Journal of Economics
and Administrative Sciences (JEAS)

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

مقارنة بين طريقة المقدرات الموزونة والطريقة المقترحة(BEMW) لتقدير انموذج الشبة معلمي في حالة البيانات الغير تامة

م.د. سعد كاظم حمزه الباحث/ رند هيثم عبد الحسين
قسم الاحصاء/ كلية الادارة والاقتصاد
/جامعة بغداد
sdkadem100@yahoo.com Rand.alwakeel@yahoo.com

Received:7/1/2020 Accepted :10/2/2020 Published :June / 2020

هذا العمل مرخص تحت اتفاقية المشاع الابداعي نَسب المُصنَّف - غير تجاري - الترخيص العمومي الدولي 4.0
Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)

مستخلص البحث:

يستعمل الاحصاء بشكل عام في مختلف مجالات العلوم وخاصة في المجال البحثي إذ تتم المعالجة الاحصائية بأساليب متعددة وفقاً لطبيعة الدراسة وأهدافها ومن أهم تلك الأساليب هي بناء نماذج أحصائية و تتم من خلال نماذج الانحدار والذي يعد من أهم الأساليب الأحصائية المستعملة اذ يدرس العلاقة ما بين متغير معتمد او يسمى (بمتغير الاستحابة) ومتغير أو متغيرات توضيحية وفي هذا البحث تم توضيح تقدير انموذج الانحدار الخطي الجزئي مع تقدير القيم المفقودة فقدان عشوائي (MAR) في الجزء المعلمي حيث تم استعمال طرائق مطوره لتقدير معلمات انموذج الانحدار الخطي الجزئي والمتمثلة بطريقة المقدرات الموزونة وطريقة المقترحة WBEM وتمت مقارنة طريقتين بأسلوب المحاكاة باستعمال ثلاث احجام (n=100,150,200) واستعمال ثلاث قيم مختلفه لتباين $\sigma^2 = (1.5, 1, 0.5)$ ومتوسط صفر وتبين ان طريقة المقترحة (WBEM)تتفوق على طريقة المقدر الموزونة.

المصطلحات الرئيسة للبحث:_الانحدار الخطي الجزئي , نداريا واتسون, المقدرات الموزونة, الطريقة المقترحة

* البحث مستل من رسالة ماجستير

1- المقدمة:

يستعمل الاحصاء بشكل عام في مختلف مجالات العلم وخاصة في المجال البحثي , إذ ان استعمال الاساليب الاحصائية بطريقة مناسبة وصحيحة من قبل الباحث في مجال بحثه يزيد من فرصة الوصول الى نتائج سليمة سيستطيع من خلالها فهم الظاهرة قيد الدراسة فهماً صحيحاً , وبناءً على ذلك سيتبين للباحث أتخاذ القرارات الصائبة والمناسبة , إذ أن المعالجة الأحصائية هي الأساس في أي دراسة وتُعد العمود الفقري الذي يعتمد عليه في تصحيح الدراسة وتحليل نتائجها.

وتتم المعالجة الاحصائية بعدة اساليب وفقاً لطبيعة الدراسة وأهدافها ومن أهم تلك الأساليب هي بناء نماذج أحصائية والذي تتم من خلال نماذج الانحدار والذي يعد من أهم الأساليب الأحصائية المستخدمة الذي يدرس العلاقة ما بين متغير معتمد او يسمى (بمتغير الاستجابة) ومتغير أو متغيرات توضيحية أو تسمى (متغيرات المستقلة), وينقسم الانحدار الى نوعين بحسب نوع البيانات وطبيعتها حيث يسمى النوع الاول بالانحدار المعلمي والنوع الثاني بالانحدار اللامعلمي الا أن التطور الهائل الذي طرأ على علم الاحصاء دفع الباحثين الى البحث عن نماذج انحدار أخرى مما أدى بهم الى إيجاد ما يسمى بالانحدار الشبه المعلمي الذي يجمع ما بين النوعين السابقين والسؤال المطروح الان أي تلك الاساليب أكثر ملائمة للدراسة والجواب أهم لكي يتم تحديد الاسلوب الامثل للدراسة لابد من معرفة هذه الاساليب المختلفة ومعايير كل أسلوب , ولعل أول تلك الاساليب التي بدأ العمل بها هي نماذج الانحدار المعلمية (parametric regression) والتي نالت الحظ الاوفر من الدراسات الا ان هذه النماذج وبالرغم من دقة وكفاءة نتائجها إلا أنها تفرض افتراضات صارمه في حال تطبيقها كونها تفرض ان العينة لها توزيع معروف مثل التوزيعات الطبيعية أو توزيعات كاما وغيرها إذ ان الافتراض المخاطىء للتوزيع المعلمي للمسألة المعطاه قد يؤدي بالطرائق الاحصائية المستعملة الى استنتاجات غير صحيحة وتقديرات غير مقنعة بالاضافة الى كون بعض البيانات تكون فئوية أي أسمية أو رتبية الامر الذي دفع الباحثين الى البحث عن نماذج انحدار أقل صرامه واكثر مرونه للتعامل مع الكم الهائل من البيانات في العالم مما أدى الى ظهور انموذج الانحدار اللامعلمي (Nonparametric regression) والتي تعد من أهم المعالجات الاحصائية التي يمكن أستعمالها في حال عدم أستبعاد افتراضات معينه حول المجتمع الذي سحبت منه العينة , وقد أشار العديد من الباحثين الى أن هذه الاساليب تقبل اكثر مرونه كونها لاتعرف نوع التوزيع أو طبيعة البيانات الا ان هذا النوع من النماذج عانت هو الاخرى من مشكلة البعديه عند تعدد المتغيرات مما دفع الباحثين عن البحث عن نماذج اكثر مرونة الامر الذي أدى الى ظهور النماذج شبه المعلمية (semiparametric models) والتي تجمع ما بين خصائص الانحدار المعلمي وخصائص الانحدار اللامعلمي وبنفس الوقت تحقق نفس الغاية المرجوه وهي الحصول على أفضل منحنى للبيانات إلا أن هذه النماذج تعاني هي الاخرى مثلها مثل بقية النماذج أنفة الذكر من عدة مشاكل ومن أهمها مشكلة فقدان البيانات , إذ تعد هذه المشكلة من المشاكل التي تؤثر على التحليل الأحصائي بشكل كبير وتؤدي الى نتائج غير دقيقة ومضللة وهنالك أسباب عديدة لفقدان البيانات بعض منها يعود الى الحروب والاهمال والتلف وأخطاء جمع البيانات وبعض منها ميكانيكي مثل الاعطال التي تحدث في ألالات القياس أو تلف الذاكره ومنها أسباب بشريه عائده الى الفشل في التخزين للبيانات أو عدم الدقة بذلك . وهنا يبرز تحدي جديد للباحثين ودفعهم الى التساؤل هل يقفون عاجزين أمام هذه المشكلة أم البحث عن أساليب وطرائق لمعالجتها وجاءت الاجابات من قبل الباحثين عن أيجاد طرائق جديدة للتعامل مع تلك المشكلة ومعالجتها لكي يتم الحصول على نتائج اكثر دقة وأكثر مقبولية .

