

استخدام قيم عتبة مختلفة في مقارنة بعض طرائق التقدير المويجي

لدالة الانحدار اللامعلمي بوجود بيانات مفقودة

أ.د. ظافر حسين رشيد

rasheddhafir@yahoo.com

جامعة بغداد - كلية الادارة والاقتصاد

م.م. سعد كاظم حمزة

sdkadem@yahoo.com

رئاسة جامعة بغداد - قسم الشؤون الادارية

المستخلص:

تعد مشكلة فقدان بعض مشاهدات العينة من اهم المشاكل التي تواجه الباحث اثناء التحليل الاحصائي ، ومن أسباب الفقدان كثيرة كأن تكون التلف والإهمال او موت مرضى كما في الدراسات السريرية . وان وجود مثل هكذا مشكلة ضمن البيانات يؤثر على التحليل وبالتالي يؤدي الى استنتاجات مظلمة ، وعلى الرغم من الكفاءة العالية للتقديرات المويجية في تقدير دالة الانحدار الا انها هي الاخرى تتأثر بمشكلة فقدان البيانات ، حيث انه بالاضافة الى تأثير مشكلة الفقدان على دقة التقدير فانه ليس بالامكان تطبيق هذه الطرائق لفقدان احد شروطها وهي حجم العينة الدايديكوي $n = 2^j$.

ونظراً للتأثير الكبيرة الناجم عن تلك المشكلة فان الكثير من الباحثين ممن كرسوا بحوثهم لمعالجة تلك المشكلة باستخدام طرائق تقليدية في معالجة البيانات المفقودة ، بينما قام الباحث باستخدام طرائق تعويض اكثر كفاءة لمعالجة البيانات المفقودة كمرحلة اولى كي تصبح البيانات جاهزة للتطبيق المويجي وقد اثبتت تجارب المحاكاة كفاءة الطرائق المقترحة على بقية الطرائق الاخرى ، كذلك تضمن البحث التصحيح التلقائي لمشكلة الحدودية عن طريق استخدام نموذج متعدد الحدود بالاضافة الى استخدام قيم عتبة مختلفة ضمن التقديرات المويجية كون ان الاختيار المناسب لقيمة العتبة يكون حاسم في دقة تلك التقديرات.

الكلمات الرئيسية: البيانات المفقودة، الانحدار المويجي، متعدد حدود، الحصين المضاعف اللامعلمي

1. المقدمة

تعد مشكلة فقدان بعض مشاهدات العينة من اهم المشاكل التي تواجه الباحث اثناء التحليل الاحصائي، ومن اسباب الفقدان كثيرة كأن تكون التلف والاهمال او موت مرضى كما في الدراسات السريرية.

وعلى الرغم من ان المعالجة الاولى لمشكلة البيانات المفقودة هي تجنب حدوثها الا انه في بعض الأحيان يضطر الى التعامل معها كونها قسرية الحدوث ولاسباب اهمها ان موضوع حدوثها يكون خارج سيطرة الباحث أو ان السيطرة عليها بشكل تام امر مكلف جداً ، وعليه يجب البحث عن طرائق لمعالجة هذه البيانات كون وجود هذه المشكلة يؤثر بشكل سلبي على تحليل البيانات وبالتالي يؤدي الى استنتاجات مظلمة ، وان هذه الاستنتاجات ناتجة من التحيز الكبير الذي تحدثه تلك المشكلة ، ونظراً للآثار الكبيرة التي تخلفها تلك المشكلة فان عدداً كبيراً من الباحثين كرس بحوثه ودراساته حول معالجة هذه المشكلة ، وقد شهد العقدين الاخيرين ازدياد أهمية النظرية المويجية بشكل منقطع النظير وذلك بسبب زيادة عدد التطبيقات والحقول الجديدة التي تدخلها مثل التشفير والبصريات والتعرف على الاصوات وفيزيولوجيا الأعصاب وتطبيقات الرياضيات البحتة وغيرها من العلوم التي يصعب حصرها ، ومن جهة اخرى التطور الحاصل بعلم الحاسوب هو الآخر ساهم بشكل كبير في زيادة تطبيقات تلك التقنيات ومن حينها أصبحت التقنية الأكثر شيوعاً للاستعمال ، الا ان هذه التقنيات تعاني هي الاخرى من مشكلة وجود البيانات المفقودة كون ان هذه المشكلة بالاضافة الى ما تسببه من تظليل في الاستنتاجات بسبب التحيز الناشئ عنها فانها تحول دون التطبيق التقديرات المويجية لانها تشترط اثناء تطبيقها تساوي المسافات بين المشاهدات (-Regular spaced data) وحجم العينة الدايدكي (Dyadic) أي $n = 2^J$ لكل J عدد صحيح . وان هذه القيود المفروضة على التقديرات المويجية دفعت الباحثين الى التفكير لايجاد الحلول لجعل البيانات ملائمة لهذا التطبيق المويجي وذلك عن طريق معالجة البيانات المفقودة وطرائق التعويض المعروفة، كذلك تضمن البحث استخدام قيم عتبة مختلفة عند تطبيق حد العتبة كون ان اختيار القيمة المناسبة للعتبة يكون حاسم في اعطاء دقة اكبر للتقديرات المويجية ، وكذلك تم استخدام نموذج الانحدار متعدد الحدود لمعالجة مشكلة الحدودية تلقائياً والتي تحدث بسبب التحويل المويجي (Wavelet Transformation).

تم تقسيم البحث الى ثمانية مباحث ، تضمن المبحث الاول المقدمة ومشكلة البحث والهدف من البحث ، اما المبحث الثاني فتضمن تعريف الانحدار المويجي وطرائق اختيار العتبة ، بينما تضمن المبحث الثالث نموذج وآلية الفقدان ، وتضمن المبحث الرابع طرائق تعويض البيانات المفقودة ، اما المبحث الخامس فقد تضمن التقدير لدالة الانحدار غير المعلومة ، بينما تضمن المبحث السادس دراسة المحاكاة ، وتضمن المبحث السابع الاستنتاجات واخيراً التوصيات ضمن المبحث الثامن .

1.1 مشكلة البحث

ان مشكلة البحث تتركز حول تأثير القيم المفقودة لمتغير الاستجابة y_i على دقة تقدير دالة الانحدار اللامعلمي غير المعلومة بالاضافة الى ان فقدان بعض القيم يجعل البيانات غير ملائمة لتطبيق التقدير المويجي كونه يشترط حجم العينة الدايديكي أي $n = 2^J$ وتساوي المسافات بين المشاهدات .

1.2 هدف البحث

يتناول البحث حالة انموذج انحدار لامعلمي

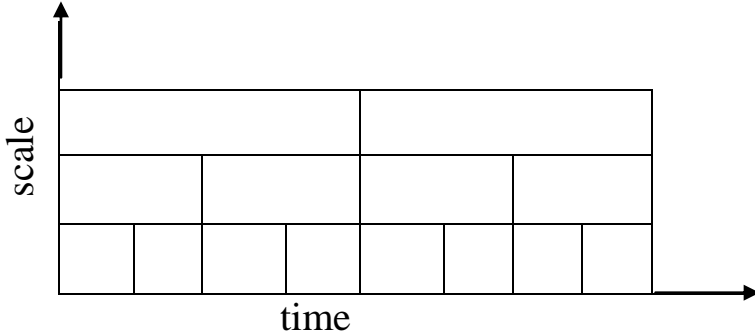
$$y_i = f(x_i) + e_i \quad \dots\dots\dots (1)$$

حيث ان $x_i = i/n$ ، أي أن $x \in [0,1]$ و f دالة غير معلومة و e_i يمثل التشويش (noise) و y_i تعاني من مشكلة فقدان في بعض مشاهداتها عشوائياً .
 ان الهدف من هذا البحث هو تقدير دالة الانحدار اللامعلمية $f(x_i)$ باستخدام الانحدار المويجي بواسطة ازالة التشويش ومعالجة مشكلة البيانات المفقودة في متغير الاستجابة والتصحيح التلقائي لمشكلة الحدودية وذلك عن طريق سلوك اتجاهان ، الاول معالجة مشكلة الحدودية التي تحصل بسبب التحويل المويجي عن طريق استخدام نموذج متعدد حدود من درجة قليلة ، والاتجاه الثاني استخدام طرائق تعويض كفاءة للبيانات المفقودة في متغير الاستجابة كمرحلة اولى ومن ثم تقدير دالة الانحدار غير المعلومة باستخدام التقديرات المويجية وقيم عتبة مختلفة .

2. تحويل المويجات

ان تحويل المويجات ظهر من اجل معالجة حالات الضعف التي تعانيها التحويلات السابقة ، اذ يعد هذا التحويل من اكثر التحويلات ذات التقنية المتقدمة في مجال معالجة الاشارة ، حيث يمكننا هذا التحويل من تحليل الاشارة الى مجموعة من المستويات متعددة الحل (Multiresolution) في كلاً من مجالي الزمن والتردد ، ويتم تطبيق هذا التحويل عن طريق استخدام مجموعة من المويجات (Wavelets) المختلفة .
 والمويجة هي اشارة محدودة الطول الزمني (الاستمرارية) تمتلك قيمة متوسطة مساوية للصفر، بخلاف التحويلات السابقة يستخدم التحويل المويجي نافذة متغيرة بدلاً من استخدام نافذة ثابتة، اذ يتم تغيير عرض النافذة باستمرار للحصول على معلومات مختلفة التردد على طول المويجة . فيتم الحصول على المويجات التي تختلف تردداتها بحسب عرض النافذة المستخدم ، رياضياً يقوم التحويل المويجي على ضغط المويجة

المراد معالجتها مع دالتين هما دالة الموجة الام $\psi(x)$ من اجل الحصول على مجموعة من المعاملات (coefficients) والتي تسمى معاملات الموجة او المعاملات التفصيلية $D(s, t)$ والثانية هي دالة القياس $\phi(x)$ وتسمى كذلك بدالة اللب من اجل الحصول على المعاملات التقريبية $A(s, t)$.



شكل (1) : يبين استخدام نافذة متغيرة في التحويل المويجي

2.1 تحويل المويجي المتقطع

ان التحويل المويجي المتقطع (DWT) هو خوارزمية كفوءة اقترحت من قبل الباحث Mallat (1989) لحساب معاملات الموجة لسلسلة من البيانات المشوشة من خلال عمليات خطية او عن طريق استخدام المصفوفات ، وان هذا التحويل يمكن تعريفه على انه عبارة عن مرشحين هما مرشح الممر الواطئ والذي يشار له بـ $H = \{hk\}$ ومرشح الممر العالي والذي يشار له بـ $G = \{gk\}$ بحيث ان gk, hk هما معاملات تلك المرشحات على التوالي.

