مقارنة بين طريقتي بيز والانحدار الموضعي لتقدير انموذج الانحدار اللوجستي شبه المعلمي

أ.م.د.قتيبة نبيل نايف/ كلية الادارة والاقتصاد/ جامعة بغداد/ قسم الاحصاء
الباحث/ علي محمد علي جيجان

المستخلص :

يلقى موضوع تحليل النماذج شبة المعلمية والذي يدمج النماذج المعلمية والنماذج اللامعلمية اهتماماً واضحا في معظم الدراسات التي تأخذ طابعا أكثر تقدما في عملية التحليل الإحصائي الدقيق الذي يهدف إلى الحصول على مقدرات ذات مستوى عالٍ من الكفاءة، في بعض الدراسات يكون المتغير المعتمد للأنموذج الشبه المعلمي ثنائي الاستجابة اما يساوي صفراً لعدم حدوث الاستجابة او يساوي واحداً لحدوث الاستجابة وهذا ما يسمى بالانحدار اللوجستي .

سنستعرض في هذا البحث طرائق شبه معلميه لتقدير أنموذج انحدار لوجستي وهذه الطرائق هي المقدر الموضعي الاصغر (classic local least estimator) وطريقة بيزية لتقدير معالم الانموذج، وتمت المقارنة بين هاتين الطريقتين بأستعمال معيار المقارنة متوسط مربعات الخطأ (MSE) .

إذ تمت المقارنة من خلال أسلوب المحاكاة باستعمال النماذج اللامعلميه واحجام عينات وتباينات مختلفة للتحقق من أداء الطرائق باستعمال معيار المقارنة، حيث كانت اهم الاستنتاجات ان الطريقة البيزية تتفوق على طريقة الانحدار الموضعي في حال كانت حجوم العينات صغيرة.

المصطلحات الرئيسية للبحث/الانحدار اللوجستي- أنموذج شبه المعلمي- المقدر(LLE)- مقدر بيز- دالة اللب-عرض الحزمة .

*بحث مستل من رسالة ماجستير

١- المقدمة Introduction

تحليل الانحدار هو اداة احصائية تقوم ببناء انموذج احصائي وذلك لتقدير العلاقة بين المتغير التابع وعدة متغيرات توضيحية،بحيث ينتج معادلة احصائية توضح العلاقة بين المتغيرات يمكن استعمال هذه المعادلة في معرفة نوع العلاقة بين المتغيرات من خلال تقدير الأنموذج.عندما تكون العلاقة في الانموذج الاحصائي بين متغير تابع واحد ومتغير مستقل واحد فان هذا الانموذج هو ابسط نماذج الانحدار ويسمى انموذج خطي بسيط $(simple\ linear\ regression)$، وعندما تكون عدد المتغيرات توضيحية اكثر من متغير واحد يسمى انموذج الانحدار المتعدد $(multiple\ regression)$.

قد ظهرت العديد من نماذج الانحدار غير الخطية التي تتعامل مع البيانات ثنائية الاستجابة ومنها انموذج الانحدار اللوجستي $(\text{logistic regression})$ الذي يكون فيه التنبؤ بين الصفر والواحد من خلال دالة الانحدار اللوجستي وهذا مالا نراه في انموذج الانحدار الخطي الذي تكون فيه التنبؤات بين $(-\infty\ ,\ +\infty)$.

وفي احيان أخرى تكون المتغيرات التوضيحية غير خطية مما دعى الباحثين الى ايجاد اسلوب اخر يتعامل مع التأثير اللاخطي لهذا المتغير وهو الانحدار اللامعلمي .

لكن من الملاحظ ان الانحدار اللامعلمي يعاني من بعض المشاكل ومنها مشكلة الابعاد $(\,the\ curse\ of\ dimensonality\,)$ والتي تحصل عند زيادة عدد المتغيرات التوضيحية، مما ادى ذلك الى ظهور بعض الاساليب الحديثة في نماذج الانحدار ومنها نماذج الانحدار شبه المعلمية $(\text{ semiparametric regression })$ لتحليل البيانات والتي تدمج المركبات المعلمية مع المركبات اللامعلميه التي تسهم الى حد ما في الحصول على قدرة عملية لتلافي مشكلة الابعاد .