2- الانحدار اللامعلمي(Nonparametric regression):

يعد انموذج الانحدار اللامعلمي جزءاً مهم من الاحصاء اللامعلمي وهو يختلف عن انموذج الانحدار المعلمي في بناء هيكل الانموذج ،فهو يعتمد بشكل اساس و مباشر على البيانات وان طبيعة دالة الانحدار لهذا انموذج يكون بشكل مرن وغير ثابت حيث لا تمتلك توزيع ويكون وفق الصيغة الاتي:

$$y_i = g(t_i) + \varepsilon_i \quad \text{... (2)}$$

حيث يمثل Y المتغير المعتمد و $g(t_i)$ تمثل الدالة التمهيدية غير المعلومة و $(\varepsilon_i)_{i=1}^{n}$ الخطأ العشوائي يتوزع توزيع طبيعي مع وسط حسابي صفر والتباين σ^2.

وهناك عده طرائق لتقدير الدالة غير المعلومة g(t) لانموذج الانحدار اللامعلمي منها مقدر Nadarya-Watson

1-2 مقدراللامعلمي Nadarya-Watson Kernel Estimator:

يعد هذا المقدر من اكثر المقدرات استعمالاً لتقدير دالة الانحدار اللامعلمي بأستعمال معدل الاوزان للبيانات والتي اقترح من قبل الباحثان Nadarya -Watson ويتم حساب دالة الوزن حسب الصيغة الاتية [9pp254]:

$$w_{ni}(t) = \frac{k\left(\frac{t-T_i}{h_n}\right)}{\sum_{j=1}^{n} k\left(\frac{t-T_i}{h_n}\right)} = \frac{k(u)}{\sum k(u)} \quad \dots (3)$$

حيث ان $w_{ni}(t)_{i=1}^{n}$ تمثل سلسلة الوزن التي يكون مجموعها مساوٍ للواحد اذا كانت موجبه وk(u) تمثل دالة kernel تلبي شروطاً معينة و h_n عبارة عن عرض الحزمة او تسمى بمعلمة التمهيد واذ ان w(.) هي دالة الوزن وهي دالة حقيقة وغير سالبة و مستمرة ومحدده ومتماثلة وتكاملها مساوٍ واحد

وان k هو شكل اوزان kernel ويتم تحديد حجم الاوزان بواسطة h ويسمى بعرض الحزمة

ويتم تقدير دالة الانحدار اللامعلمي وفق الصيغة التالية :

$$\hat{g}_h(t) = \frac{n^{-1}\sum_{i=1}^{n} k_h(t-T_i)yi}{n^{-1}\sum_{i=1}^{n} k_h(t-T_i)} \quad \dots (4)$$

حيث $k(u)$ داله Kernel وهي دالة تقليل لـ u والتي تتغير حسب بعد او قرب البيانات المشاهدة t_i عن قيمة المشاهدة المدروسة T, و h تسمى عرض الحزمة او معلمة التمهيد $h > 0$

ان تقديرات معادلة انحدار kernel تأتي من حقيقة ان مقدر دالة الانحدار في t_i يتم الحصول عليها بواسطة المعدل الاوزان لقيم yj حيث الاوزان w_{ni} انتجت بواسطة داله kernel $k(u)$ كما أكد الباحث Hardel (1990) ان اختيار معلمة التمهيد (عرض الحزمة) مهم جدا في تقدير دالة kernel يتم ذلك عادة هو اختيار الدالة غير سالبة ومتماثلة حول الصفر ومستمرة ولديها المشتقة الثانية.

ومن الخصائص دالة kernel

$$1)\int K(u)du = 1 \quad \dots (5)$$

$$2)\int K(u)u\, du = 0 \quad \dots (6)$$

اما التحيز والتباين فيكون كالاتي

$$Bias = \frac{h^2 m''(t)}{2} dk \quad \dots (7)$$

$$Variance = \frac{\sigma^2(t)}{nh} ck \quad \dots (8)$$

وتعرف ck وdk

$$ck = \int k^2(u)du \quad \dots (9)$$

$$dk = \int u^2 k(u)du \quad \dots (10)$$

ck يستخدم لايجاد دالة kernel ذات الاقل تباين

dk يمثل التحيز المحاذي

3-انموذج الانحدار شبه المعلمي (Simeparamatric regression model)

يَعد انموذج شبه المعلمي هو دمج بين انموذج الانحدار المعلمي وانموذج الانحدار اللامعلمي حيث قدم عدد من الباحثين عدة بحوث حول هذا الانموذج وهناك عدة اسباب لأهتمام الباحثين بهذا انموذج ومن احد هذه الاسباب انه يكون اكثر مرونة من انموذج المعلمي لانه يجمع الاثنين معاً المعلمي واللامعلمي والسبب الاخر انه دائماً يعطي تفسير اسهل لتأثير كل متغير مقارنة مع الانحدار اللامعلمي التام بسبب علاقته التقليدية بانموذج الانحدار المعلمي ويعد افضل من انموذج الانحدار اللامعلمي لانه يتجنب مشكلة البعدية (curse of dimensionality), اذ أن متغير الاستجابة Y يرتبط بعلاقه خطية مع المتغير المستقل x ولكن يرتبط بشكل اللاخطي بمتغير المستقل آخر T.ومن اشهر النماذج شبه معلمية هو انموذج الانحدار الخطي الجزئي (plm) الذي يوفر حلاً وسطاً لتقدير كل من الجزء المعلمي β والجزء اللامعلمي g(t) وفي عام (1986) قام Engle

