ISSN (1681- 6870)

تقدير انموذج الانحدارالخطي الجزئي الشبة معلمي بطرائق تقدير مختلفة بوجود بيانات غير تامة

رند هيثم عبد الحسين الوكيل[1] , م.د. سعد كاظم حمزه[2]
[1],[2] جامعة بغداد،كلية الادارة والاقتصاد

المستخلص:
يستعمل اسلوب الانحدار في قياس العلاقة بين متغيرين على هيئة دالة، للعلاقة بين متغير تابع والذي يرتبط بمتغير او متغيرات توضيحية وفي هذا البحث تم توضيح انموذج الانحدار الخطي الجزئي شبة المعلمي الذي دمج بين انموذج الانحدار المعلمي وانموذج الانحدار اللامعلمي والذي لاقه قبولاً واسعاً في الكثير من الدراسات حيث تم استعمال طرائق تقدير مطوره لتقديرانموذج الانحدار الخطي الجزئي شبة المعلمي مع فقدان في بعض مشاهدات في متغير توضيحي في الجزء المعلمي والمتمثلة بطريقة معايرة الانموذج MCM بالاضافة الى طريقة المقترحة MCBEM وتم اجراء اسلوب المحاكاة باستعمال ثلاث احجام (n=60,90,120) واستعمال ثلاث قيم افتراضية للتباين $\sigma^2 = (1.5,1,0.5)$ وقد تم التوصل الى ان طريقة المقترحة MCBEM اعطت نتائج اكثر دقة وذات اداء جيد عند التقدير .

الكلمات المفتاحية: الانحدار الخطي الجزئي، مقدر اللبي اللامعلمي، طريقة معايرة انموذج، EM algorithm

Estimating A Semi - Parametric Partial Linear Regression Model with Different Estimation Methods with Incomplete Data

Rand Haitham Abdel-Hussein **Dr. Saad Kazem Hamza**
University of Baghdad / College of Administration and Economics / Department of Statistics / Iraq.
rand.alwakeel@yahoo.com sdkadem100@yahoo.com

Abstract:

The regression method is used to measure the relationship between two variables in the form of a function, for the relationship between a dependent variable, which is related to one or explanatory variables. In this research, a parasympathetic partial linear regression model that represents the median state between the parameter regression model and the Non-parametric regression model has found wide acceptance in many Among the studies where methods of estimating a developer have been used to estimate the semi-linear partial linear regression model with a loss in the parameter part represented by the MCBEM model calibration method in addition to the MCB model calibration method proposed by the researcher Qi-HuaWang.

Keywords: Semi-Linear Partial Linear Regression Model, The Parameter Regression Model, The Non-Parametric Regression Model, Incomplete Data.

ISSN (1681- 6870)

1- المقدمه:

يستعمل اسلوب الانحدار في قياس العلاقة بين متغيرين على هيئة دالة، يسمى أحد المتغيرات متغير تابع والذي يرتبط بمتغير اوعدة متغيرات توضيحية، ، إذ أن المعالجة الأحصائية هي الاساس في أي دراسة وتعتبر العمود الفقري الذي يعتمد عليه في تصحيح الدراسة وتحليل نتائجها.

حيث يُجيب تحليل الانحدار على الاسئله حول أفضل انموذج يمثل ظاهرة ما ،والذي سيكون المدخل الاساس لفهم هذه الظاهرة وتحديد معالمها الرئيسية وينقسم الانحدار الى عدة انواع منها الانحدار المعلمي والانحدار اللامعلمي والانحدارشبة المعلمي الذي يدمج بين الجزء المعلمي والجزء اللامعلمي . ومن أهم النماذج شبة المعلمية النماذج الخطية الجزئية (partial linear models)(PLM) وهي احد نماذج الانحدار التجميعية الذي يسمح بتفسير اسهل لتأثير المتغيرات التوضيحية (1990) Hastis, ,Tibshrani,(1985)stone و يكون افضل من نماذج الانحدار المعلمية كون بعض النماذج تعاني من مشاكل عديدة منها مشكلة عدم تجانس في التباين او الارتباط الذاتي او التعدد الخطي وغيرها وكذلك الحال بالنسبة لنماذج الانحدار اللامعلمية كونها تعاني من مشكلة البعدية (the curse of dimensionality) وهذه المشاكل التي قادت الباحثين الى تفكير في التعامل مع النماذج شبة المعلمية التي تكون اكثر مرونة و توفر حلاً وسطاً بين النماذج المعلمية واللامعلمية.