ولتوضيح كيفية اجراء هذا التحويل ، لتكن f دالة معرفة على مشاهدات ذات مسافات متساوية $x_k, k = 0, 1, \dots, n$ وليكن $c_{j,k} = f(x_k)$ فان التحويل المويجي لـ $c_{j,k}$ يمكن حسابه باستخدام العلاقة التالية :

$$c_{j-1,k} = \sum h_n - 2k c_{j,n} \quad (2)$$

$$d_{j-1,k} = \sum g_n - 2k c_{j,n} \quad (3)$$

بحيث ان $d_{j-1,k}$ تسمى معاملات الموجة (معاملات التفصيل) ، بينما $c_{j-1,k}$ تسمى معاملات القياس ، اما عبارة $2k$ تشير الى ان ترشيح البيانات يتم باستخدام المرشحين H, G ومن ثم نسحب كل مشاهدة تسلسلها زوجي ولكل الاعداد الصحيحة ، وبالتالي فان كل مرشح سينتج نصف طول البيانات الاصلية الا ان الجزء الاهم من المخرجات هو الذي يتم الحصول عليه عن طريق الممر الواطئ (g) كونه يحتوي على اغلب المعلومات المحتواة في الاشارة الاصلية وبعبارة اخرى أي يكون الاهتمام منصب عادةً على التكرارات الواطنة كونها تعطي الاشارة بشكلها الموحد ، بينما سبب اهمال مخرجات الممر العالي يعود الى انه يضم التشويشات في الاشارة والتي تكون غير مرغوب بها وبالتالي يتم استبعادها ، ويتم الحصول على الاشارة الاصلية بالتجميع المتسلسل لكل العوامل الناتجة سابقاً (التقريبية و التفصيلية) بدءاً من آخر مرحلة تحليل [11,12,10] .

2.2 الانحدار اللامعلمي التقليدي

في هذا المبحث سنوضح بشكل موجز طريقة من طرائق التقدير اللامعلمي كونها تستخدم ضمن البحث وهي طريقة التقدير اللبي (Kernel Estimation) وهي ابسط اشكال الانحدار اللامعلمي ويمكن كتابتها بصورة عامة وفق الصيغة الآتية [6]

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{x-x_j}{h}\right) y_j = \frac{1}{n} \sum w_j y_j \quad (4)$$

حيث ان

$$w_j = K\left(\frac{x-x_j}{h}\right) / h \quad (5)$$

حيث ان $K_h(x)$ هي الدالة اللبية ، و h هي عرض الحزمة . وقد عدلت هذه الطريقة بواسطة مقدر (Nadaraya-Watson) ليكون بالصيغة التالية

$$\hat{f}_h(x) = \frac{\sum w_i y_i}{\sum w_i} \quad (6)$$

2.3 الانحدار المويجي

يعد تقدير الانحدار المويجي من الأساليب الحديثة جداً في تقدير منحنى الانحدار والذي قدم بواسطة Donoho and Jonstone في عام (1994) وما تزال

منطقة توسعه في البحوث جارية، وان أهم الخطوط العريضة لهذه الطريقة هو أنه عادةً ما يتم إفتراض تساوي المسافات بين النقاط (x_1, x_2, \dots, x_n) خلال الفترة $[0, 1]$. حيث أن $x_i = i/n$ و أن n تكون بالشكل $n = 2^j$ ، $J=0,1,2, \dots, n$ ويمكن أن تعرف تقديرات الانحدار المويجي وفق الخطوات التالية [5]:
ليكن لدينا المشاهدات $y_i = (y_1, \dots, y_n)'$ معطاة بالصيغة التالية :

$$y_i = f\left(\frac{i}{n}\right) + \varepsilon_i \quad (7)$$

أو بصيغة المصفوفات

$$\underline{y} = \underline{f} + \underline{\varepsilon} \quad (8)$$

حيث أن $y = (y_1, \dots, y_n)'$ و $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ يمثل التشويش، والهدف هو تقدير الدالة f غير المعلومة حيث أن $f = (f_1, \dots, f_n)'$.

1. حساب قيم معاملات الموجة w بواسطة تطبيق التحويل المويجي المتقطع على البيانات $(y_1, \dots, y_n)'$ وفق الصيغة التالية

$$w = Wy \quad (9)$$

حيث ان W هي مصفوفة التحويل المويجي من درجة $(n \times n)$ لها علاقة بقاعدة الموجة المتعامدة التي يتم اختيارها .

2. نعدل معاملات الموجة التي تم ايجادها من الخطوة (1) وذلك من خلال تمريرها عبر عتبة (Thresholding) ومن ثم نحسب المعاملات المعدلة w^* .

3. اخيراً نجد تقدير الدالة f بواسطة ايجاد معكوس التحويل المويجي المتقطع (IDWT) وفق الصيغة الآتية [4]

$$\hat{f}(x) = W^T w^* \quad (10)$$

2.4 قوانين العتبة

ان الخطوة الثانية من خطوات تقدير دالة الانحدار المويجي هي ازالة التشويش الموجود في الاشارة عن طريق حد العتبة، وباستخدام التحويل المويجي وتحديد الخطوة الثانية منه يتم وضع عتبة ترددية مناسبة بحيث تلغي هذه العتبة معاملات التشويش وتحافظ على معاملات الاشارة الاصلية. [5]

2.4.1 قانون قطع العتبة الناعم (Soft) :

يتم فيها إنهاء القيم ما دون العتبة الى الصفر والمحافظة على القيم الأعلى من العتبة ، وتعرف رياضياً بالعلاقة التالية :

$$S_{\lambda}^S(\hat{w}_{jk}) = \begin{cases} 0 & \text{if } |\hat{w}_{jk}| \leq \lambda \\ \hat{w}_{jk} - \lambda & \text{if } \hat{w}_{jk} > \lambda \\ \hat{w}_{jk} + \lambda & \text{if } \hat{w}_{jk} < -\lambda \end{cases} \quad (11)$$

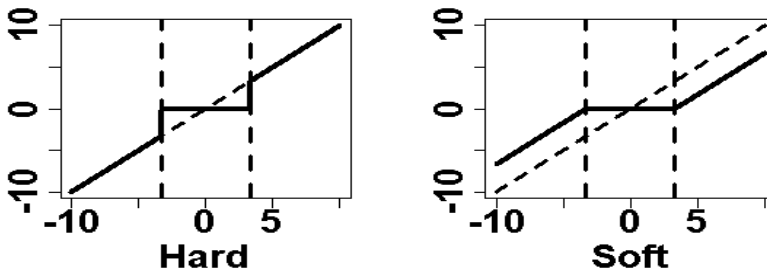
2.4.2 قانون قطع العتبة الصلب (Hard) :

يتم فيها تصغير القيم ما دون العتبة والمحافظة على القيم الأعلى من العتبة ويتم تعريفها رياضياً بالعلاقة التالية :

$$S_{\lambda}^H(\hat{w}_{jk}) = \begin{cases} 0 & \text{if } |\hat{w}_{jk}| \leq \lambda \\ \hat{w}_{jk} & \text{if } |\hat{w}_{jk}| > \lambda \end{cases} \quad (12)$$

حيث أن λ هي قيمة العتبة (Thresholding value) .

مع ان معادلة العتبة الناعمة أعقد من معادلة العتبة الصلبة إلا ان العديد من الباحثين اكدوا من خلال تجاربهم ان نتائج استخدام العتبة الناعمة أفضل من نتائج العتبة الصلبة .



شكل (2) يمثل حد العتبة الصلب والمرن

2.5 قيمة العتبة

كما تم الذكر ان الخطوة الثانية من خطوات التقدير باستخدام الانحدار المويجي هي حد العتبة وان الجزء المهم في هذه الخطوة هو اختيار قيمة العتبة λ ، حيث ان الاختيار المناسب لهذه القيمة يكون حاسماً في دقة التقدير ويعود السبب في ذلك الى ان معاملات الموجة الناتجة من التحويل المتقطع تمرر من خلال حد عتبة وفي هذه الحالة فان القيم الكبيرة جداً للعتبة تؤدي الى ان العديد من المعاملات لن يسمح لها ان تكون ضمن عملية اعادة البناء (التقدير)، مما يؤدي الى منحى ممدد اكثر من اللازم (فوق التمهيد)، وبالعكس اذا كانت قيمة العتبة صغيرة جداً سوف يسمح للكثير من المعاملات ان تكون موجودة في عملية اعادة البناء مما يؤدي الى مقدر متذبذب.

وعليه فان الاختيار المناسب لقيمة العتبة سيلعب دوراً هاماً في جودة المقدر، وفي ادناه شرح موجز لبعض الطرائق المستخدمة في اختيار قيمة العتبة

2.5.1 العتبة الشاملة (Universal Thresholding)

طريقة العتبة الشاملة قدمت من قبل Donoho and Jonstone والمعطاة وفق الصيغة التالية [5]

$$\lambda_{universal} = \sigma \sqrt{2 \log(n)} \quad (13)$$

حيث ان :

n : عدد نقاط البيانات الكلي (مكافأة لعدد معاملات المويجي).

σ : الانحراف المعياري لمستوى التشويش والذي يكون على الأغلب غير معلوم ويمكن استبداله بتقدير حصين هو $\hat{\sigma}$ وهذا التقدير الحصين هو وسيط الانحرافات المطلق (Median Absolute Deviation) لمعاملات المويجي عند المستوى الاول (Finest) ($j = \log(n) - 1$)

$$\hat{\sigma} = \frac{\text{median}(|w_{j-1,k} - \text{median}(w_{j-1,k})|)}{0.6745} \quad (14)$$

2.5.2 العتبة (Sure Thresholding)

هذه الطريقة قدمت بواسطة Donoho and Jonst [5] والتي استندت في اختيار قيمة λ على تقليل تقدير المخاطرة غير المتحيزة لـ (Stein Unbiased Risk Estimation) (SURE) لكل مستوى موجة j حيث اعتبر الباحثان ان معاملات

المويجة عند كل مستوى على حدة كمشكلة تقدير متعدد متغيرات طبيعي مستقل ، حيث اوضح Stein انه ذلك تقدير غير متحيز للمخاطرة هو

$$SURE(\lambda_j, d_{jk}) = N - 2 \sum_{k=1}^N I(|d_{jk}| \leq \lambda_j) + \sum_{k=1}^N \min(|d_{jk}|, \lambda_j)^2 \quad (15)$$

وعليه فان قيمة العتبة للـ (SURE) يمكن ايجادها من الصيغة التالية

$$\lambda_{j,SURE} = \arg \min_{0 \leq \lambda \leq \sqrt{2 \log N}} SURE(\lambda_j, d_{jk}) \quad (16)$$

2.5.3 طريقة التقاطع الشرعية (Two Fold Cross Validation) [11,12]

تعد طريقة تقاطع العبور (C.V) من الطرائق شائعة الاستخدام احصائياً والتي تسمى أيضاً (Leave one-out) ، إلا أن هذه الطريقة لا يمكن تطبيقها مباشرة على طرائق التحويل المويجي السريعة والسبب في ذلك هو اشتراط هذه التحويل أن تكون البيانات ذات حجم دايديكي أي $n=2J$ ، وبالتالي فانها تقوم باستبعاد بعض نقاط البيانات مما يؤدي الى اختلال احد شروط تطبيق التحويل المويجي المتقطع ، ولذلك اقترح Nason طريقة (Two Fold C.V) يقوم من خلالها باستبعاد نصف البيانات حتى يضمن بقاء حجم العينة $n=2J$ ، وسيتم توضيحها وفق الصيغة الآتية :

ليكن لدينا $y_1^0, y_2^0, \dots, y_{n/2}^0$ تمثل نقاط البيانات الفردية و $y_1^E, y_2^E, \dots, y_{n/2}^E$ تمثل نقاط البيانات الزوجي .