هناك بعض التحليلات الاحصائية عند عدم توفر معلومات كاملة عن العينة قيد الدراسة لذلك يلجأ الباحثون الى وضع افتراضات تتلخص بكون المعلمات المراد تقديرها تكون عشوائية وهذا مما يتطلب الحصول على معلومات مسبقة حول المعلمة قبل الحصول على العينة العشوائية حيث يمكن صياغة هذه المعلومات المسبقة بشكل توزيع احتمالي يسمى التوزيع الاولي $(\text{ prior distribution })$ وبالتالي من ممكن الحصول على دالة الكثافة الاحتمالية اللاحقة (posterior p.d.f) لهذه المعلمات وذلك بدمج دالة الامكان للمشاهدات **(Likelhood function)** مع دالة الاحتمالية الاولية والذي يحتوي على جميع المعلومات حول المعلمات المراد تقديرها وهذا ما يسمى بالنظرية البيزية $(\text{ Bayesian theorem })$

٢- هدف البحث

أن هدف البحث يتمحور حول تقدير أنموذج الانحدار اللوجستي شبه المعلمي والمقارنة بين بعض طرائق تقديره والمتمثلة بالمقدر الموضعي الاصغر **local least estimator** وطريقة بيز .

٣- الجانب النظري

٣-١ إنموذج الانحدار شبه المعلمي semiparametric Regression Model [17 ,7]

يعد انموذج الانحدار الخطي الجزئي **(Partial linear model)** احد نماذج الانحدار شبه المعلمي أذ أقترح هذا النموذج من قبل الباحث speckman ١٩٨٨م[14] وان هذا النموذج يعتمد على متغيرات خطية (معلمية) واخرى غير خطية (لامعلمية) اذ حيث أن هذه المتغيرات الخطية واللاخطية تؤثر في متغير الاستجابة **Y**، ويعد حالة خاصة من النماذج التجميعية (additive model) وكذلك يتميز بميزة وهي امكانية تجنب مشكلة الابعاد والتي تحدث في النماذج اللامعلمية عند زيادة عدد المتغيرات التوضيحية، ويمكن تمثيل إنموذج الانحدار شبه المعلمي بالصيغة الاتية :-

$$Y = XB + g(t) + \varepsilon \quad ...(1)$$

حيث أن :

- Y يمثل المتغير المعتمد من درجة $(n \times 1)$.
- X يمثل مصفوفة المتغيرات التوضيحية من درجة $(n \times p)$.
- B متجة المعالم من درجة $(p \times 1)$.
- t متغير مستمر ويمثل المتغير لامعلمي من درجة $(n \times 1)$.
- $g(t)$ تمثل دالة تمهيدية غير معروفة من درجة $(n \times 1)$.
- ε متجة الاخطاء العشوائية من درجة $(n \times 1)$

في هذا البحث سوف نأخذ المتغير المعتمد **Y** ثنائي الاستجابة اما احتمال ان يساوي واحداً للحصول على الاستجابة او صفراً لعدم حصول الاستجابة وهذا ما يسمى إنموذج الانحدار اللوجستي [2, 3, 12, 13, 16]

لذلك يكون المتغير **Y** يتبع توزيع برنولي والذي يمتلك دالة كثافة احتمالية بالصيغة الاتية :-

$$f(y_i) = p_i^{y_i}(1-p_i)^{1-y_i} \quad ...(2)$$

حيث أن :

- y_i متغير ثنائي الاستجابة (1, 0).
- p_i احتمال حدوث الاستجابة عندما y_i يساوي واحداً.

وعلية فأن أنموذج الانحدار اللوجستي يعطى بالصيغة الاتية

$$y_i = p_i + \varepsilon_i \ , \ i = 1,2,\dots .n \quad ...(3)$$

اذ ان P_i تمثل دالة الانحدار اللوجستي (احتمال الاستجابة) والتي يمكن التعبير عنها بالصيغة الاتية :-

$$P_i = P(y = 1/x, t) = P(XB + g(t)) = \frac{\exp\left(XB+g(t)\right)}{1+\exp\left(XB+g(t)\right)} \quad ...(4)$$

وان q_i تمثل احتمال عدم الاستجابة والتي يمكن التعبير عنها بالصيغة الاتية :-

$$q_i = P((y = 0/x, t) = 1 - P(XB + g(t)) = \frac{1}{1 + exp\left(XB + g(t)\right)} \quad ...(5)$$

اما بالنسبة للحد الخطأ الذي سوف يكون له متوسط صفر كما في الصيغة الاتية .

$$\varepsilon_i = y_i - p_i \quad ...(6)$$

$$E(\varepsilon_i) = E(y_i) - E(p_i) = p_i - p_i = 0 \quad ...(7)$$

أما تباين حد الخطأ فانه يساوي تباين المتغير المعتمد ثنائي الاستجابة .

$$V(\varepsilon_i) = V(y_i) = p_i(1-p_i) \quad ...(8)$$

لذلك فان حد الخطأ له متوسط صفر وتباين $p_i(1-p_i)$.