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 أم عدم الدقة بذلك . وهنا يبرز تحدي جديد للباحثين ودفعهم الى التساؤل هل يقفون عاجزين أمام هذه المشكله 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- نماذج الانحدار الشبه معلميه

يعد انموذج الانحدار شبة 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 بهذ انموذج ومن احد هذه الاسباب انه يكون اكثر مرونة من انموذج الانحدار المعلمي وانموذج الانحدار اللامعلمي 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 المتغيرالمستقل x ولكن يرتبط بشكل اللاخطي بمتغير المستقل آخر T.ومن اشهر نماذج شبة المعلمية هو انموذج الانحدار الخطي الجزئي (plm) الذي يوفر حلاً وسطاً لتقدير كل من الجزء المعلمي β والجزء اللامعلمي$g(T)$ [10,14]

ISSN (1681- 6870)

ويحدد الانموذج بالصيغه التاليه :

$$\underline{Y} = \underline{X_i^T}\ \underline{\beta} + g(T) + \underline{\varepsilon} \qquad \dots (1)$$

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

2-1 الانحدار المعلمي (parametric regression)

اذا نعتبر y هو متغير تابع او متغير استجابة الذي يرتبط خطياً بمتغير مستقل وفقاً للمعادلة تالية:

$$y = \beta_\circ + \beta_1 X_{i1} \dots + \beta k X_{ik} \dots + \varepsilon \qquad \dots (2)$$

تمثل المعادلة مايعرف بنموذج الانحدار الخطي،وعندما يكون لدينا n من المشاهدات لمتغيرالاستجابة y وعدد من المتغيرات المستقلة كأن تكون k من المتغيرات ويمكن كتابة المعادلة بصيغة المصفوفات :

$$y = X\beta + \varepsilon \qquad \dots (3)$$

حيث y متجه ابعاده $n \times 1$ يمثل متغير الأستجابة و ε متجه الأخطاء يتوزع توزيع طبيعي مع متوسط حسابي يساوي صفر وتباين σ_I^2 ابعاده $n \times 1$ ، تمثل X مصفوفة المتغيرات المستقلة ابعادها $n \times p$ ، وتقدر β بطريقة المربعات الصغرى OLS حسب الصيغة الاتية:

$$\hat{\beta} = (X^T X)^{-1} X^T y \qquad \dots (4)$$

ISSN (1681- 6870)

2-2 الانحداراللامعلمي مقدر Nadarya-Watson Estimator

الفلسفة الرئيسية للانحدار اللامعلمي لتقدير دالة الانحدار $g(T)$ هي أستخدام معدل الاوزان للبيانات الخام حيث تكون الاوزان هي دالة المسافة في t على وجه الخصوص ان الاوزان هي دالة المسافة المتناقصة التي اقترحت من قبل Nadarya 1964-Watson 1964 ويتم حسابها حسب الصيغة الاتية [6]:

$$w_{ni}(t) = \frac{k(\frac{t-T_i}{h_n})}{\sum_{j=1}^{n} k(\frac{t-T_i}{h_n})} \quad \dots (5)$$

حيث ان $w_{ni}(t)_{i=1}^{n}$ تمثل سلسلة الوزن التي يكون مجموعها مساوٍ للواحد و(.)k تمثل دالة kernel و h_n عبارة عن عرض الحزمة اوتسمى بمعلمة التمهيد

واذ ان (.)w هي دالة الوزن وهي دالة حقيقة غير سالبة مستمره ومحدده ومتماثلة وتكاملها مساوٍ للواحد

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

وان شكل مقدر Nadarya-Watson يكون كالاتي:

$$\hat{g}_h(t) = \frac{n^{-1}\sum_{i=1}^{n} k_h(t-T_i)(Y_i - X_i'\beta)}{n^{-1}\sum_{i=1}^{n} k_h(t-T_i)} \quad \dots (6)$$

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 عليه بواسطة المعدل الاوزان لقيم Y_i 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[4,9].