وليكن \hat{f}^0, \hat{f}^E يمثل التقديرات المويجية لنقاط البيانات الفردية والزوجية توالياً . باستخدام حذف بيانات المؤشر الفردي، فان البيانات الفردية المشوشة تكون وفق الصيغة الآتية :

$$\tilde{y}_i^0 = \begin{cases} \frac{1}{2}(y_{2i-1} + y_{2i+1}) & , i=1, 2, \dots, n/2-1 \\ \frac{1}{2}(y_{n-1} + y_i) & , i=n/2 \end{cases} \quad (17)$$

أما الصيغة للبيانات المشوشة الزوجية تكون:

$$\tilde{y}_i^E = \begin{cases} \frac{1}{2}(y_{2i-2} + y_{2i}) & , i=2, \dots, n/2 \\ \frac{1}{2}(y_n + y_2) & , i=1 \end{cases} \quad (18)$$

أما التقدير النهائي لطريقة (C.V) للمخاطرة $\mu(\lambda)$ هو

$$\mu(\lambda) = \sum \left\{ \left(f_{\lambda,j}^E \left(\frac{2i}{n} \right) - \tilde{y}_i^0 \right)^2 + \left(\hat{f}_{j,\lambda} \left(\frac{2i-1}{n} \right) - \tilde{y}_i^E \right)^2 \right\} \quad (19)$$

إذا $\lambda_{n/2}$ تقلل من $\mu(\lambda)$ إذن قيمة العتبة النهائية تعطى وفق الصيغة التالية :

$$\lambda_n = \left(1 - \frac{\log 2}{\log n} \right)^{-1/2} \lambda_{n/2} \quad (20)$$

2.6 الانحدار المويجي متعدد الحدود Polynomial Wavelet [15] Regression

لاختزال تأثير مشكلة الحدودية على تقدير دالة الانحدار المويجي اقترح الباحثون Lee & Navean and Oh (2001) طريقة مبسطة تدعى طريقة الانحدار المويجي متعدد الحدود (Polynomial Wavelet Regression) ومختصره (PWR) ، وان الفكرة الاساسية لهذه الطريقة هي الجمع ما بين التقدير المويجي $\hat{f}_w(x)$ مع دالة متعدد حدود ذات درجة قليلة $\hat{f}_p(x)$ والصيغة الرياضية لهذه الطريقة تكون كالآتي :

$$\hat{f}_{pw}(x) = \sum_{n=1}^d \hat{\alpha}_n x^n + \sum_{k=1}^{2^j-1} \hat{c}_0 k \phi_k(x) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \hat{d}_j^s k \psi_{,k}(x) \quad (21)$$

حيث ان الجزء الأول يمثل متعدد الحدود ، اما الجزء الثاني والثالث فيمثلان التقدير المويجي . ويمكن كتابة الصيغة اعلاه بالشكل الآتي :

$$\hat{f}_{pw}(x) = \hat{f}_p(x) + \hat{f}_w(x) \quad (22)$$

حيث ان :

$\hat{f}_{pw}(x)$: هو التقدير النهائي للدالة .

$\hat{f}_p(x)$: تقدير نموذج متعدد الحدود الذي يقوم بإزالة مشكلة اللادورية من مشاهدات y_i .

$\hat{f}_w(x)$: التقدير المويجي للإشارة المتبقية والمتخفية في البواقي $y_i - \hat{f}_p(x)$.

أما درجة المتعدد فهناك طرائق عديدة لتحديد لها ومن هذه الطرائق يتم تحديدها من قبل الباحث اعتماداً على دراسات سابقة.

3. نموذج والية الفقدان

ان النموذج وكما ذكرنا سابقاً هو نموذج انحدار لامعلمي (1) ، وبناءً عليه فان اساس الاستدلال يبدأ بافتراض عينة عشوائية من (x_i, y_i) والتي يمكن اعادة كتابتها في حالة وجود بيانات مفقودة ، حيث ان (x_i) قد تم افتراضها تامة المشاهدة ، واما (y_i) فانها تكون معتمدة على مؤشر δ_i بحيث ان اذا $\delta_i = 0$ فان متغير الاستجابة (y_i) يكون مفقوداً ، و عدا ذلك $\delta_i = 1$.

اما آلية الفقدان فهي الجزء الذي يوضح العلاقة بين احتمالية فقدان القيمة لمتغير ما مع بقية المتغيرات في مجموعة من البيانات.

ويوجد في ادبيات الاحصاء ثلاث انواع من آليات الفقدان شائعة الاستخدام وهي :

1. الفقدان العشوائي التام MCAR .

2. الفقدان العشوائي MAR .

3. الفقدان غير العشوائي NMAR .

وسيتم استخدام آلية الفقدان MAR والذي يتطلب احتمالية الآلية والتي يتم كتابتها وفق الصيغة التالية [9]:

$$p(\delta = 1 / y, x) = p(\delta = 1 / x) = p(x) \quad (23)$$

4. بعض طرائق تعويض البيانات المفقودة

4.1 تعويض المتوسط – الوسيط Mean – Median Imputation

في هذه الطريقة يتم استبدال القيم المفقودة بالمعدل لمتغير الاستجابة، ومن خلال استخدام المتوسط الا انه يتأثر بالقيم المتطرفة في بعض الحالات، وبالتالي يمكن استخدامه بالوسيط .

في حالة معالجة جزء كبير من المشاهدات أي ان نسبة الفقدان عالية اقترح الباحث (2009) Chatterjee اضافة قيم اضافية عن طريق توليد بيانات بشكل عشوائي من التوزيع الطبيعي بمتوسط $(M = Median)$ وتباين صغير $(\sigma^2 = 0.01)$ نوع آخر من تعويض المتوسط يسمى تعويض المتوسط العشوائي والذي يعرف بـ (zero order regression) $(ZOR +)$ ، والذي يمكن استخدامه لملى المشاهدات المفقودة لمتغير

الاستجابة (y_i) وببساطة فان صيغته هي عبارة عن معدل قيم (y_i) مضافاً له قيمة عشوائية مولدة من توزيع طبيعي $(\bar{y} + N(0, \sigma^2))$ وبالتالي ان القيمة التقديرية لكل (y_i) مفقودة تكون وفق الصيغة التالية [14].

$$y_i = \bar{y} + N(0, \sigma^2) \quad (24)$$

حيث ان (σ^2) التباين للبواقي لنموذج الانحدار، انظر (2003) Nittner [14]

4.2 تعويض التمهيد اللامعلمي Non Parametric Bootstrap Imputation

قدمت هذه الطريقة لأول مرة عن طريق Efron (1979) ومنذ ذلك الحين تتالت بحوث كثيرة واكثر تعقيداً حول طرائق التمهيد وبالخصوص حول معالجة مشكلة البيانات المفقودة وفيما يلي الخطوات الاساسية لهذه الطريقة لتقدير قيم متغير الاستجابة المفقودة [1,2].

a. استبدال كل القيم المفقودة لمتغير الاستجابة بقيمة (\bar{y}) .

b. سحب B من العينات الممهدة (Bootstrap Samples) المستقلة وحساب المتوسط لكل عتبة وليكن $(\bar{y}_i \quad i = 1, 2, \dots, B)$.

c. حساب متوسط التمهيد الكلي والتباين كالاتي

$$\bar{y}_b = (\sum_{i=1}^B \bar{y}_i) / B, S_b^2 = \frac{1}{B-1} \sum_{i=1}^B (\bar{y}_i - \bar{y}_b)^2 \quad (25)$$

d. نولد m من نقاط البيانات عشوائياً من التوزيع الطبيعي بمتوسط $(M = \bar{y}_b)$ وتباين $(\sigma^2 = S_b^2)$ لاستبدال m من المشاهدات المفقودة.

4.3 طريقة التقدير الحصينة المضاعفة اللامعلمية نوع (HT)

قام الباحث (2012) P. Cheng باقتراح مقدر لامعلمي حصين لمعالجة القيم المفقودة في متغير الاستجابة y_i عن طريق تعديل مقدر (HT) من خلال اضافة بواقي الانحدار الموزونة نوع (HT) الى دالة الانحدار المقدر بالطرائق اللبية ، ويمكن كتابة هذا المقدر وفق الصيغة التالية [13]

$$m_{DR}(x) = \hat{m}(x_i) + \frac{\delta_i(y_i - \hat{m}(x_i))}{w_i} \quad (26)$$

حيث ان

$$\hat{m}(x_i) = \sum_{j=1}^n w_h(x_i, x_j) \delta_j y_j / w_h(x_i, x_j) \delta_j \quad (27)$$

حيث ان w هي دالة كثافة احتمالية وان

$$w_h(u, x) = h^{-1} w((u - x) / h) \quad (28)$$

حيث ان h تمثل معلمة التمهيد ، اما الحد الثاني من المعادلة فما هو الا صيغة (HT) باستثناء استبدال y_i لبواقي الانحدار حيث صيغة (HT) هي

$$HT = \frac{y_i \delta_i}{w_i} \quad (29)$$

حيث يتم تقدير قيم y_i المفقودة وفق الصيغة التالية المقترحة من قبل الباحث Cheng [13,3]

$$y_i^* = \delta_i y_i + (1 - \delta_i) \hat{m}_{DR}(x) \quad (30)$$

5. طرائق التقدير

5.1 طريقة : Bootstrap-Iterative Polynomial Wavelet

ان هذه الطريقة مقدمة من قبل الباحثان A. M. Taher & M. T. Ismail (2012) ، وان الفكرة الاساسية لهذه الطريقة هي اعتماد تنبأ متعدد الحدود المكرر في استبدال القيم المفقودة مع ما يقاربها من قيم متوقعة من ذلك النموذج مضافاً اليها خطأ عشوائي ، وتتلخص خطوات هذه الطريقة وفق الآتي :

1. نبدأ اولاً بتقدير اولي للبيانات المفقودة مستخدمين طريقة التمهيد اللامعلمي والموضحة في (2-8-2) [2,1,15] :
2. نجد القيم التقديرية لـ f باستخدام نموذج الدراسة وهو نموذج متعدد الحدود ذات درجة ثانية .

$$\hat{f} = \hat{y}_j^c \quad (31)$$

يتم ذلك طبقاً للبيانات التامة التي تم الحصول عليها من الخطوة (1) حيث ان

$$\hat{y}_j^c = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2 \quad (32)$$

حيث تم تقدير معالم النموذج أعلاه باستخدام المربعات الصغرى .