في تحليل العلاقة بين الطقس ومبيعات الكهرباء وهو من بين اول من طبق هذا الانموذج واذ افترض الباحث العلاقة بين متغير الاستجابة والمتغيرات المستقلة وفي عام (2011) قدمت الباحثة اسيل مسلم بحثاً في مقارنة بعض المقدرات شبة المعلمية لتقدير دالة استهلاك الطاقة الكهربائية لمدينة بغداد وايضاً وفي العام (2012) قدم كل من دكتور مناف يوسف واسيل مسلم بحثاً في مقارنة بعض المقدرات شبه ألمعلميه لتقدير دالة الانحدار, و يعد انموذج الانحدار الخطي الجزئي(plm) حاله خاصه من نماذج الانحدار التجميعيه (Additive models) Stone(1985),Tibshrani,Hastie(1990)

ويحدد الانموذج بالصيغة التالية :

$$\underline{Y} = \underline{X_i^T}\,\underline{\beta} + g(T) + \underline{\varepsilon} \quad \ldots (11)$$

Y متجه المشاهدات ويمثل متغير الاستجابه من درجة (n*1)

T المتغير اللامعلمي من درجة (n*1)

X متغير التوضيحي وهو متجه من درجة (p*1)

β متجه الجزء المعالم المجهوله من درجة (p*1)

g(T) داله تمهيديه غير المعلومة من درجة (n*1)

حيث ε يمثل الخطأ العشوائي ويتوزع توزيع طبيعي $N \sim (0, \sigma^2)$

4- البيانات المفقودة :

تعد مشكلة البيانات المفقودة مشكلة واسعة الآنتشار في العديد من مجالات الدراسة , حيث تحدث هذه المشكلة ضمن دراسات اساسية كالدراسات ذات البيانات الطولية والدراسات التجريبية وأيضا تحدث عادة هذه المشكلة التي ضمن المسوحات ويعود حدوث مشكلة فقدان البيانات الى عدة اسباب منها الرفض في الإجابة عن معلومة مفيدة كأن تكون تلك المعلومة تسبب احراج او يحدث لاسباب غير مقصودة مثل الحرائق او الضياع او غيرها من الاسباب او انها تكون مقصودة والغرض من ذلك التضليل على المعلومة المراد الحصول عليها من الدراسات لأسباب متعلقة بالدراسة وبما أن لكل عملية فقدان نمط يتم بموجبة الفقدان وفق احتمال معين لذلك قام (Rubin) (1976) بتصنيف ألية الفقدان الى ثلاثة فئات حيث أنه يرى بأن كل نقطة بيانات لديها أحتمال أن تكون مفقودة منها وتسمى العملية التي تحكم هذه الاحتمالات ألية البيانات المفقوده[3][21pp48] .

حيث تعتمد الطريقة المناسبة لمعالجة البيانات المفقودة على نمط و ألية الفقدان لتلك البيانات وبناءاً على ما تقدم توجد ثلاثة انواع من البيانات المفقودة وهي[19pp7] :

1- الفقدان العشوائي التام (MCAR)

2- الفقدان العشوائي(MAR)

3- الفقدان الغير عشوائي (NMAR)

حيث تم التطرق في هذا البحث الى عدة طرائق لتقدير القيمة المفقودة ومن هذه الطرائق هي :

1- طريقة المقدرات الموزونة

2- طريقة المقترحة(BEMW)

5- طرائق معالجة البيانات المفقودة

5-1 حذف الحالة الكامل (complet case cc):

تعد هذه الطريقة من اكثر الطرائق انتشاراً واستعمالاً والتي وضحت من قبل الباحثان Little and Rubin اذ تستعمل هذه الطريقة في معالجة القيم المفقودة في متغير (X) عن طريق حذف كل الازواج (X_i, Y_i) التي تعاني فقدان اي حذف كل (Y_i) مقابل (X_i) مفقودة , ومن ثم معاملة بقية البيانات على أنها تامة حيث نحصل على بيانات تامة ولكن بحجم اقل من البيانات الاصلية , وبالرغم من كون هذه الطريقة تعاني الضعف كونها تضحي ببعض المشاهدات الا انها تبقى مطلباً متاحاً ومن ضمن طرائق المعالجة [17pp48][13pp362] .

5-2 طريقة Expectation-Maximization with Bootstrapping (EM)، [19pp63]

ان الفكرة الاساسية لهذه الطريقة مستمدة من طريقة[20](Takahashi and Ito(2018 واخرون وهي طريقه هجينه مكونه من طريقة Bootstrap وخوارزمية EM اذا يتم في بادئ الامر توظيف طريقة Bootstrsapوالتي قدمت من قبل الباحث Efron(1979) كأداة لتقدير توزيع العينة عندما لا يمكن تطبيق الطرائق التقليدية او لا نملك معلومات حول توزيع العينة, حيث يتم من خلال طريقة Bootstrap توليد عدد

من العينات ولتكن Bعينه وبعد ذلك يتم استعمال طريقة خوارزمية EM لغرض معالجة القيم المفقودة وتقديرها والتي تتلخص بخطوتين هما :

الخطوة E:هي مرحله التوقع اذ يتم فيها حساب الداله Q-function من خلال لوغارتم الامكان الاعظم لمعدل البيانات التامة على التوزيع التنبؤي للبيانات المفقوده

$$Q(\theta / \theta^{t}) = \int l(\theta/x) pr(x_{miss}/x_{obs}, \theta^{(t)} dx_{miss} \qquad \dots (12)$$