ومن خصائص دالة kernel

$$1) \int K(u)du = 1 \quad \dots (7)$$

ISSN (1681- 6870)

$$2) \int K(u)u\,du = 0 \qquad \ldots (8)$$

وهناك بعض دوال kernel الشائعه[6]:

Kernel	Explict form
Gaussian kernel	$K(u)=\frac{1}{\sqrt{2\pi}}\exp(-u^2), u\in[-\infty,\infty]$
Uniform kernel	$k(u)=\frac{1}{2}, u\in[-1,1]$
Traingular	$(1-\lvert u\rvert), u\in[-1,1]$
Epanechnikv	$\frac{3}{4}(1-u^2), u\in[-1,1]$
Quatric	$\frac{15}{16}(1-u^2)^2, u\in[-1,1]$
Triweight	$\frac{35}{32}(1-u^2)^3, u\in[-1,1]$

3-البيانات المفقوده

مشكلة البيانات المفقودة تعد من احدى المشاكل المهمة التي تحدث اثناء التحليل الإحصائي وهي مشكلة واقعية في الدراسات الإحصائية المختلفة، حيث تحدث هذه المشكلة لاسباب مختلفة منها عدم الرد على الاجابة في حالة جمع البيانات عن طريق الاستبيان او المقابله او تحدث مشكلة الفقدان في الوثائق والسجلات في المستشفى بسبب عدم اكتمال الوثائق وأيضا تحدث عادة هذه المشكلة ضمن المسوحات او تحدث لاسباب غير مقصودة مثل الحرائق او الضياع او غيرها من الاسباب او انها تكون مقصودة والغرض من ذلك التضليل على المعلومة المراد الحصول عليها .

لذلك في الممارسات العملية عادةً ماتتعرض المتغيرات سواء كانت المستقلة او متغيرالاستجابه الى الفقدان و لاسباب مختلفة كون تلك البيانات وكما ذكرنا يتم جمعها اما عن طريق المقابلات المباشرة عن طريق مستندات مثل سجلات المستشفى ، وعادة ما تتعرض تلك المستندات (الطبلات) الى فقدان كأن يكون متعرض لاهمية وحرج بعض البيانات أو أن فقدانها يعود لأسباب الاهمال وعدم التوثيق الدقيق وهنالك العديد من الاسباب التي تؤدي الى فقدان والتي يصعب حصرها في هذه الدراسة [18].

ISSN (1681- 6870)

قام كلاً من Little و Rubin (1987) بتميز بين أليات الفقدان الثلاثة حيث وضح الباحثان أنه اذا لم يرتبط سبب فقدان البيانات بالقيم المرصوده للمتغيرات الاخرى يسمى حينها الفقدان العشوائي (MAR) وهي الالية الاكثر انتشاراً وحدوثاً وواقعية اما اذا كان سبب الفقدان مستقلاً عن القيم المفقوده كما أنه لايعتمد على أي قيمة من قيم المشاهدات الأخرى في العينه (MCAR) واما الية الفقدان الغير عشوائي (NMAR) فأنها تعتمد على القيم الغير مشاهدة او القيم المفقودة[16].

ان مشكله البيانات المفقوده في المتغيرات المستقلة كانت محط اهتمام كبير للعديد من الباحثين ،اذ تم اقتراح العديد من الطرق لمعالجة تلك المشكلة وأحد أكثر الطرق شيوعاً (طريقة الحذف) وطريقة تحليل الانحداروطرائق التمهيد وطريقة تعويض المتوسط وغيرها، الا ان هذه الطرائق تعاني في اغلب الأحيان عند معالجتها لمشكلة الفقدان اما بسبب الفقدان الموجودة ضمن البيانات او انها تحدث تحيزاً كبيراً في عملية الاستدلال الاحصائي لتلك البيانات لذلك سعى العديد من الباحثين الى ايجاد طرائق كفؤه بأمكانها معالجة مشكلة الفقدان و بأكبر دقة ممكنة للحصول على استنتاجات بعيده عن التحيز والتضليل ليتمكنوا من الخروج بنتائج كفؤه .