3. نستبدل القيم المقدرة مع القيم المتوقعة من نموذج متعدد الحدود حيث ان

$$y_j = \hat{y}_j^c + N(\mu = 0, \sigma^2 = MSE) \quad (33)$$

MSE هو لنموذج متعدد الحدود الذي تم حسابه في ثانية .

4. لـ $j = 1, 2, 3, \dots$ نكرر الخطوتين السابقتين حتى نصل الى التقارب بالنسبة MSE لـ

5. من اجل تقريب التقدير النهائي للقيم المفقودة سيكون تقدير القيم المفقودة النهائي والمقصود هنا هو آخر تكرار + قيمة عشوائية مولدة من توزيع طبيعي بمتوسط (صفر) وتباين مساوي الى MSE_j أي متوسط مربعات الخطأ لنموذج متعدد الحدود من آخر تكرار.

اصبح الآن لدينا بيانات تامة، وعليه سيتم استخدام طرائق تقدير المربعات الصغرى على نموذج متعدد الحدود قيد الدراسة. وبعد ذلك نقوم بتقدير نموذج متعدد الحدود من الدرجة الثانية للحصول على $\hat{fp}(x)$ باستخدام طريقة المربعات الصغرى ، وبعدها نجد البواقي عن طريق تطبيق الصيغة التالية:

$$e_i = y_i - \hat{fp}(x) \quad (34)$$

وبعد ايجاد البواقي نجد دالة الانحدار المويجي $\hat{fw}(x)$ ، ومن ثم نجد التقدير النهائي لدالة الانحدار المويجي متعدد الحدود .

$$\hat{fpw}(x) = \hat{fp}(x) + \hat{fw}(x) \quad (35)$$

5.2 الطريقة المقترحة Doubly Robust Polynomial Wavelet

ان فكرة هذه الطريقة مستمدة من فكرة الباحثان T. C. M. Lee & X. L. Meng [8] والباحث Dohon Kin [7] ، حيث قام كلاً من الباحثين باستخدام طرائق تعويض تقليدية كخطوة اولى مثل طرائق التعويض باستخدام الوسيط او المتوسط وطريقة EM او التعويض المتعدد ومن ثم تطبيق طريقة الانحدار المويجي كخطوة ثانية بعد ان اصبحت البيانات تامة وملائمة لهذه الطريقة ، الا اننا استخدمنا طرائق تعويض اخرى تمثلت بتوظيف طريقة الباحثان Cheng & Ning وفكرة هذه الطريقة انهما قاما بتعديل طريقة التقدير اللامعلمي في تعويض البيانات المفقودة Horvitz-

Thompson (HT) الى مقدر جديد أطلق عليه اسم مقدر الحصين المضاعف (DR) اللامعلمي والموضحة في (2-8-3) لمعالجة مشكلة البيانات المفقودة في متغير الاستجابة كخطوة اولى بغية جعل البيانات ملائمة لتطبيق طرائق الانحدار المويجي ، اما في الخطوة الثانية فيتم تطبيق طرائق الانحدار المويجي متعدد الحدود وان هذه الطريقة قد تم تسميتها بطريقة (DRPW) حيث انه بعد معالجة القيم المفقودة باستخدام الصيغة الموضحة في المعادلة (30)، حيث ان $\hat{m}DR$ يتم الحصول عليها من المعادلة (27) وبعد ذلك نطبق الخطوة الثانية من التقدير والتي تتلخص بايجاد دالة متعدد الحدود $\hat{f}p(x)$ وذلك عن طريق تقدير معالم هذا النموذج ذات الدرجة الثانية باستخدام طريقة المربعات الصغرى ، حيث ان

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i + \hat{b}_2 x_i^2 \quad (36)$$

$$\hat{y}_i = \hat{f}p(x) \quad \text{حيث ان}$$

بعد ذلك نجد البواقي عن طريق الصيغة التالية :

$$e_i = \hat{y}_i - \hat{f}p(x) \quad (37)$$

واخيراً نطبق الانحدار المويجي التقليدي الموضح في (2-3) على البواقي للحصول على التقدير النهائي لدالة الانحدار

$$\hat{f}drpw(x) = \hat{f}w(x) + \hat{f}p(x) \quad (38)$$

6. المحاكاة

ان تطبيق ما جاء في الجانب النظري يتم عن طريق استخدام اسلوب المحاكاة (Simulation) من اجل محاكاة اكبر قدر من الحالات التي يمكن ان تواجهنا في الواقع العملي بغية الوصول الى نتائج اكثر عمومية ، وان اللجوء الى استخدام اسلوب المحاكاة كان لعدة اسباب اهمها ان هذا الاسلوب يوفر لنا اختصاراً في الوقت والكلفة بجميع اشكالها المادية والبشرية الذي تتطلبه التجارب الواقعية كالتجارب الطبية والفلكية وغيرها من التجارب التي تحتاج الى وقت كبير وتكلفة باهضة جداً.

6.1 توليد المتغيرات

لانموذج الانحدار اللامعلمي في المعادلة (1) حيث ان $x_i = i/n$ ، $i = 1, \dots, n$ ، حيث x_i ذات مسافات متساوية ضمن الفترة $[0,1]$ وان e_i يتوزع توزيع طبيعي بمتوسط صفر وتباين ثابت σ ، $f(x_i)$ دالة الانحدار التي يراد تقديرها في ظل وجود فقدان في قيم متغير الاستجابة y ، وعليه لتوليد البيانات لـ (x_i, y_i) وفق

الانموذج الانحدار اللامعلمي (1) مع دوال الاختبار الموضحة في (3-4)، حيث ان المتغير المعتمد يتم توليده من خلال دوال الاختبار الموضحة في (3-4) مضافاً اليه حد الخطأ e_i .

ولتنفيذ تجارب المحاكاة جرى استخدام مستويات مختلفة من العوامل الآتية :

1. حجوم العينات n ، حيث تم استخدام ثلاثة حجوم للعينات وهي $2^6 = 64$ ، $2^7 = 128$ ، $2^8 = 256$ ، كون حجم العينة هنا يجب ان يكون $n = 2^j$ ، حيث ان J عدد صحيح موجب .
2. نسب الاشارة الى التشويش (SNR) حيث تم استخدام اثنان من نسب التشويش $SNR = 5, 10$.
3. دوال الاختبار $f(x_i)$ ، حيث جرى استخدام ثلاثة دوال اختبار مختلفة موضحة في (3-4).
4. نسب الفقدان، حيث تم استخدام ثلاثة نسب للفقدان وهي (15% , 25%) .
5. درجة متعدد الحدود من الدرجة الثانية أي $d = 2$.

6.2 دوال الاختبار Test Function [11]

وتتميز هذه الدوال كونها دوال اختبار قياسية ونموذجية ومناسبة لاستخدامها في تجارب المحاكاة كونها صممت لتعرض مجموعة من الظواهر التي غالباً ما تحدث في مجموعة البيانات المأخوذة من الواقع العملي وان هذه الدوال تكون معرفة على الفترة [0,1] وسوف نعرض تلك الدوال [5] وكالاتي :

1. دالة Doppler

$$f_1(x) = \{x(1-x)\}^{1/2} \sin\{2\pi(1+\varepsilon)/(x+\varepsilon)\}, \varepsilon = 0.05 \quad (39)$$

2. دالة Heavisine

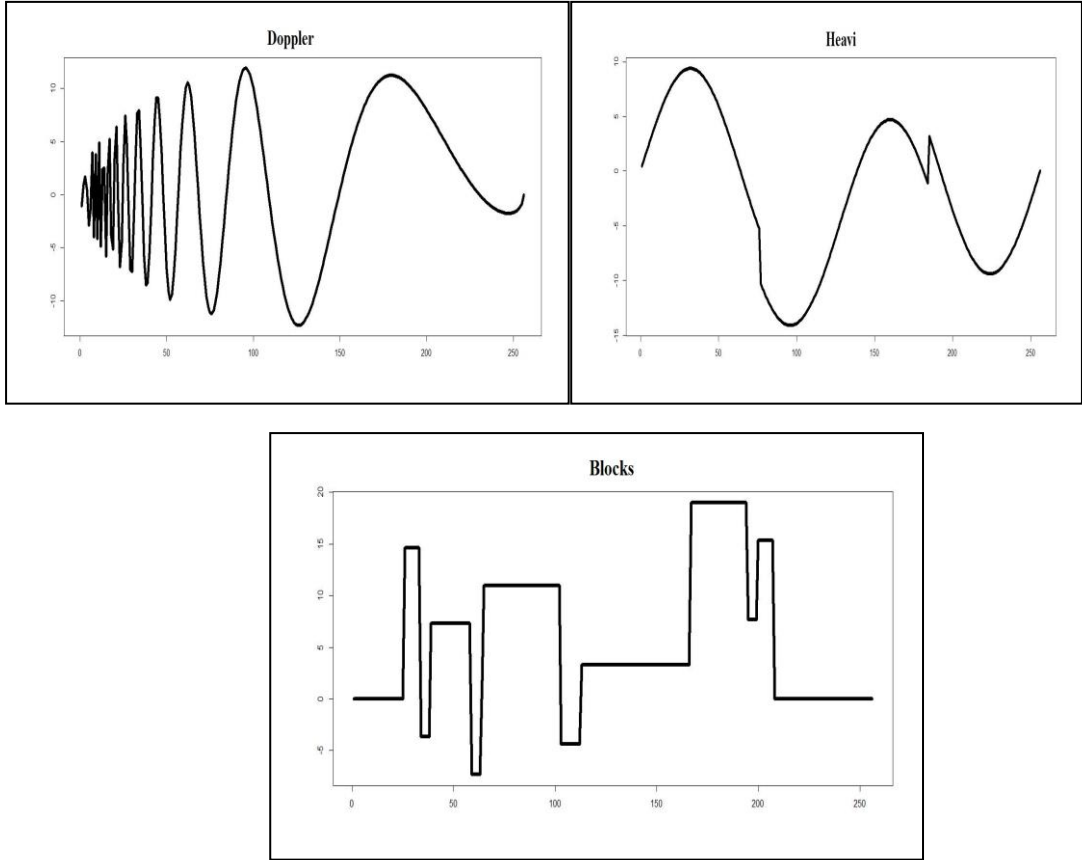
$$f_2(x) = 4 \sin 4\pi x - \text{sgn}(x - 0.3) - \text{sgn}(0.72 - x) \quad (40)$$