الخطوة M :هي خطوه التعظيم لقيمه المعلمة التي تم ايجادها عند التكرار (t+1) بواسطه تعظيم Q-function

$$\theta^{(t+1)} = argmax\, Q(\theta/\theta^{(t)}) \qquad \dots (13)$$

6- طريقة المقدرات الموزونة

:(Weighted estimator method)

وهي طريقه مقترحه من قبل الباحث Qi-HuaWang [21pp53] وتعتبر طريقه مطوره لتقدير انموذج الانحدار الجزئي (PLM) عندما تكون بعض المتغيرات التوضيحية x_i مفقوده حيث عمد الباحث الى تطوير الاساليب المستعملة في تقدير كلاً من تقدير الجزء المعلمي (β) والجزء اللامعلمي $g(.)$ عند وجود تلك المشكلة ويمكن توضيح عمل هذه الطريقة وفق الاتي :

تحت افتراض الفقدان (MAR)

$$\beta = E^{-1}[(X\text{-}E[X|T])(X\text{-}E[X|T])^{T} E[(X\text{-}E|T])(Y\text{-}E[Y|T])]$$

$$E[(X - E[X\text{-}E[X|T])(X\text{-}E[X|T])^{T}] = E[\frac{\delta}{\Delta(z)}(X - E[X|T])(X - E[X|T])^{T}]$$

$$= E[\frac{\delta}{\Delta(z)}(X - E[X|T])(Y - E[X|T])]$$

و

$$E[X|T] = E[\frac{\delta X}{\Delta(z)}|T]$$

ويدمج طريقة الملئ "plug in" وطريقة العزوم الحصينة "samp moment" بالامكان تقدير β

$$\widetilde{\beta}_w = \widetilde{B}_n^{-1}\widetilde{A}_n \qquad \dots (14)$$

حيث ان

$$\widetilde{A}_n = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i(X_i - \widehat{g}_{1,n}(T_i))(Y_i - \widehat{g}_{2,n}(T_i)}{\Delta_n(Z_i)} \qquad \dots (15)$$

و

$$\widetilde{B}_n = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i(X_i - \widehat{g}_{1,n}(T_i))(X_i - \widehat{g}_{1,n}(T_i))}{\Delta_n(Z_i)} \qquad \dots (16)$$

ولتوضيح أكثر ان

$$Z_i = (Y_i, T_i) \qquad for\ i = 1, 2, \dots . n$$

اذن $\Delta(z)$ تقدر

$$\Delta_n(z) = \frac{\sum_{i=1}^{n}\delta_i k(\frac{z - Z_i}{h_n})}{\sum_{i=1}^{n} k(\frac{z - Z_i}{h_n})} \qquad \dots (17)$$

ولتكن $\omega(.)$ تمثل دالة **kernel** و h_n تمثل عرض الحزمه ولآيجاد الاوزان

$$W_{nj}(t) = \frac{\omega(\frac{t-T_j}{h_n})}{\sum_{j=1}^{n} \omega(\frac{t-T_i}{h_n})} \quad \dots (18)$$

يمكن تقدير $g(t)$ بالصيغة الاتية:

$$\hat{g}_w(t) = \hat{g}_{2,n}(t) - \hat{g}_{1,n}(t)\hat{\beta}_w \quad \dots (19)$$

ولتقدير $g_1(t)$ و $g_2(t)$ وفق الصيغة الاتية:

$$\hat{g}_{1,n}(t) = \sum_{j=1}^{n} W_{nj}(t)\frac{\delta_i X_i}{\Delta_n(Z_j)} \quad \dots (20)$$

$$\hat{g}_{2,n}(t) = \sum_{j=1}^{n} W_{nj}(t)Y_j \quad \dots (21)$$

7- **الطريقة المقترحة** Expectation-Maximization with Bootstrapping Weighted (BEMW)

ان هذه الطريقة مشابه لطريقة المقدرات الموزونة الا انه تم اقتراح بدلاً من معالجة الفقدان بطريقة الحالة التامة (complete case) يتم معالجة الفقدان بطريقة BEM ومن ثم تطبيق الطريقة المقدرات الموزونة على بيانات تامة.

8-**الجانب التجريبي**:

لتطبيق ما جاء في الجانب النظري يتم عن طريق استعمال اسلوب المحاكاة(Simulation) ويقصد بالمحاكاة هي عملية استعمال نماذج منطقية رقمية لنظام أو مفهوم أو لعملية للكشف عن السلوك المتوقع فيها عبر الزمن حيث المحاكاة هي طريقة او اسلوب اساسة اجراء التجارب عديدة في ظروف مختلفة لتقريب الى العالم الواقعي.

توليد المتغيرات العشوائية

تم تنفيذ تجارب المحاكاة بأستعمال ثلاثة حجوم للعينات (n=60,90,120) وبتكرار (Replicates=1000) لكل تجربة وكالاتي :

1- تم توليد المتغير التوضيحي X وفق التوزيع الطبيعي والمتغير T وفق التوزيع المنتظم تم استعمال متوسط القيم والتباين من البيانات الحقيقة كقيم اولية للتوليد وقيم المعاملات ايضاً

2- الاخطاء العشوائيه تتوزع توزيعاً طبيعاً بمتوسط صفر وتباين σ^2 $\epsilon \sim N(0,\sigma^2)$ وقد تم افتراض ثلاث قيم لتباين الخطأ هي (0.5 , 1 , 1.5).