4- طرائق معالجة البيانات المفقودة

4-1 حذف الحالة الكامل Complete case (cc):

تعد هذه الطريقة من اقدم الطرائق المعالجة والتي أقترحت من قبل الباحثان Little and Rubin اذ تعد هذه الطريقة في معالجتها للقيم 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 إلا أنها تكون بحجم اقل من الحجم الاصلي ، وبالرغم من كون هذه الطريقة تعاني الضعف كونها تضحي ببعض المشاهدات الا انها تبقى مطلباً متاحاً ومن ضمن طرائق المعالجة[11] .

4-2 طريقة Expectation-Maximization with Bootstrapping (BEM) :

ان الفكرة الاساسية لهذه الطريقة مستمدة من [17](Takahashi and Ito(2018 واخرون وهي طريقه هجينه مكونه من طريقة Bootstrap وخوارزمية EM اذا يتم في باديء الامر توضيف طريقة Bootstrsap والتي قدمت من قبل الباحث Efron(1979) كأداة لتقدير توزيع العينه عندما لايمكن تطبيق الطرائق التقليديه او لا نملك معلومات حول توزيع العينه، حيث يتم من خلال طريقة Bootstrap توليد عدد من العينات ولتكن Bعينه وبعد ذلك يتم استعمال طريقة خوارزمية EM لغرض معالجة القيم المفقوده وتقديرها والتي تتلخص بخطوتين هما :

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

$$Q(\theta \,/\, \theta^t) = \int l(\theta/x) pr(x_{miss}/x_{obs}, \theta^{(t)} dx_{miss} \qquad \dots (9)$$

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

$$\theta^{(t+1)} = argmax\, Q\left(\theta/\theta^{(t)}\right) \qquad \dots (10)$$

ISSN (1681- 6870)

5- طرائق التقدير

منذ فترة طويلة والباحثين وبالخصوص الاحصائيين يدأبون في العمل على ايجاد طرائق فعالة ودقيقة للحصول على تقديرات كفوءة ذات استنتاجات معقولة في ظل العديد من المشاكل وفي مقدمتها مشكلة البيانات المفقودة تتنوع تلك الاساليب بحسب نوع الفقدان وطبيعة البيانات التي تحدد نوع انموذج الدراسة وبالتالي ستتطرق في هذا البحث الى اهم طرائق التي استخدمها الباحث في تقدير دالة الانحدار شبة المعلمي في ظل مشكلة البيانات المفقودة في الجزء المعلمي.

5-1 طريقة معايرة انموذج MCM

5-2 طريقة مقترحة MCBEM

5-1 طريقة معايرة الانموذج

(Model calibration method)

وهي طريقه مقترحه من قبل(2009) **Qi-HuaWang** وتعتبر طريقة مطوره لتقدير انموذج الانحدار الجزئي (PLM) عندما تكون بعض المتغيرات التوضيحية x_i مفقوده حيث عمد الباحث الى تطوير الاساليب المستخدمة في تقدير كلاً من تقدير الجزء المعلمي (β) والجزء اللامعلمي $g(.)$ عند وجود تلك المشكلة ويمكن توضيح عمل هذه الطريقة وفق الاتي[18] :

تحت افتراض الفقدان (MAR) والذي يكون وفق الدالة الاتية :

$$\Delta(Y,t) = p(\delta = 1|Y = y, T = t) \qquad \dots (11)$$

نلاحظ ان

$$\mathrm{E}[\delta_i X_i / \triangle (Y_i, T_i)|X_i, Y_i, T_i] = X_i \qquad for\ i = 1,2, \dots n$$

حيث ان (X_i, Y_i, T_i) يتبع الانحدار شبه المعلمي لكل قيم $i = 1,2, \dots n$ ،ولتقديركلاً من الجزء المعلميβ والجزء اللامعلمي المتمثل بالدالة اللامعلمية $g(T)$ تحت افتراض الفقدان العشوائي MAR ويكون كالاتي :

$$E[Y_i - X_i^T B - g(T_i)|\ X_i, T_i)$$

$$= E\{E[Y_i - X_i^T B - g(T_i)|\ X_i, Y_i, T_i]|X_i, T_i\} = 0 \qquad \dots (12)$$