3. دالة Blocks

$$f_3(x) = \sum h_j k(x - x_j), \quad k(x) = \{1 + \text{sgn}(x)\} / 2 \quad (41)$$

$$(x_j) = (0.1, 0.13, 0.15, 0.23, 0.25, 0.40, 0.44, 0.65, 0.76, 0.78, 0.81)$$

$$(h_j) = (4, -5, -4, 5, -4.2, 2.1, 4.3, -3.1, 2.1, -4.2)$$



شكل (3) : دوال الاختبار

6.3 تجارب المحاكاة

هنا سيتم وصف وافي لتجارب المحاكاة بحسب دوال الاختبار المستخدمة في نموذج الانحدار اللامعلمي (1) لتوليد البيانات (x_i, y_i) من ثم اجريت تجارب المحاكاة لتقدير الانموذج من خلال توظيف عدد من طرائق معالجة القيم المفقودة في متغير الاستجابة ولحجوم عينات ونسب اشارة مختلفة، حيث تم توظيف طرائق المستخدمة في التقدير وهي ستة طرائق تم تكرار التجربة (500) بثبات جميع العوامل عدا المتغير العشوائي والذي يعاد توليده عند تكرار كل تجربة .

في كل مرة يجري فيها توليد للبيانات المشوشة ، يتم استخدام آلية فقدان العشوائي MAR لفقدان بعض مشاهدات متغير الاستجابة عشوائية وفق الصيغة الموضحة في المعادلة (23).

أولاً: تطبيق طريقة (BIPW): ان ملخص عمل هذه الطريقة والموضحة في الخوارزمية (1) (ملحق رقم 1) يكون وفق الخطوات التالية :

1. بعد توليد كلاً من المتغير التوضيحي والمتغير المعتمد الموضحة في (4-1) و اجراء عملية فقدان العشوائي على متغير الاستجابة وفق الصيغة الموضحة في المعادلة (23) يتم معالجة قيم y_i المفقودة باستخدام المعادلة (33) .

2. بعد ان تم معالجة البيانات المفقودة لمتغير الاستجابة نجد دالة نموذج متعدد الحدود $\hat{fp}(x)$ من الدرجة الثانية عن طريق تقدير معاملات النموذج باستخدام المربعات الصغرى.

3. نجد البواقي عن طريق الصيغة الموضحة في المعادلة (34).

4. نطبق الانحدار المويجي الموضح في (2-3) على البواقي.

5. نجد التقدير النهائي لدالة الانحدار المويجي متعدد الحدود وفق الصيغة الموضحة في المعادلة (22) وكذلك الحال هنا سيتم الحصول على ثلاث تقديرات لدالة الانحدار المويجي متعدد الحدود بحسب نوع قيمة العتبة كما موضح في الخوارزمية.

ثانياً: تطبيق طريقة (DRPW) : ان ملخص عمل هذه الطريقة والموضحة في الخوارزمية (2) (ملحق رقم 2) يكون وفق الخطوات التالية :

1. تطبيق طريقة تعويض الانحدار اللامعلمي باستخدام طريقة (DR)، حيث يتم تطبيق الصيغة الموضحة في المعادلة (30) للحصول على قيم متغير الاستجابة المفقودة.

2. نجد قيمة $\hat{fp}(x)$ لنموذج متعدد الحدود من الدرجة الثانية عن طريق تقدير معاملاته باستخدام المربعات الصغرى.

3. نجد البواقي عن طريق الصيغة الموضحة في المعادلة (37).

4. نطبق الانحدار المويجي على البواقي وخطواته نفس ماتم في الطريقة الاولى لغاية الحصول على التقدير النهائي، الجدير بالذكر انه عند استخدام طريقة

(DRPW) الموضحة في (2-3)، وقد تم استخدام دالة (Epenchenkov)،
 واما معلمة التمهيد فهي طريقة (Plug-in dircet).

6.4 تحليل النتائج

1. دالة الاختبار $f_1(x)$

بصورة عامة نلاحظ انه باختلاف كلاً من نسب فقدان ونسب التشويش ($SNR=1$) وقيم العتبة وحجوم العينات تفوق التقدير المقترح DRPW على بقية التقديرات.

2. دالة الاختبار $f_2(x)$

بصورة عامة وعند نسبة فقدان 15% وحجم عينة 64 ونسبة تشويش $SNR=5$ نلاحظ تقارب اداء التقديرات وباختلاف قيم العتبة مع تفوق بسيط للتقدير المقترح DRPW، اما بقية احجام العينة فنلاحظ تفوق تقدير BIPW. اما عند نسبة تشويش $SNR=10$ نلاحظ تفوق التقدير المقترح DRPW وباختلاف قيم العتبة وحجوم العينات.

اما عند نسبة فقدان 25% وحجم عينة 64 فنلاحظ تفوق التقدير المقترح DRPW عند $SNR=10$ وقيمة عتبة SURE, CV، بينما يتفوق تقدير BIPW عند نسبة تشويش $SNR=5$ وقيمة عتبة CV, Universal، اما عند حجم (128، 256) فنلاحظ تفوق تقدير BIPW على بقية التقديرات وباختلاف نسب التشويش وقيم العتبة.

3. دالة الاختبار $f_3(x)$

بصورة عامة وعند نسبة فقدان 15% وحجوم عينة (64، 128) نلاحظ تفوق التقدير المقترح DRPW وباختلاف قيم العتبة، بينما عند حجم عينة 256 فانه باستثناء التقدير BIPW عند قيمة عتبة SURE ونسبة تشويش $SNR=10$ نلاحظ تفوق التقدير المقترح DRPW. اما عند نسبة فقدان 25% نلاحظ تفوق التقدير المقترح DRPW وباختلاف قيم العتبة وحجوم العينات ونسب التشويش.

SNR^1 (نسبة الاشارة الى التشويش) : هي مقياس يتم بواسطته المقارنة بين قيمة الاشارة وقيمة التشويش (Noise) المحمولة معها او بتعريف آخر هي النسبة بين قيمة الاشارة الى قيمة ما تحويه من تشويش ويتم حسابها وفق الصيغة الآتية: $SNR = \frac{\sigma_{signal}}{\sigma_{noise}}$

جدول (1)

معييار MSE لمقارنة التقديرات لدالة Doppler المشوشة ونسبة فقدان 15% لحجوم عينات n=256, n=128, n=64 ونسب إشارة إلى تشويش SNR=10, SNR=5

	Sure		Universal		CV	
	SNR=5	SNR=10	SNR=5	SNR=10	SNR=5	SNR=10
n=64						
BIPW	0.01290761	0.008635101	0.01304060	0.009021309	0.01260377	0.009279891
DRPW	0.009156261	0.007350796	0.009156261	0.007350796	0.009208967	0.007231386
n=128						
BIPW	0.01129779	0.008880039	0.01506875	0.010206173	0.01503667	0.010225460
DRPW	0.007141616	0.004305472	0.007141616	0.004888565	0.006582957	0.004586508
n=256						
BIPW	0.009576080	0.010070339	0.011462096	0.009019977	0.009761725	0.008565829
DRPW	0.006971813	0.005905501	0.006971813	0.005178063	0.005574153	0.005111043

جدول (2)

معييار MSE لمقارنة التقديرات لدالة Doppler المشوشة ونسبة فقدان 25% لحجوم عينات n=256, n=128, n=64 ونسب إشارة إلى تشويش NR=10

	Sure		Universal		CV	
	SNR=5	SNR=10	SNR=5	SNR=10	SNR=5	SNR=10
n=64						
BIPW	0.01767312	0.01347703	0.01758670	0.01307456	0.01763996	0.01244311
DRPW	0.009824934	0.008596772	0.009824934	0.008596772	0.009854863	0.009106415
n=128						
BIPW	0.01866458	0.01637375	0.01903258	0.01736255	0.01624562	0.01565867
DRPW	0.009872897	0.009842858	0.009872897	0.009842858	0.010103937	0.009952577
n=256						
BIPW	0.01612434	0.01282049	0.01700527	0.01421086	0.01512836	0.01181037
DRPW	0.009032292	0.006946507	0.009032292	0.006553716	0.008950463	0.006296763

جدول (3)

معياري MSE لمقارنة التقديرات لدالة Heavisine المشوشة ونسبة فقدان 15% لحجوم عينات $n=256$, $n=128$, $n=64$ ونسب إشارة إلى تشويش $SNR=10$, $SNR=5$

	Sure		Universal		CV	
	SNR=5	SNR=10	SNR=5	SNR=10	SNR=5	SNR=10
$n=64$						
BIPW	0.007346267	0.004682865	0.006613479	0.003749266	0.006507761	0.003673009
DRPW	0.005214697	0.002079536	0.005214697	0.002079536	0.005798040	0.002679815
$n=128$						
BIPW	0.005223748	0.006005323	0.004902374	0.004068223	0.004837363	0.003968068
DRPW	0.005138795	0.005807593	0.006826065	0.005404053	0.007142345	0.005523906
$n=256$						
BIPW	0.004156575	0.009517706	0.004021926	0.005092062	0.004002106	0.004997047
DRPW	0.005053950	0.007468338	0.005053950	0.005673272	0.004151547	0.005793894

جدول (4)

معياري MSE لمقارنة التقديرات لدالة Heavisine المشوشة ونسبة فقدان 25% لحجوم عينات $n=256$, $n=128$, $n=64$ ونسب إشارة إلى تشويش $SNR=10$, $SNR=5$

