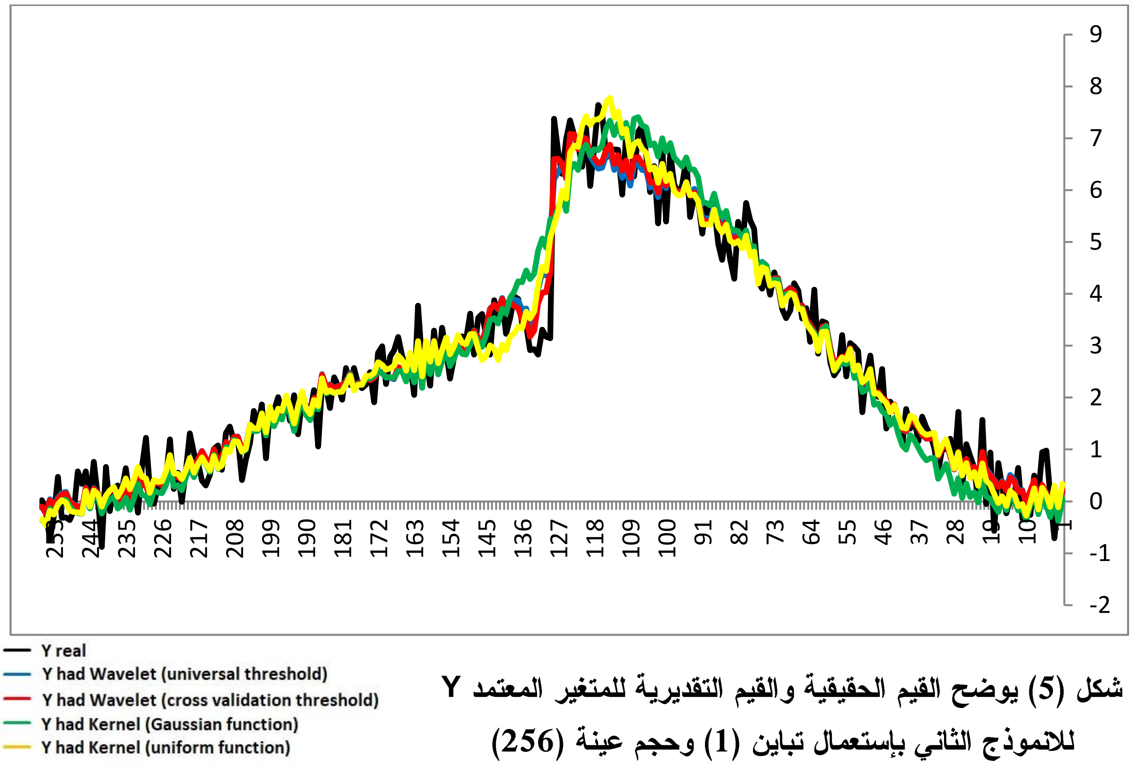
6- من خلال شكل (4) يتضح ان القيم التقديرية للمتجه Y بإستعمال طريقة التمهيد المويجي تكون ممهدة اكثر وذات تشتت اقل بسبب صغر قيمة متوسط مربعات الخطأ كما ان القيم التقديرية للمتجه Y بإستعمال طريقة التمهيد الخطي الموضعي لكل من دالة Gaussian ودالة Uniform تكون متطابقة. وهذا يثبت ان طريقة التمهيد المويجي هي افضل من طريقة التمهيد اللبي في تقدير الانموذج الخطي الجزئي (PLM).

5-2 الانموذج الثاني:

في ادناه نتائج محاكاة الانموذج الثاني عند اختيار دالة متعددة الحدود المجزئة **Piecewise Polynomial** كدالة لامعلميه:

σ	N	Wavelet (Cross Validation)	Wavelet (Universal)	Local Linear (Uniform function)	Local Linear (Gaussian function)
0.5	64	0.0961	0.0972	0.2426	0.4333
	128	0.0862	0.0945	0.1673	0.3367
	256	0.0770	0.0924	0.1337	0.2286
	512	0.0713	0.0883	0.1156	0.1592
1	64	0.1718	0.1758	0.2987	0.4902
	128	0.1495	0.1745	0.2281	0.3998
	256	0.1432	0.1730	0.1980	0.2942
	512	0.1390	0.1685	0.1813	0.2272
2	64	0.3118	0.3118	0.4066	0.6048
	128	0.2768	0.3276	0.3468	0.5259
	256	0.2744	0.3246	0.3243	0.4271
	512	0.2729	0.3238	0.3142	0.3616