حيث ان U_i

ISSN (1681- 6870)

$$U_i = \delta_i X_i / \triangle (Y_i, T_i) \quad \dots (13)$$

وبناءً على اعلاه يصبح انموذج الانحدار الخطي الجزئي plm وفق الاتي:

$$Y_i = X_i^T B + g(T_i) + e_i \quad \dots (14)$$

وبواسطة Speckman (1988) انموذج في المعادلة (11) يكون مكافأً للتالي:

$$Y_i - E[Y_i|T_i] = (U_i - E[Y_i|T_i])^T \beta + e_i \quad \dots (15)$$

ليكن

$$g_1(t) = E[X|T = t] \quad \text{و} \quad g_2(t) = E[Y|T = t]$$

اذن $g_1(.), \Delta(.,.)$ و $g_2(.)$ دوال معلومة عند ذلك بالامكان تطبيق طريقة المربعات الصغرى على (2-27) لتقدير β

$$\tilde{\beta}_{MCM} = B_n^{-1} A_n$$

حيث

$$B_n = \frac{1}{n}\sum_{i=1}^{n}[(U_i - g_1(T_i))(U_i - g_1(T_i))^T] \quad \dots (16)$$

و

$$A_n = \frac{1}{n}\sum_{i=1}^{n}(U_i - g_1(T_i))(Y_i - g_2(T_i)) \quad \dots (17)$$

اذ ان k(.) تمثل دالة kernel و h_n تمثل عرض الحزمة

وللتوضيح أكثر ان

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

ISSN (1681- 6870)

اذن $\Delta(z)$ تقدر

$$\Delta_n(z) = \frac{\sum_{i=1}^{n} \delta_i k(\frac{z - Z_i}{h_n})}{\sum_{i=1}^{n} k(\frac{z - Z_i}{h_n})} \quad \dots (18)$$

ولتكن $\omega(.)$ تمثل دالة kernel و b_n تمثل عرض الحزمة ولأيجاد الاوزان

$$W_{nj}(t) = \frac{\omega(\frac{t - T_j}{b_n})}{\sum_{j=1}^{n} \omega(\frac{t - T_i}{b_n})} \quad \dots (19)$$

ولتقدير $g_1(t)$ و $g_2(t)$

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

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

ويمكن ايجاد مقدر β ليقال $\hat{\beta}_{MC}$ وليكون $\tilde{\beta}_{MC}$ و$\Delta(.,.)$ و $g_1(.), g_2(.)$ تستبدل ب

$\Delta_n(.,.), \hat{g}_{1,n}(t), \hat{g}_{2,n}(t)$ على توالي

$$\hat{\beta}_{MCM} = \hat{\beta}_n^{-1} \hat{A}_n \quad \dots (22)$$

واخذ توقع الشرطي ل T المعطى في المعادلة $(26 - 2)$ تحت فقدان MAR

$$g(t) = g_2(t) - g_1^T(t) \quad \dots (23)$$

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

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

وان g(.) تقدر :

ISSN (1681- 6870)

$$\hat{g}_{MC}(t) = \hat{g}_{2,n}(t) - \hat{g}_{1,n}^{T}\hat{\beta}_{MC} \quad \dots (24)$$

5-2 الطريقة مقترحة MCBEM

ان هذه الطريقة مشابه لطريقة معايرة الانموذج الا انه تم اقتراح بدلاً من معالجة الفقدان بطريقة الحالة التامة (complete case) يتم معالجة الفقدان بطريقة BEM الموضحة في 2-4 ومن ثم تطبيق الطريقة المذكورة على بيانات تامة.