	Sure		Universal		CV	
	SNR=5	SNR=10	SNR=5	SNR=10	SNR=5	SNR=10
$n=64$						
BIPW	0.009705328	0.006967265	0.008793474	0.006280820	0.008679318	0.006173189
DRPW	0.008859227	0.004187244	0.008859227	0.004187244	0.009248620	0.004149591
$n=128$						
BIPW	0.005793481	0.007156381	0.005397311	0.006547807	0.005341070	0.006486727
DRPW	0.00972677	0.007251965	0.01008933	0.007867633	0.01023063	0.008053544
$n=256$						
BIPW	0.007349494	0.009452139	0.007349494	0.005423241	0.007355703	0.005352997
DRPW	0.008925085	0.009076386	0.008925085	0.007066985	0.008469894	0.007051746

جدول (5)

معيار MSE لمقارنة التقديرات لدالة Blocks المشوشة ونسبة فقدان 15% لحجوم عينات
 SNR=10, SNR=5 ونسب إشارة إلى تشويش n=256, n=128, n=64

	Sure		Universal		CV	
	SNR=5	SNR=10	SNR=5	SNR=10	SNR=5	SNR=10
n=64						
BIPW	0.02613088	0.02097905	0.02614797	0.02170400	0.02625743	0.02024140
DRPW	0.01355233	0.004761464	0.01355233	0.009880912	0.01397248	0.010273640
n=128						
BIPW	0.02120379	0.01774562	0.02153016	0.01886903	0.01739452	0.01882397
DRPW	0.006346031	0.008306738	0.009050161	0.009455502	0.009424994	0.009671268
n=256						
BIPW	0.01667140	0.009538898	0.01862466	0.013101437	0.01332884	0.010298473
DRPW	0.006889723	0.01184778	0.008926612	0.01030538	0.007922947	0.01028037

جدول (6)

معيار MSE لمقارنة التقديرات لدالة Blocks المشوشة ونسبة فقدان 25% لحجوم عينات
 SNR=10, SNR=5 ونسب إشارة إلى تشويش n=256, n=128, n=64

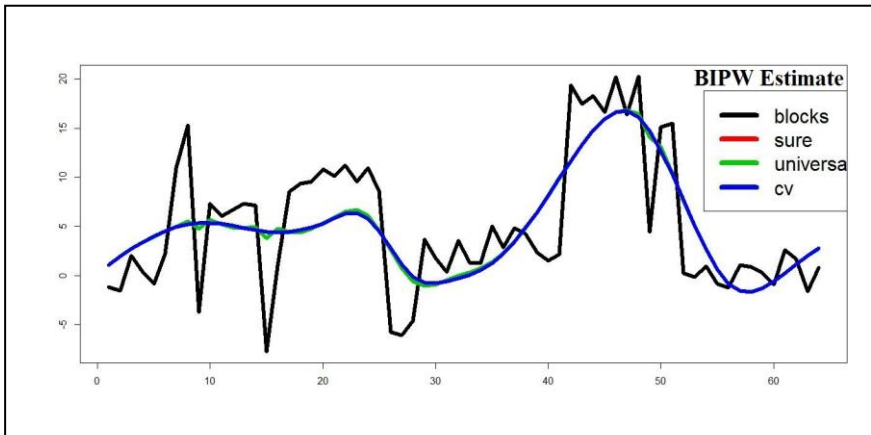
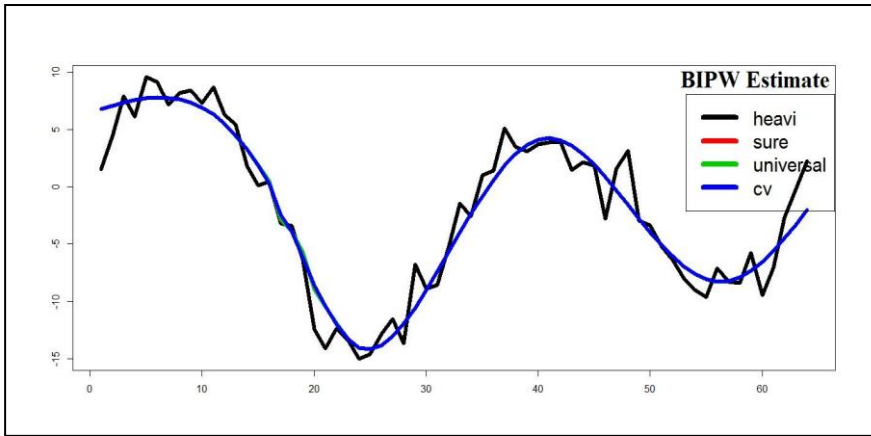
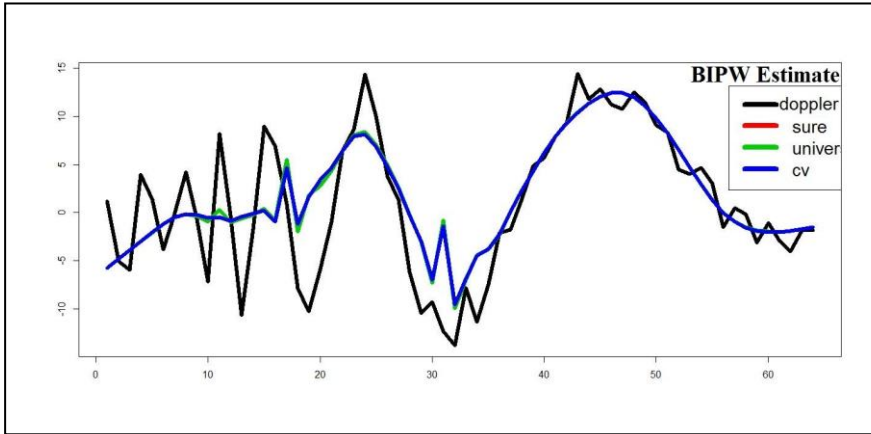
	Sure		Universal		CV	
	SNR=5	SNR=10	SNR=5	SNR=10	SNR=5	SNR=10
n=64						
BIPW	0.02542030	0.02427281	0.02578266	0.02435907	0.02534633	0.02393166
DRPW	0.01425165	0.01884875	0.01425165	0.01536147	0.01427282	0.01544421
n=128						
BIPW	0.02583084	0.02292590	0.02586177	0.02328932	0.02569930	0.02028087
DRPW	0.01629328	0.02085763	0.01608421	0.01819619	0.01606907	0.01827777
n=256						
BIPW	0.02052275	0.01637269	0.02098447	0.01790606	0.01485425	0.01403207
DRPW	0.01727069	0.01528143	0.01421087	0.01181023	0.01409739	0.01188012

7. الاستنتاجات

1. تفوق التقدير المقترح DRPW عند استخدام دالة الاختبار Doppler وباختلاف نسب التشويش والفقدان وحجوم العينة.
2. تفوق التقدير BIPW عند نسبة تشويش $SNR=5$ وحجوم عينات (128، 256)، بينما يتقارب اداء التقديرات عند حجم عينة 64، اما عند $SNR=10$ تفوق التقدير المقترح DRPW.
3. تفوق التقدير BIPW عند حجوم العينة (128، 256) ونسبة فقدان 25% ودالة اختبار Heavesine ، بينما يتقارب الاداء عند حجم عينة 64 .
4. عند دالة الاختبار Blocks نلاحظ تقارب الاداء عند نسبة فقدان 15%، بينما يتفوق التقدير المقترح DRPW عند نسبة فقدان 25% وباختلاف نسب التشويش وقيم العتبة وحجوم العينة.
5. تتناقص قيمة MSE بازدياد نسب التشويش وكذلك عند زيادة حجم العينة.
6. عند مقارنة قيمة العتبة لكل تقدير على حدة نلاحظ انه افضل اداء عند استخدام قيمة عتبة CV يليها قيمة عتبة SURE.

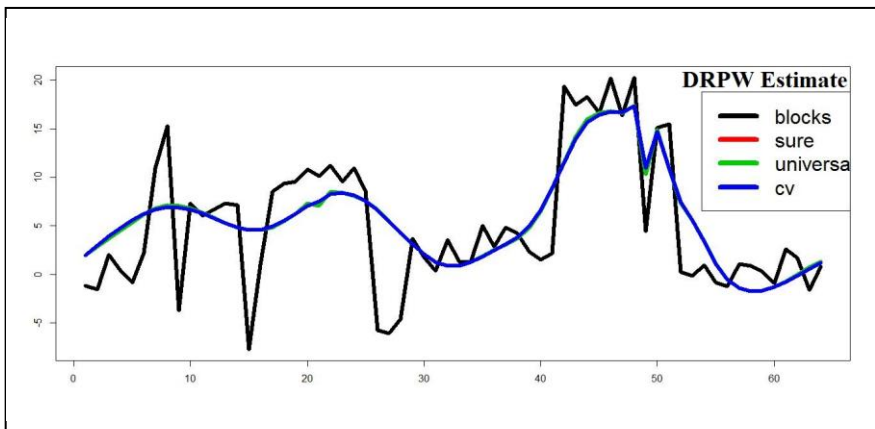
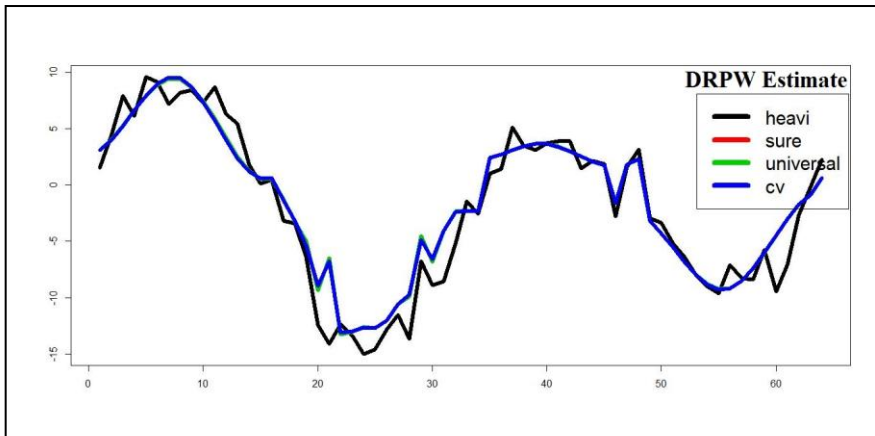
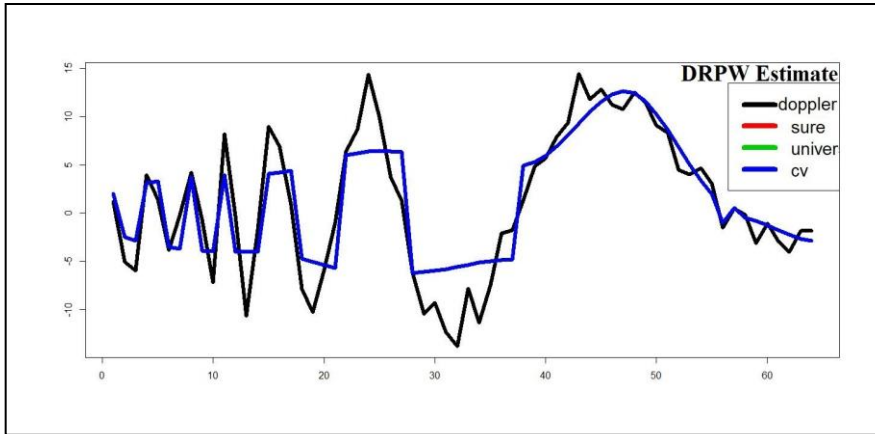
8. التوصيات

1. توظيف الطرائق المقترحة واستخدامها في معالجة حالي الفقدان والبيانات الشاذة في آن واحد.
2. التركيز على دراسة تأثير قيمة العتبة بشكل موسع لما لها من تأثير على اداء الطرائق المويجية في التقدير والتركيز على دراسة العتبات الحصينة.