3- المغير المعتمد Y سيتم حسابه باستعمال دالة الانحدار شبه المعلمية بدلالة المتغيرات التوضيحية التي تم توليدها من الفقرة (1) والفقرة (2)

9- **النماذج المستعملة في المحاكاة**

تم اختيار انموذج للدالة التمهيدية g(t) [21pp56]:

$$\mathbf{g(t) = 3.5\left(exp(-(4T-1)^2) + exp(-(4T-3)^2)\right) - 1.5}$$

اما الدالة اللبية (kernel) المستعملة فهي دالة Quatric kernel[21pp56]:

$$k(u) = \frac{15}{16}\,(1-u^2)$$

في حالة البيانات المفقودة وحسب الية الفقدان (MAR) فتكون وفق الصيغة الاتية[21pp56] :

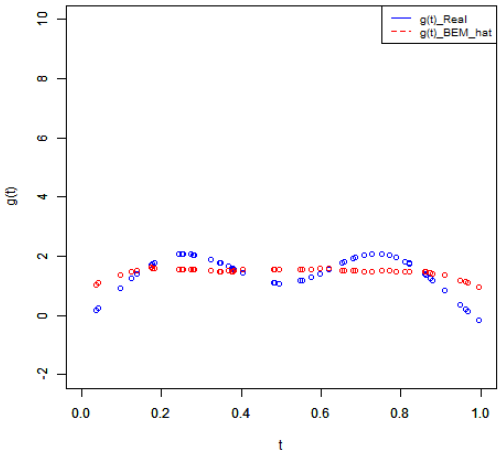
$$\boldsymbol{p(x,t) = p(\delta = 1|X = x)}$$
$$\boldsymbol{= -1/(1 + exp(-ln(9) - 0.1(y - mean(y)) - 0.2(t - mean(t))}$$

حيث يكون فقدان البيانات بشكل عشوائي (MAR) اذا كان سبب الفقدان له علاقة بقيم المتغيرات الاخرى فقط ومستقل عن القيمة المفقودة.

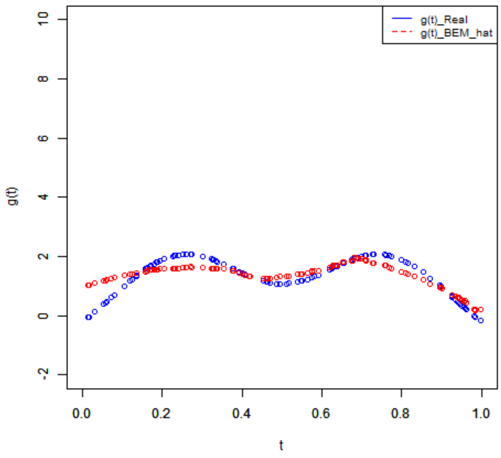
الجدول (1)

يبين متوسط مربعات الخطأ (MASE) لانموذج الانحدار الخطي الجزئي

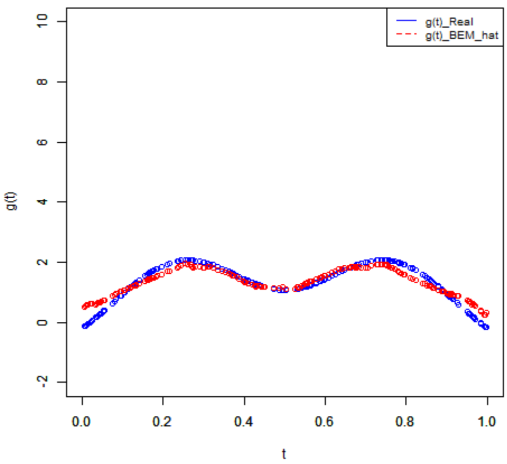
	method	W			BEMW		
نسبة الفقدان	σ^2 / n	0.5	1	1.5	0.5	1	1.5
10%	60	0.26259	0.96556	2.57168	0.16583	0.65813	2.24334
	90	0.23667	1.22777	1.99748	0.21157	1.04426	2.16375
	120	0.27888	0.90554	2.35014	0.27376	0.24392	2.12205
20%	60	0.33282	1.02106	2.20908	0.31646	0.94773	2.43136
	90	0.28651	0.94061	2.17916	0.24524	1.05183	1.97421
	120	0.31905	0.86583	1.81797	0.23042	0.24996	2.44971
30%	60	1.14194	1.12286	2.28411	0.32843	0.81313	2.33472
	90	0.26909	0.96211	3.08116	0.28442	0.71712	2.02691
	120	0.32945	0.86407	2.08746	0.25314	0.70456	1.66041



n=60



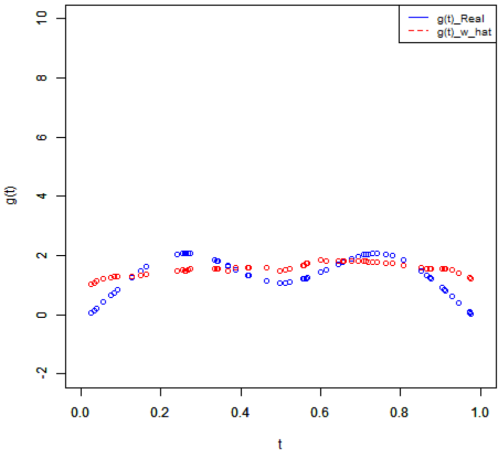
n=90



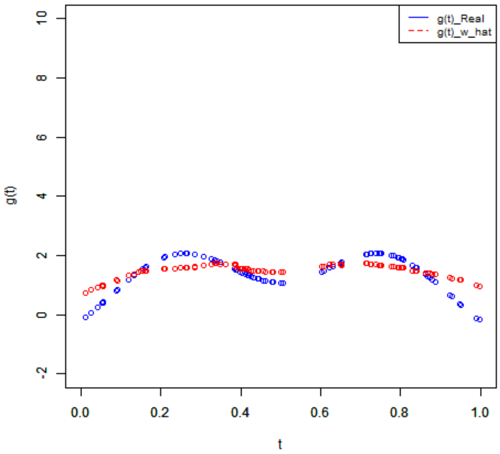
n=120

شكل (1)