جدول (3) يوضح معدل متوسط مربعات الخطأ MASE للانموذج الثاني



شكل (5) يوضح القيم الحقيقية والقيم التقديرية للمتغير المعتمد Y للانموذج الثاني بإستعمال تباين (1) وحجم عينة (256)

ويكون تفسير النتائج المذكورة انفاً كالاتي:

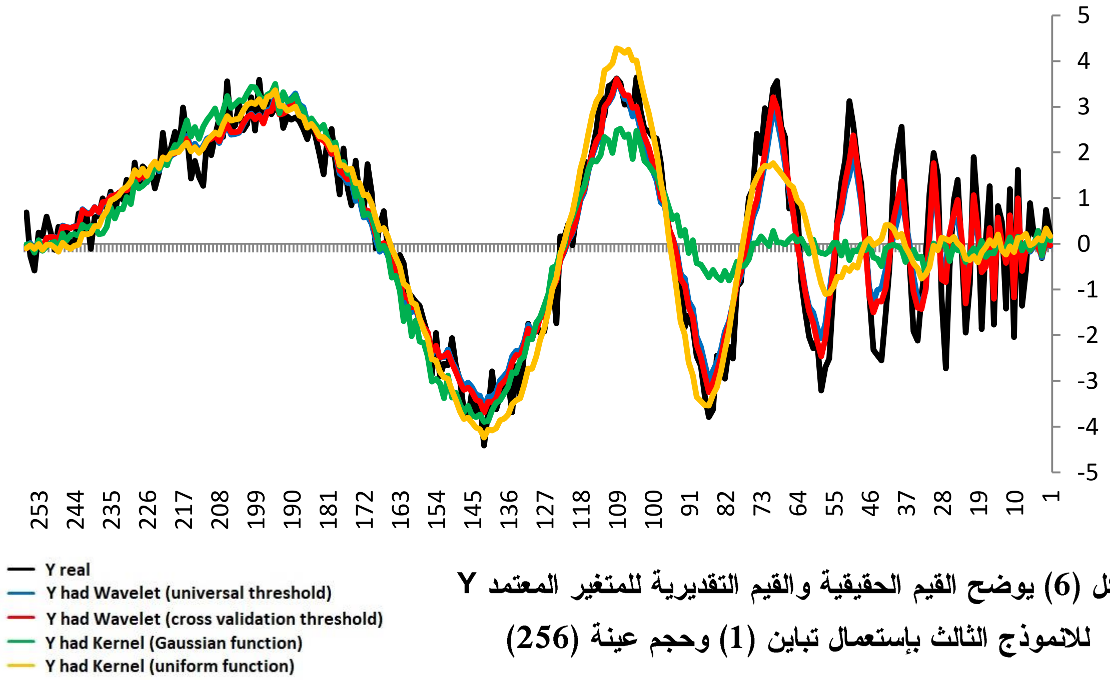
1- ان جميع قيم MASE للانموذج الثاني هي صغيرة ومتقاربة ونسبة التغيير فيها عند تغيير حجوم العينات او التباين تكون قليلة جداً مقارنة مع باقي الدوال المستعملة.
2- طريقة التقدير المويجي (Wavelet Estimation) للانموذج الخطي الجزئي (PLM) هي افضل من طريقة التقدير الخطي الموضعي (Local Linear Smoother) بإختلاف التباينات وعند جميع احجام العينات.
3- التقدير المويجي (Wavelet Estimation) للانموذج الخطي الجزئي (PLM) بإستعمال طريقة العبور الشرعي (Cross-Validation method) تكون افضل من طريقة العتبة الشاملة (Universal threshold method) عند جميع حجوم العينات ولجميع قيم التباين.
4- طريقة التمهيد الخطي الموضعي (Local Linear Smoother) للانموذج الخطي الجزئي (PLM) بإستعمال دالة Uniform تكون افضل من استعمال دالة Gaussian ولجميع قيم التباين وحجوم العينات.
5- تزداد قيمة MASE بزيادة قيمة التباين
6- تقل قيمة MASE عند زيادة حجم العينة
7- من خلال شكل (5) يتضح ان القيم التقديرية للمتجه Y بإستعمال طريقة التمهيد المويجي لكلتا الحالتين (العبور الشرعي Cross-Validation و العتبة الشاملة Universal threshold) تكون متطابقة تقريباً وممهدة اكثر وذات تشتت اقل بسبب صغر قيمة MASE, وهذا يثبت ان طريقة التمهيد المويجي هي افضل من طريقة التمهيد اللبي في تقدير الانموذج الخطي الجزئي (PLM)

5-3 الانموذج الثالث:

في ادناه نتائج محاكاة الانموذج الثالث عند اختيار دالة Doppler كدالة لامعلميه:

σ	N	Wavelet (Cross Validation)	Wavelet (Universal)	Local Linear (Uniform function)	Local Linear (Gaussian function)
0.5	64	0.5567	0.5312	1.3714	2.0888
	128	0.3765	0.3486	1.1532	1.6935
	256	0.2138	0.2256	0.8305	1.2821
	512	0.0898	0.1666	0.6089	0.9418
1	64	0.6763	0.7283	1.4164	2.1406
	128	0.5015	0.5180	1.2152	1.7597
	256	0.2844	0.3755	0.8953	1.3473
	512	0.1499	0.2879	0.6756	1.0086
2	64	0.8580	1.0295	1.5287	2.2655
	128	0.6563	0.7950	1.3361	1.8829
	256	0.4147	0.6331	1.0192	1.4761
	512	0.26731	0.49925	0.8080	1.1419

جدول (4) يوضح معدل متوسط مربعات الخطأ MASE للانموذج الثالث



شكل (6) يوضح القيم الحقيقية والقيم التقديرية للمتغير المعتمد Y للانموذج الثالث بإستعمال تباين (1) وحجم عينة (256)

ويكون تفسير النتائج المذكورة انفاً كالاتي:

1- طريقة التقدير المويجي (Wavelet Estimation) للانموذج الخطي الجزئي (PLM) هي افضل من طريقة التقدير الخطي الموضعي (Local Linear Smoother) بإختلاف التباينات وعند جميع احجام العينات.

2- التقدير المويجي (Wavelet Estimation) للانموذج الخطي الجزئي (PLM)

- عند قيمة تباين 0.5: تكون طريقة العتبة الشاملة (Universal threshold method) افضل من طريقة العبور الشرعي (Cross-Validation method) عند حجوم العينات (n = 64, 128) ولكن تكون طريقة العبور الشرعي (Cross-Validation method) افضل من طريقة العتبة الشاملة (Universal threshold method) عند حجوم العينات (n= 256, 512)

- عند قيمة تباين (1) و (2): تكون طريقة (Cross-Validation method) افضل من طريقة العتبة الشاملة (Universal threshold method) لجميع حجوم العينات

3- طريقة التمهيد الخطي الموضعي (Local Linear Smoother) للانموذج الخطي الجزئي (PLM) بإستعمال دالة Uniform افضل من استعمال دالة Gaussian ولجميع قيم التباين واحجام العينات.

4- تزداد قيمة MASE بزيادة قيمة التباين

5- تقل قيمة MASE عند زيادة حجم العينة

6- من خلال شكل (6) يتضح ان القيم التقديرية للمتجه Y بإستعمال طريقة التمهيد المويجي لكلتا الحالتين (العبور الشرعي Cross-Validation و العتبة الشاملة Universal threshold) تكون متطابقة تقريباً وممهدة اكثر وذات تشتت اقل بسبب صغر قيمة متوسط مربعات الخطأ وهذا يثبت ان طريقة التمهيد المويجي هي افضل من طريقة التمهيد اللبي في تقدير الانموذج الخطي الجزئي (PLM)

المصادر:

1.Antonini, M., Barlaud M., Mathieu P., Daubechies I., 1992. "Image Coding Using Wavelet Transform". IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. I, NO 2., 205-220.

2.Changa X, Qub L., 2004. "Wavelet estimation of partially linear models". Computational Statistics & Data Analysis 47 31–48.

3.CHOEN, A.; DAUBECHIES, I.; VIAL, P.; 1993. "Wavelet on the interval and fast wavelets transform". Applied and Computational Harmonic Analysis 1, 54-81.

4.Donoho D.L., Johnstone I.M., 1994. "Ideal spatial adaptation by wavelet shrinkage". Biometrika 81, 425–455.

5.Engle F., C. W. J. Granger, Rice J. and Weiss A., 1986. "Semiparametric Estimates of the Relation Between Weather and Electricity Sales". Journal of the American Statistical Association Vol. 81, No. 394 pp. 310-320.

6.Eubank R.L., Kambour E.L., Kim J.T., Klipple K., Reese C.S., Schimek M., 1998. "Estimation in partially linear models. Comput. Statist". Data Anal. 29, 27–34.

7.FAN J., 1993. "Local Linear Regression Smoothers and Their Minimax Efficiencies". The Annals of Statistics, Vol. 21, No. 1, 196-216.

8.Hamilton S.A., Truong Y.K., 1997. "Local linear estimation in partly linear models". J. Multivariate Anal. 60 (1), 1–19.

9.Hardle W., Kerkyacharian G., Picard D., Tsybakov A., 1997. "Wavelets, Approximation and Statistical Applications". Seminars Berlin-Paris.

10. Hardle W., Müller M., Sperlich S., Werwatz A., 2004. "Nonparametric and Semiparametric Models". Springer-Verlag Berlin Heidelberg.