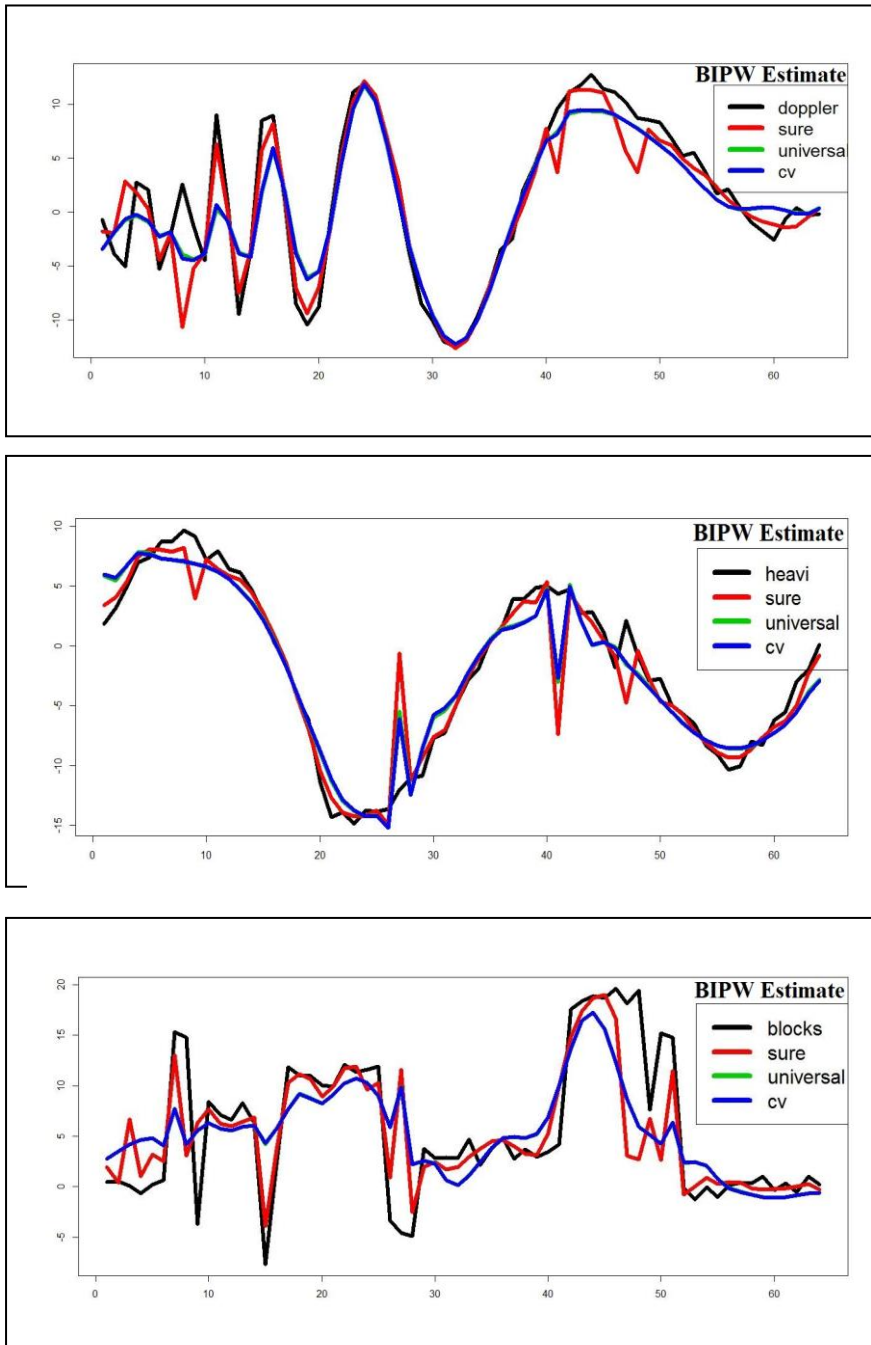


شكل (4)

نتائج تجارب المحاكاة باستخدام تقديرات الدوال (Doppler, Heavi, Blocks) المشوشة عند انواع قيم عتبة مختلفة ونسبة فقدان 15% ونسبة تشويش 5 وحجم عينة 64

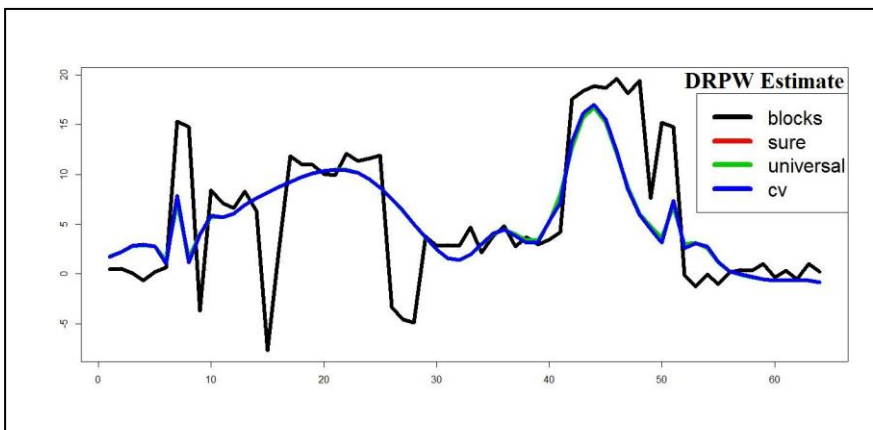
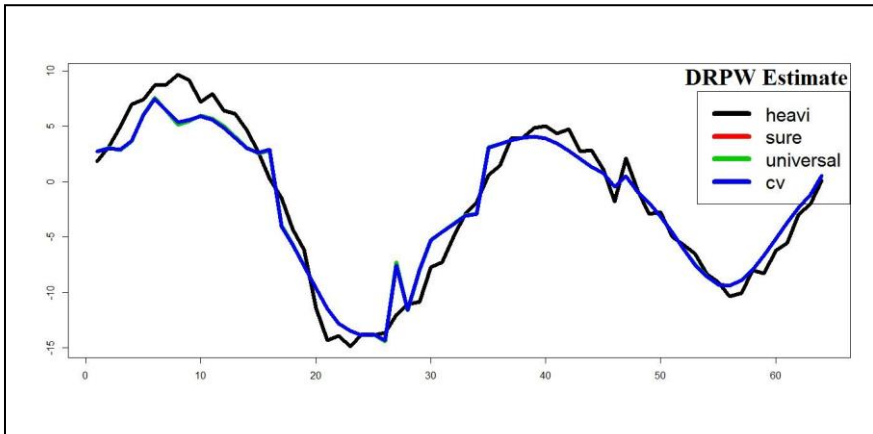
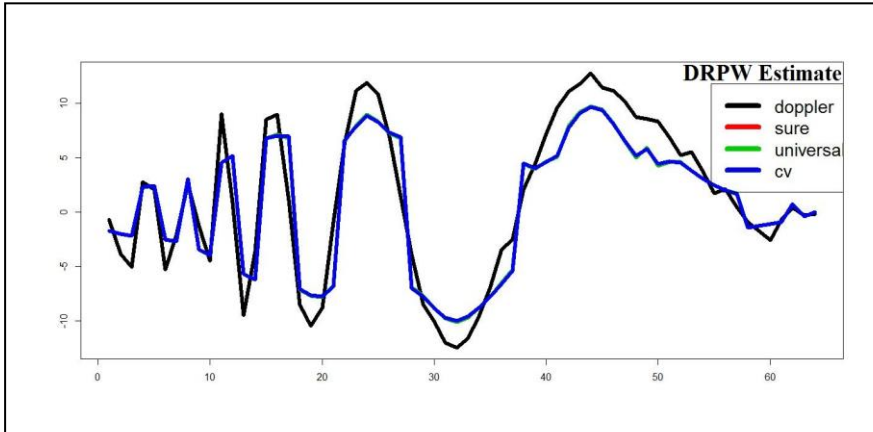


تتمة شكل (4)



شكل (5)

نتائج تجارب المحاكاة باستخدام تقديرات الدالة (Doppler, Heavi, Blocks) المشوشة عند انواع قيم عتبة مختلفة ونسبة فقدان 15% ونسبة تشويش 10 وحجم عينة 64