6- الجانب التجريبي:

لتطبيق ماجاء 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 االعشوائية :

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)

4- تم اختيار انموذج للدالة التمهيدية g(t) [18]:

$$g(t) = 3.5\big(exp(-(4T-1)^2) + exp(-(4T-3)^2)\big) - 1.5$$

5- امالدالة اللبية (kernel) المستعملة فهي دالة Quatric kernel[6]:

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

6- في حالة البيانات المفقودة وحسب الية الفقدان (MAR) فتكون وفق الصيغة الاتية[18]:

ISSN (1681- 6870)

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

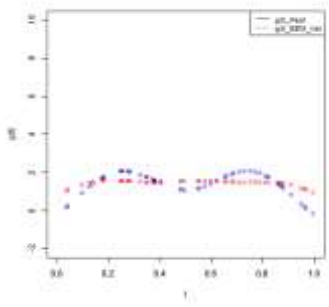
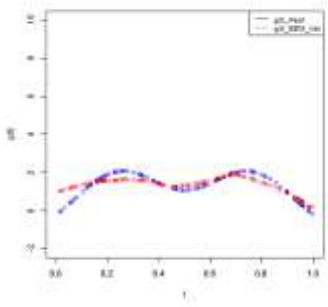
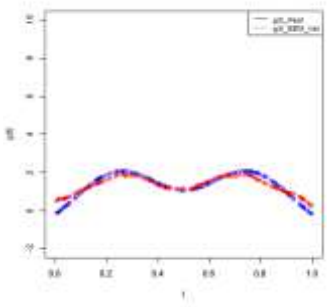
حيث يكون فقدان البيانات بشكل عشوائي (MAR) اذا كان سبب الفقدان له علاقة بقيم المتغيرات الاخرى فقط ومستقل عن القيمة المفقودة.

الجدول (1)

جدول يبين متوسط مربعات الخطأ (MSE) للانموذج شبة المعلمي

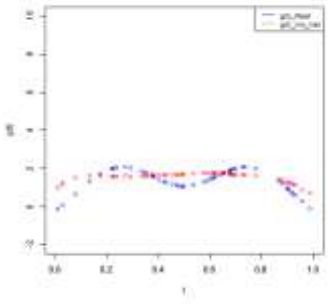
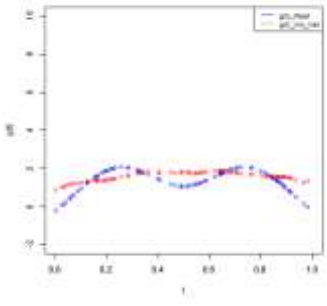
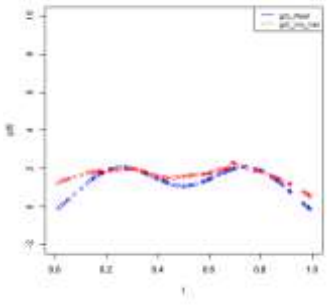
	method	MCM			MCBEM		
نسبة الفقدان	σ^2 / n	0.5	1	1.5	0.5	1	1.5
10%	60	0.26378	0.97056	2.57079	0.16583	0.65813	2.24334
	90	0.24035	1.22104	2.00399	0.21157	1.04426	2.16375
	120	0.29789	0.93216	2.36659	0.27376	0.24392	2.12205
20%	60	0.33086	1.04993	2.21841	0.31646	0.94773	2.43136
	90	0.33338	0.94253	2.22376	0.24524	0.56641	1.97421
	120	0.31201	0.86534	1.86451	0.23042	0.24996	2.44971
30%	60	1.15795	1.27191	2.37429	0.32843	0.81313	2.33472
	90	0.33627	0.98532	3.10779	0.28442	0.71712	2.02691
	120	0.33367	0.86401	2.17551	0.25314	0.70456	1.66041

ISSN (1681- 6870)

		
MCBEM -60	**MCBEM -90**	**MCBEM -120**

الشكل رقم (1)

نتائج طرق المحاكاة لطريقة MCBEM وعند نسبة فقدان 10%

		
MCM-60	**MCM-90**	**MCM-120**

الشكل رقم (2)

نتائج طرق المحاكاة لطريقة MCM وعند نسبة فقدان 10%

ISSN (1681- 6870)

7- الاستنتاجات:

يمكن تحديد الاستنتاجات المستخلصة من البحث بالنقاط الاتية :

ان المرونه التي يوفرها الانموذج الشبة معلمي في توصيف البيانات بصورة عامة تكون كبيرة جداً مقارنة بالانموذج الخطي

1- اظهرت نتائج في الجدول (1) ان طريقة المقترحة MCBEM هي الافضل من طريقة MCM في جميع نسب الفقدان وجميع الحجوم والتباينات ماعدا عند تباين $\sigma^2 = 1.5$ ونسبة فقدان 10% وحجم عينة 90 وعند نسبة فقدان 20% وحجم عينة (120,60) ظهرت ان طريقة MCM هي الافضل.