11. Li C., 2013. "A Partially Linear Model with Its Applications". Communications in Statistics—Simulation and Computation R [1], 42: 1673–1680.

12. Liang H., 2006. "Estimation in partially linear models and numerical comparisons". Computational Statistics & Data Analysis 50, 675–687.

13. Mallat S.G., 1989. "A theory for multiresolution signal decomposition: the wavelet representation". IEEE Trans. Pattern Anal. Mach. Intell. 11, 674–693.
14. Mohammed Habib Kadhum Al- Sharoot, 2006, " Wavelet Analysis For Estimating Nonparametric Regression Curve"", A dissertation for the PH.D Degree of Science in Statistics, University of Baghdad / College of Administration and Economic.
15. Munaf Y. Hmood; Aseel M.; 2012. "A comparison Of Some Semiparametric Estimators For consumption function Regression". Journal of Economic and Administrative Sciences Vol.12, NO.67, PP 273-288.
16. Munaf Y. Hmood; Mayasa M., 2014.: A comparison of the Semiparametric estimator's model using different smoothing methods". Journal of Economic and Administrative Sciences Vol.20, NO.75, PP 376-394.
17. Munaf Y. Hmood; Mayasa M.; 2015. "Estimate the General Trend using Semiparametric Regression Models with a Forecast Value of GDP in Iraq". Journal of Al-Rafidein College. NO.35, PP 26-44.
18. Munaf Y. Hmood, 2000, " Comparing nonparametric Kernel Estimators to Estimate Regression Functions", A thesis for degree of MSc in Statistic, University of Baghdad / College of Administration and Economic.
19. Nason G.P., 1996. "Wavelet shrinkage using cross-validation". J. Roy. Statist. Soc. Ser. B 58, 463–479.
20. Nason G.P., 2008. "Wavelet Methods in Statistics with R". Springer Science +Business Media, LLC.
21. Prada-Sanchez J. M., Febrero-Bande M., Cotos-Yanez T., Gonzalez-Manteiga W., Bermudez-Cela J. L. and Lucas-Dominguez T., 2000. "Prediction of SO2 pollution incidents near a power station using partially linear models and an historical matrix of predictor-response vectors". Environmetrics; 11: 209-225.
22. Qu L., 2003. "Wavelet thresholding in partially linear models: a computation and simulation". Appl. Stochastic Models Bus. Ind; 19:221–230 (DOI: 10.1002/asmb.499).
23. Ruch D. K., Van Fleet P. J., 2009. "WAVELET THEORY an Elementary Approach with Applications". John Wiley & Sons.
24. Robert F. Engle, C. W. J. Granger, John Rice and Andrew Weiss; 1986. "Semiparametric Estimates of the Relation between Weather and Electricity Sales" Journal of the American Statistical Association Vol. 81, No. 394, pp. 310-320.
25. Saad Kadhem, 2015, " Wavelet Analysis For Regression Model With Missing Data", A dissertation for the PH.D Degree of Science in Statistics, University of Baghdad / College of Administration and Economic.
26. Schindler A., Köhler M. and Sperlich S., 2014. "A Review and Comparison of Bandwidth Selection Methods for Kernel Regression ", International Statistical Review, 82, 2, 243–274.
27. Schmalensee R., Stoker T. M., 1999. "Household Gasoline Demand in the United States". Econometrica, Vol. 67, No. 3, 645-662.
28. Speckman, P., 1988. "Kernel smoothing in partial linear models". J. Roy. Statist. Soc. Ser. B 50, 413–436.
29. Wand M.P., Jones M.C., 1995. "Kernel Smoothing", SPRINGER.
30. Zeger, S. L., & Diggle, P. J., 1994. "Semiparametric Models for Longitudinal Data with Application to CD4 Cell Numbers in HIV Seroconverters". Biometrics, Vol. 50, No. 3, pp. 689-699.

Estimate the Partial Linear Model Using Wavelet and Kernel Smoothers

Younis Amer Hassan
Younis.amer91@gmail.com

Full Prof. Dr. Munaf Yousif Hmood
munaf_yousif@yahoo.com

University of Baghdad - College of Administration and Economics Statistic Department

Received :14/12/2019 Accepted :15/1/2020 Published :April / 2020

Abstract:

This article aims to estimate the partial linear model by using two methods, which are the Wavelet and Kernel Smoothers. Simulation experiments are used to study the small sample behavior depending on different functions, sample sizes and variances. Results explained that the wavelet smoother is the best depending on the mean average squares error criterion for all cases that used.

Key Words: **Partial Linear Regression Models, Wavelet Smoother, Local Linear Smoother.**