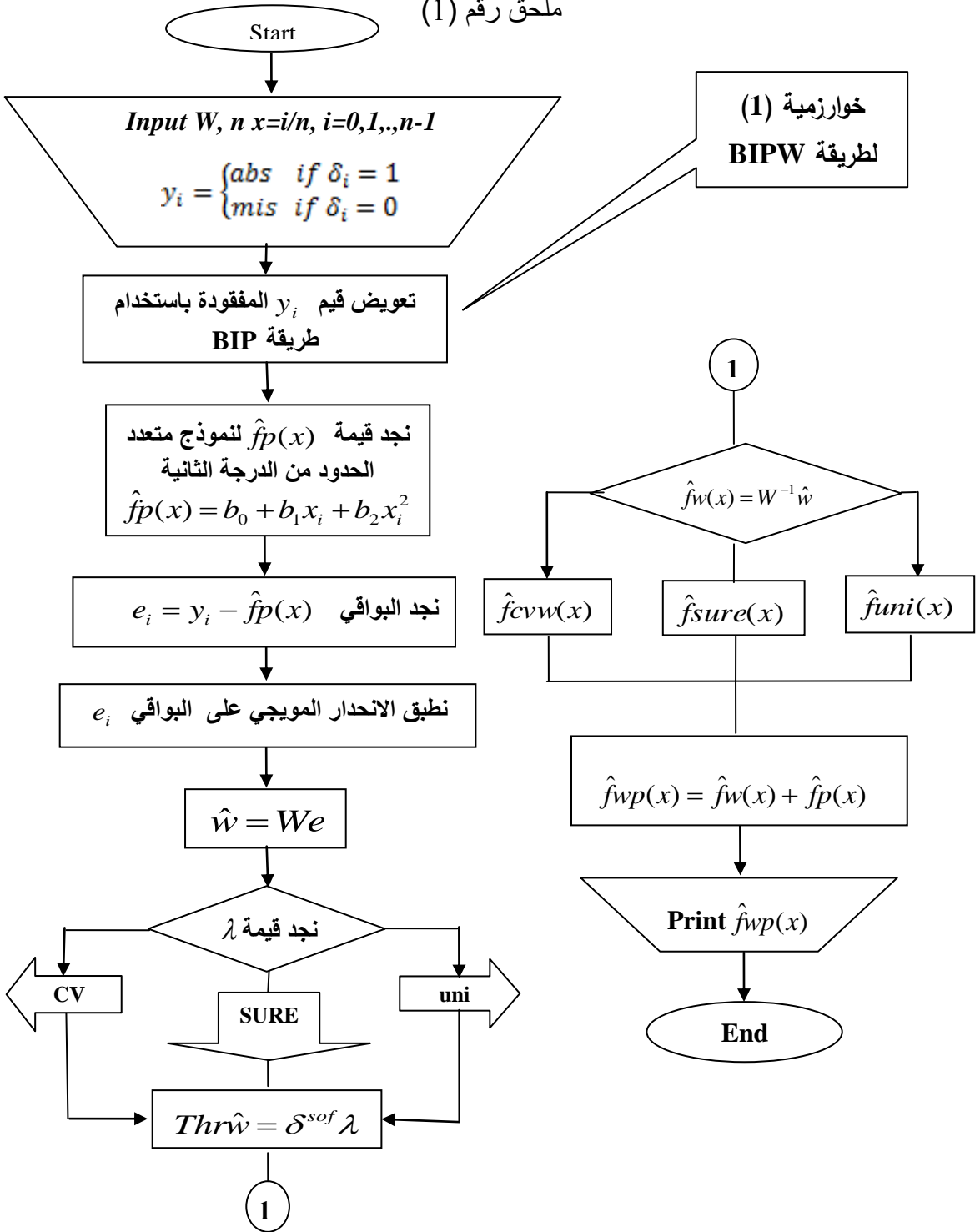
تتمة شكل (5)

المصادر

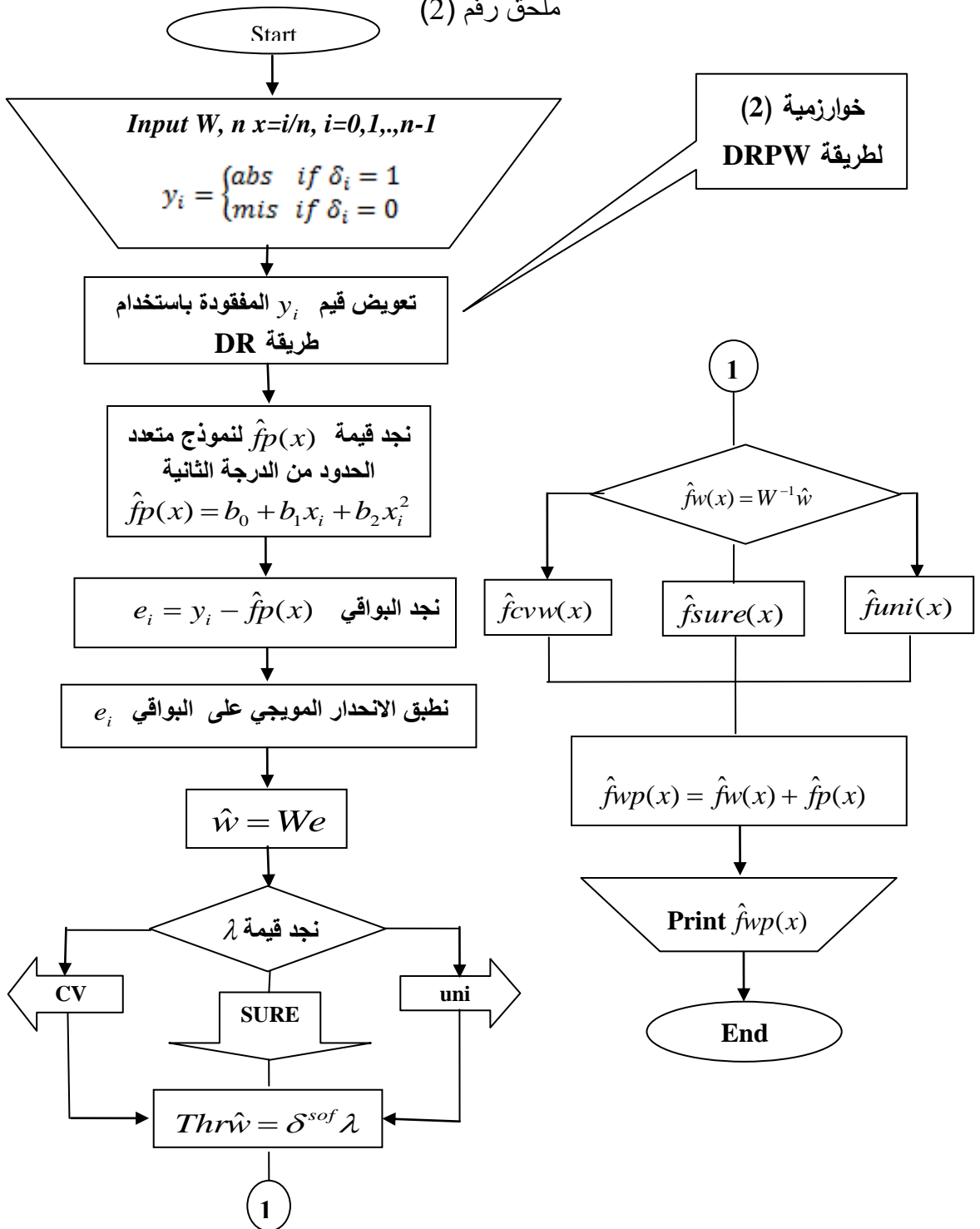
- [1] Altaher, M. A. , Ismail, T. M. , (2011), “A New Method on Treating Missing Values in Polynomial Wavelet Regression”, Proceedings of the Annual International Conference on Operations Research and Statistics (ORS), copyright © GSTF & ORS , ISBN : 978-981-08-8407-9 .
- [2] Altaher, M.A., (2012), “Local Polynomial Wavelet Regression with Missing at Random”, Applied Mathematical Sciences, Vol. 6, no. 57, 2805-2819.
- [3] Cheng, E. P. , (1994), “Nonparametric Estimation of Mean Functional with Data Missing at Random”, Journal of the American Statistical Association, Vol. 89, No. 425 .
- [4] Daubechies, I. , (1992), “Ten Lectures on Wavelets”, CBMS-NSF regional conference series in applied mathematics .
- [5] Donoho, L. D. , Johnstone, M. I. , (1994), “Ideal spatial adaptation by wavelet shrinkage”, Biometrika, 81, 3, pp. 425-55.
- [6] Hardle, W. , (1990), “Applied Nonparametric Regression”, Gambridg – MA : Cambridg University Press .
- [7] Kim, D. , Lee, Y. and Oh, S. H. , (2006), “Hierarchical-Likelihood-Based Wavelet Method for Denoising Signals With Missing Data”, IEEE SIGNAL PROCESSING LETTERS, VOL. 13, NO. 6 .
- [8] Lee, M. C. T. , Meng, L. X. , (2007), “Self Consistency : A General Recipe for Wavelet Estimation With Irregularly-spaced and / or Incomplete Data”, arXiv : math / 0701196v1 [math. ST].
- [9] Little, A. J. R. , Rubin, B. D. , (2002), “Statistical Analysis with Missing Data”, John Wiley & Sons, INC.
- [10] Mallat, G. S. , (1989), “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation ”, Ieee Transactions on Pattern Analysis and Machine Intelligence. Vol. 11, No. 7.
- [11] Nason,G.B.,(2008),“Wavelet Methods in Statistics with R”, Springer.

- [12] Nason, G.P., (1996), “ Wavelet shrinkage using cross-validation ” , Journal of the Royal Statistical Society Series B. 58, 463-479 .
- [13] Ning, J. , Cheng, E. P. , (2012) “A Comparison Study of Nonparametric Imputation Methods”, Statistical Compute, Springer, PP. 273-285.
- [14] Nittner, T. , (2003), “Missing at Random (MAR) in Nonparametric Regression” .
- [15] Oh, S. H. , Naveau, P. , Lee, G. , (2001), “Polynomial Boundary Treatment for Wavelet Regression”, Biometrika, 88, 1, pp. 291-298 .
- [16] Vidakovic, B. , (1999), “Statistical Modeling by Wavelet”, John Wiley & Sons, Inc .

ملحق رقم (1)



ملحق رقم (2)



خوارزمية (2)
لطريقة DRPW

Using Different Threshold Value in Comparison Some of Methods Wavelet Estimation for Non Parametric Regression Function with Missing Data

Prof. Dr. Dhafir H. Rasheed

rasheddhafir@yahoo.com

Baghdad University – College of Administration and
Economics

Saad K. Hamza

sdkadem@yahoo.com

Baghdad University - Administration Affairs

Abstract: *The problem of missing of some of sample observations is one of the main problems that face researcher during the statistical analysis , the main problem of missing data are as follows damage , negligence, death and morbidity as in the case of clinical studies*

The presence of such a problem within the data may influence on the analysis and accordingly it may lead to misleading conclusions despite the fact that the wavelet estimations are of high efficiency in estimating the regression function , but it may be influenced by the problem of missing data , in addition to the impact of the problem of miss of accuracy estimation it is not possible to apply these methods because of the miss of one of its conditions which is dyadic sample size $n = 2^j$.

According to the great impact stem from that problem , many researchers who devoted their study to process this problem by using traditional methods

in processing missing data , where as the researcher used imputation methods more efficient and effective to process the missing data as a primary stage so that these data will be ready and available to wavelet application, as a result simulation experiment proved that the suggested methods (DRPW) are more efficient and superior to other methods , this paper also includes the auto correction of boundaries problem by using polynomial models , and using different threshold values in wavelet estimations , SINCE the suitable choice of this value is decisive accuracy of these estimations .

Keywords: Missing data, Wavelet regression, Polynomial, Non parametric Doubly Robust.