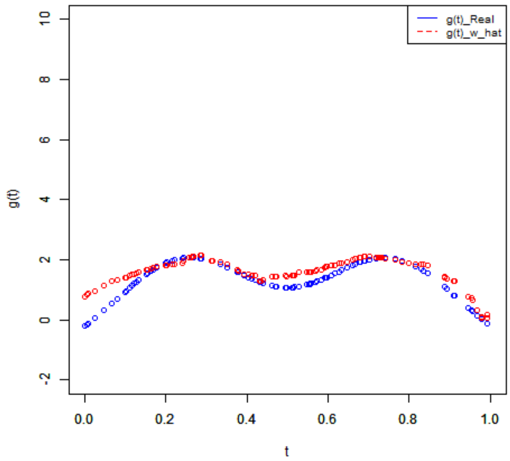
نتائج طريقة المقترحة BEMW عند تباين 0.5 ونسبة فقدان 10%



n=60



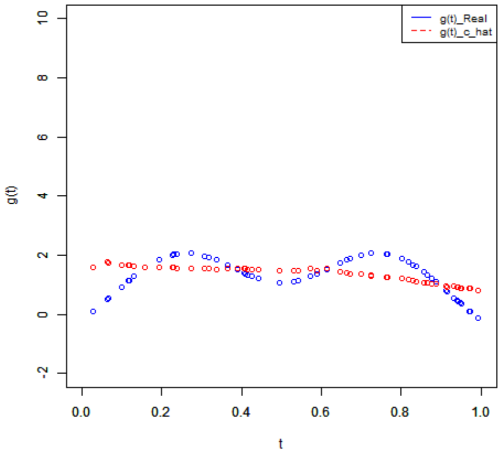
n=90



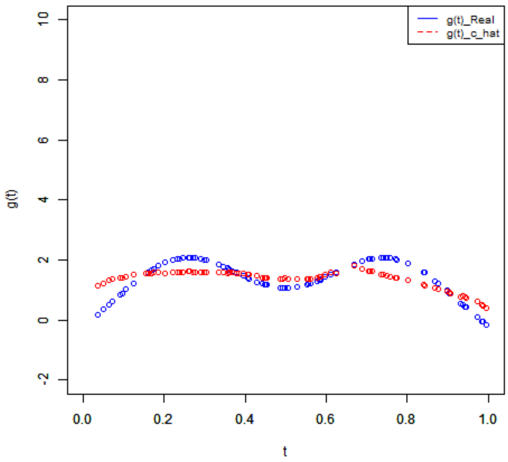
n =120

شكل (2)

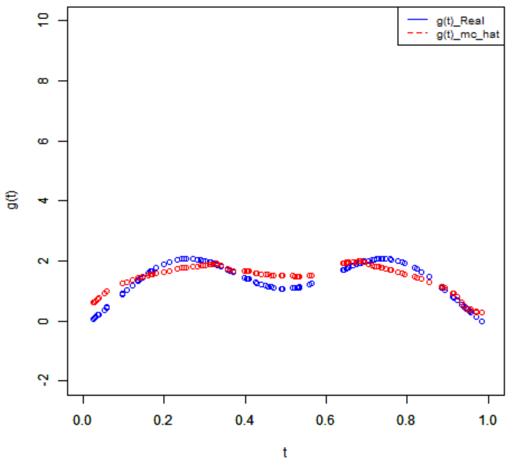
نتائج طريقة المقدرات الموزونة w عند تباين 0.5 نسبة فقدان 10%



n=60



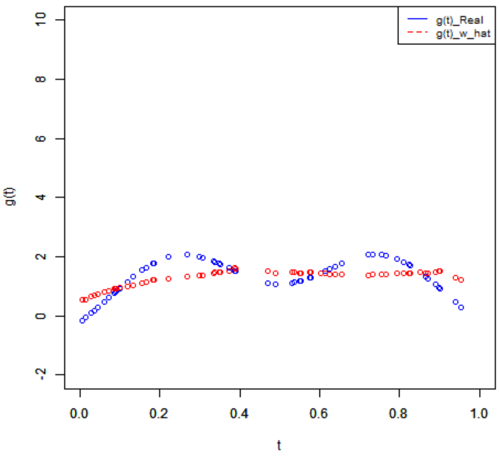
n =90



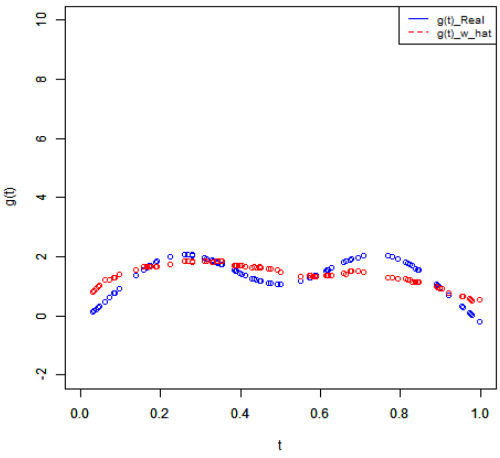
n =120

شكل (3)

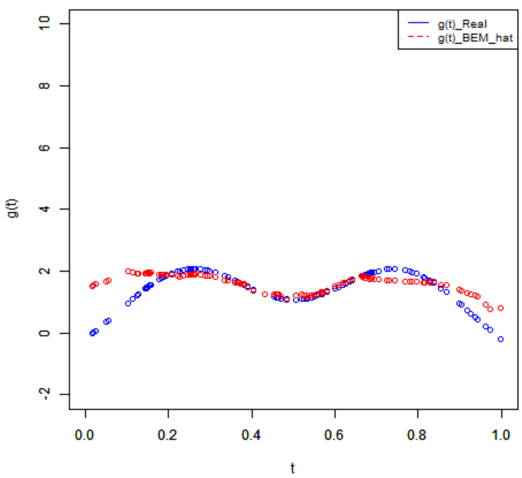
نتائج طريقة المقترحة BEMW عنده تباين 1 نسبة فقدان 20%