2- نلاحظ في جدول (1) بصورة عامة ان MSE تقل في طريقة MCM كلما زاد حجم العينة ماعدا عند تباين $\sigma^2 = 0.5$ وحجم عينة 120 ونسبة فقدان 30% وعند نسبة فقدان 20% وتباين $\sigma^2 = 1$ وحجم عينة 90 وفي طريقة MCBEM $\sigma^2 = 0.5$ وعند حجم العينة 120وحجم عينة 20%,30% على توالي .

3- تذبذب قيم MSE عند زيادة تباين الخطأ ولجميع التباينات وحجوم العينات

ISSN (1681- 6870)

المصادرالعربية:

[1] القزاز، قتيبة نبيل نايف (2007) . مقارنة اساليب بيز الحصين مع طرائق اخرى لتقدير معالم أنموذج الانحدار الخطي المتعدد في حالة بيانات غير التامة . اطروحة دكتوراه فلسفة في الاحصاء ، كلية الادارة والاقتصاد ، جامعة بغداد.

[2] حمزة، سعد كاظم (2009) . مقارنة بعض الطرائق أللبية في تقدير نماذج النحدار اللامعلمية بوجود بيانات تامة وغير تامة. رسالة ماجستير في الإحصاء كلية الادارة والاقتصاد، جامعة بغداد.

[3] حمود، مناف يوسف (2000) . مقارنة مقدرات kernel اللامعلمية لتقدير دوال الانحدار. رسالة ماجستير في الإحصاء، كلية الادارة والاقتصاد، جامعة بغداد.

[4] كاطع، مياسة محمد (2014) . مقارنة النماذج اللامعلمية وشبة المعلمية بوجود قيم مفقودة مع تطبيق عملي للنتائج المحلي الاجمالي العراقي للمدة (1971-2010) . رسالة ماجستير في الإحصاء كلية الادارة والاقتصاد، جامعة بغداد.

المصادر الاجنبية:

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

[6] Aydin,D.,"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 ,(2007).

[7] Delleji,T., Zribi,M., and Hamida,A,B.," 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.,pp. 3796-3805, (2008).

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

[9] Härdle ,W., "Applied Nonparametric regression " Cambridge, Cambridge University press ,(1990).

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

[11] Liang, H.,Wang, S., Robins, J.M., Carroll, R. J. "Estimation in partially linear models with missing covariates" *Journal of the American Statistical Association*, Vol. 99, No. 466 ,pp. 357-367,(2004).

[12] -Muller, M., "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. ,(2014)
[13] Pigott,.D, " A Review of Methods for Missing Data" Loyola University Chicago, Wilmette, IL, USA ,Vol. 7, No. 4, pp. 353-383, (2001).
[14] Qub,L., and Change, Xiao-Wen.," Wavelet estimation of partially linear models", Computational Statistics & Data Analysis 47,pp.31-48,(2004).
[15] SCHEVE,k., JOSEPH,A., HONAKER,J., and KING,G.," Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation" American Political Science Review Vol. 95, No. 1 March, pp. 49- 69 ,(2001).
[16] Takahash, M.," 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, (2017).
[17] Takahashi,&M.,Ito,T., " Multiple Imputation of Missing Values in Economic Surveys: Comparison of Competing Algorithms", National Statistics Center, Tokyo, Japan, (2018).
[18] Wang, Q.," Statistical estimation in partial linear models with covariate data missing at random" Ann Inst Stat. Math , 61:pp.47-84.,(2009).
[19] Zainuri, N,A., Jemain, A,A., and Muda, N., "A Comparison of Various Imputation Methods for Missing Values in Air Quality Data" Sains Malaysiana Vol.44,NO.3,pp 449-456,(2015).