n =60



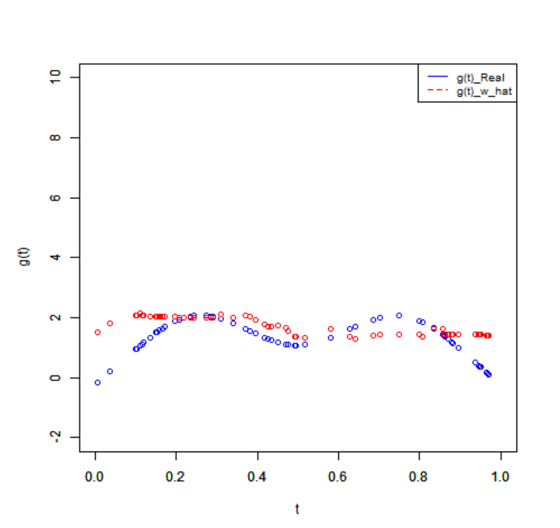
n =90



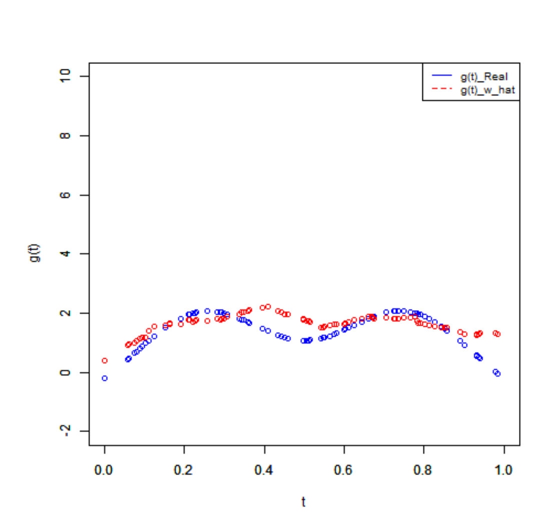
n =120

شكل (4)

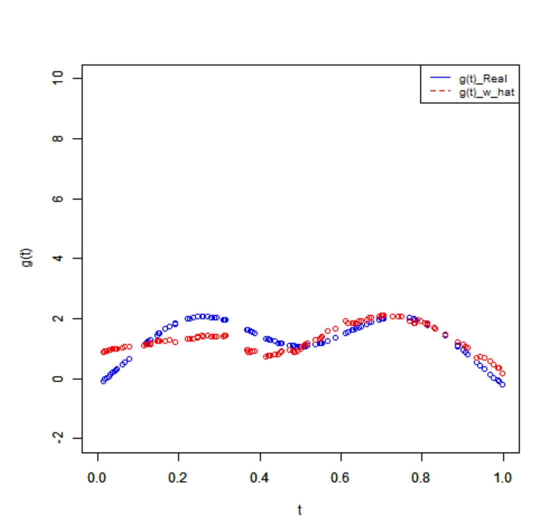
نتائج طريقة المقدرات الموزونة w عند تباين 1 نسبة فقدان 20%



n =60



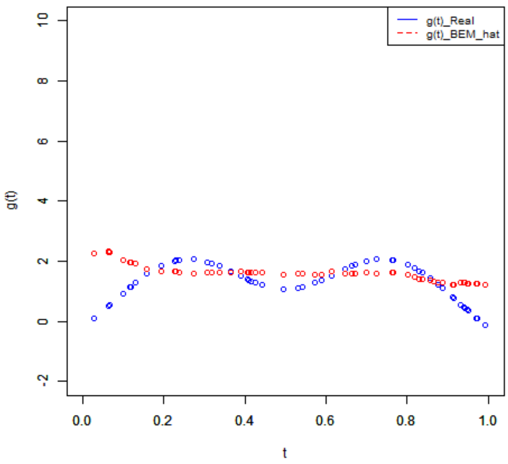
n =90



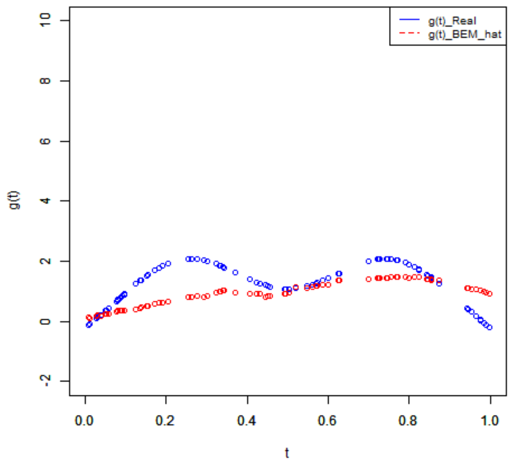
n =120

شكل رقم (5)

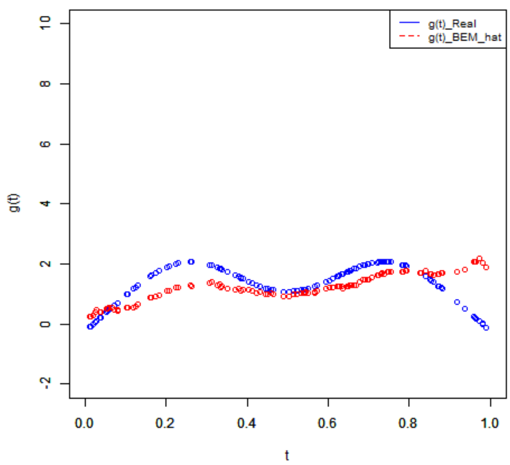
نتائج طريقة المقدرات الموزونة w عند تباين 1.5 ونسبة فقدان 30%



n =60



n =90



n =120

شكل رقم (6)
نتائج طريقة المقترحة BEMW عنده تباين 1.5 نسبة فقدان 30%

10- **الاستنتاجات**:

يمكن تحديد الاستنتاجات المستخلصة من البحث بالنقاط الاتية :

ان المرونة التي يوفرها الانموذج شبه المعلمي في توصيف البيانات بصورة عامة تكون كبيرة جداً مقارنة بالانموذج الخطي المعلمي والانموذج اللامعلمي حيث اظهرت نتائج كالاتي:

1- اظهرت نتائج على الاغلب ان قيمة MASE تقل عند زيادة حجم العينة عند جميع نسب الفقدان والتباينات ماعدا عند حجم عينة n=120 وعند تباين 0.5 نلاحظ تذبذب بقيم MASE لطريقة المقدرات الموزونة وعند n=120 ونسبة فقدان 20% وتباين 0.5 لطريقة BEMW نلاحظ تذبذب بقيمة MASE وايضاً عند نسبة فقدان 20% وحجم عينة 120 وتباين (1.5) وتباين (1) نلاحظ تذبذب بقيم MASE لطريقة BEMW

2- نلاحظ في جدول رقم (1) ان قيمة MASE على الاغلب تزداد كلما زادت نسبة الفقدان.

3- اشارت النتائج في جدول رقم (1) ان قيم MASE على الاغلب تزداد بزيادة قيم التباين.

4- اظهرت نتائج على الاغلب ان طريقة المقترحة BEMW افضل من طريقة المقدرات الموزونة ماعدا عند نسبة فقدان 20% وتباين 1 وعند تباين 1.5 وايضاً عند نسبة فقدان 30% وتباين 0.5 وتباين 1.5 نلاحظ ان طريقة المقدرات الموزونة هي الافضل.

المصادر العربية:

1- Hamza,saad kadhem,(2009)"A comparison of some Estimation Methods for Kernel Models in complete and Incomplete Data" Master Thesis in Statistics, College of Administration and Economics, University of Baghdad.

2- Hmood, Munaf Yousif., Issa, Aseel Muslim, (2011)" A comparison Of Some Semiparametric Estimators For consumption function Regression Journal of Economics and Administrative Sciences 18 (67),pp. 273-273.

3- Hmood, Munaf Yousif., Kattea,Mayasa Mohammed.,(2014)" Comparing the estimators of a semiparametric model using different smooth methods"Journal of Economics and Adminidtrative Sciences 20 (75),pp. 376-394.

4-Hmood, Munaf Yousif.,(2000)" A Comparison of non-parametric kernel estimators for estimating regression functions" Master Thesis in Statistics, College of Administration and Economics, University of Baghdad.

5-Issa, Aseel Muslil., (2011)" A comparison of Some Semiparametric Estimators For consumption function Estimation of electricity energy for Baghdad city" Master Thesis in Statistics, College of Administration and Economics, University of Baghdad.

6-Kattea,Mayasa Mohammed,(2014)"Acomparison of Nonparametric and semiparametric models in the presence of missing values with practical application of the Iraqi GDP(1971-2010)" Master Thesis in Statistics, College of Administration and Economics, University of Baghdad.

7-Qutaiba Nabeel .Nayef.,(2007)" Acomparison of Robust Bayesian method with other methodsTo estimate the parameters of the multiple linear regression model in case of incomplete data "Ph.D in statistics, College of Administration and Economics, University of Baghdad.

المصادر الاجنبية:

8-Altaher,A., and Ismail,M,T.,(2012)" Local Polynomial Wavelet Regression with Missing at Random" See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/232241949

9-Aydin,D.(2007)"Acomparison of the nonparametric regrestion model using smoothing spline and kernel regression" See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/242527986

10-Delleji,T., Zribi,M., and Hamida,A,B.,(2008)" On the EM Algorithm and Bootstrap Approach Combination for Improving Satellite Imag Fusion" World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering Vol:2, No:11.

11-Fan.J.(1992)," Design-adaptive Nonparametric Regression ",Journal of the American Statistical Association, Vol. 87, No. 420. (Dec., 1992), pp. 998-1004

12-Härdle ,W. (1990) "Applied Nonparametric regression " Cambridge, Cambridge University press

13-Härdle,W., Mori,Y.& Vieu,Ph.,(2007)" Statistical Methods for Biostatistics and Related Fields"Springer-Verlag Berlin Heidelberg.

14-Liang, H.,Wang, S., Robins, J.M., Carroll, R. J. (2004). Estimation in partially linear models with missing covariates. *Journal of the American Statistical Association*. *99*, 357–367.
15-Muller, M. (2014), "An Introduction to the estimation of GPLMs and Data Examples for the R gplm Package", https://cran.r-project.org/web/packages/gplm/vignettes/gplm-examples.pdf.
16-Pigott,.D(2001)" A Review of Methods for Missing Data" Loyola University Chicago, Wilmette, IL, USA ,Vol. 7, No. 4, pp. 353383
17-Qub,L.,and Change,Xiao-Wen.,(2004)" Wavelet estimation of partially linear models"Computational Statistics & Data Analysis 47 ,31 – 48
18-SCHEVE,k., JOSEPH,A., HONAKER,J.,and KING,G.,(2001)" Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation" American Political Science Review Vol. 95, No. 1 March 2001.
19-Takahash,M.,(2017)" Incomplete Data Analysis for Economic Statistics" The International University of Kagoshima See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/327338273.
20-Takahashi,&M.,Ito,T.,(2018)" Multiple Imputation of Missing Values in Economic Surveys: Comparison of Competing Algorithms" National Statistics Center, Tokyo, Japan
21-Wang,Q.,(2009)" Statistical estimation in partial linear models with covariate data missing at random" Ann Inst Stat Math , 61:47–84.
22-Zainuri,N,A., Jemain,A,A.,and Muda,N.,(2015)" A Comparison of Various Imputation Methods for Missing Values in Air Quality Data" Sains Malaysiana 44(3): 449–456

Comparison of weighted estimated method and proposed method (BEMW) for estimation of semi-parametric model under incomplete data

Dr.Saad Kadhem **Rand Haitham abd alhssin**

sdkadem100@yahoo.com **Rand.alwakeel@yahoo.com**

Received:7/1/2020 **Accepted :10/2/2020** **Published :June / 2020**

This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)]

Abstract

Generally, statistical methods are used in various fields of science, especially in the research field, in which Statistical analysis are carried out by adopting several techniques, according to the nature of the study and its objectives. One of these techniques is building statistical models, which is done through regression models. This technique is considered one of the most important statistical methods for studying the relationship between a dependent variable, also called (the response variable) and the other variables, called covariate variables. This research describes the estimation of the partial linear regression model, as well as the estimation of the "missing at random" values (MAR). Regarding the parametric part, a method has been developed to estimate the parametric of the partial linear regression model represented by weighted estimators as well as by the suggest method (EMBW). Two methods of simulation were compared using three sizes (n = 100,150,200) and using three different values for $\sigma^2 = (1.5, 1, 0.5)$ and zero mean and it was found that the proposed method (EMBW) was superior to the weighted estimator method.

Keywords:_partial linear regression, Nadarya-Watson, Weighted estimators, suggest method (Expectation-Maximization with Bootstrapping Weighted) (EMBW)