ولغرض توضيح هذا النوع من النماذج يتم تقدير الجزء المعلمي باستعمال احد طرائق التقدير المعلمية المعروفة وهي طريقة المربعات الصغرى الاعتيادية (OLS) والتي يمكن وضعها كقيم أولية ابتدائية لقيم المعلمات المجهولة وبعدها يتم تقدير الجزء الاخر اللامعلمي على وفق طريقة classic local least estimator .

٢-٣ طريقة المربعات الصغرى الاعتيادية (OLS) [5 ,6 ,14]

أن انموذج الانحدار الخطي الجزئي والذي ناقشه الباحث (Speckman) عام (1988)[14] للدالة اللامعلميه اذ يعمل على وصف المقدر اللامعلمي بالمصفوفة W لتمثل عناصر $kernel$ والتي تكون كامله الرتبة من الدرجة $(n \times p)$ و γ تشير الى موجة المعالم المضافة. ويمكن إعادة كتابة النموذج (1) بالشكل الاتي :

$$Y = X\beta + W\gamma + \varepsilon \qquad \dots (9)$$

ε : وهي الأخطاء العشوائية وتكون مستقلة وذات متوسط صفر وتباين منتهي $E({\varepsilon_i}^2) = \sigma_i^2$.

وعلية يتم تقدير كل من γ و β بطريقة OLS وكما يأتي:

$$\varepsilon'\varepsilon = (Y - X\beta - W\gamma)'(Y - X\beta - W\gamma)$$

$$\frac{\partial\varepsilon'\varepsilon}{\partial\beta'} = -X'\left(Y - X\hat{\beta} - W\hat{\gamma}\right) = 0$$

$$\frac{\partial\varepsilon'\varepsilon}{\partial\gamma'} = -W'\left(Y - X\hat{\beta} - W\hat{\gamma}\right) = 0$$

$$X'X\hat{\beta} = X'(Y - W\hat{\gamma}) \qquad \dots (10)$$

$$W'W\hat{\gamma} = W'\left(Y - X\hat{\beta}\right) \qquad \dots (11)$$

وبضرب المعادلة (11) ب $(W'W)^{-1}$ ينتج:

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

و بضرب طرفي المعادلة في W ينتج

$$W\hat{\gamma} = W(W'W)^{-1}W'\left(Y - X\hat{\beta}\right) \qquad \dots (12)$$

لنفرض إن

$$P_{\omega} = W(W'W)^{-1}W' \qquad \dots (13)$$

والتي تشير مصفوفة التقدير والتي تكون متماثلة و صماء لذلك فتصبح المعادلة (12) كما يأتي :

$$W\hat{\gamma} = P_{\omega}\left(Y - X\hat{\beta}\right) \qquad \dots (14)$$

وبعد تعويض المعادلة (14) في معادلة (10) يصبح لدينا ما يأتي :

$$X'X\hat{\beta} = X'\left(Y - P_{\omega}\left(Y - X\hat{\beta}\right)\right)$$

$$X'X\hat{\beta} = X'Y - X'P_{\omega}Y + X'P_{\omega}X\hat{\beta}$$

$$X'X\hat{\beta} - X'P_{\omega}X\hat{\beta} = X'Y - X'P_{\omega}Y$$

$$X'(I-P_\omega)X\hat{\beta} = X'(I-P_\omega)Y$$

$$\hat{\beta} = (X'(I-P_\omega)X)^{-1}X'(I-P_\omega)Y \qquad \dots(15)$$

وإن P_ω مصفوفة متماثلة وصماء فيمكن كتابة معادلة (15) كما يلي :

$$\hat{\beta} = (X'(I-P_\omega)'(I-P_\omega)X)^{-1}X'(I-P_\omega)'(I-P_\omega)Y$$

ويمكن فرض $\tilde{Y} = (I-P_\omega)Y$ و $\tilde{X} = (I-P_\omega)X$ لذلك يصبح المقدر المعلمي بهذه الصيغة :

$$\therefore\ \hat{\beta}_{LS} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} \qquad \dots(16)$$

٣-٣ مقدر <u>classic local least estimator</u> لتقدير الأنموذج شبه المعلمي [9]

يعد هذا المقدر من المقدرات اللامعلمية والتي تعتمد على متسلسلة الاوزان المقترحة من قبل الباحثين $Murphy\ S.A., Van\ Der\ Vaart$ عام ٢٠٠٠م وتتلخص هذه الطريقة في دوال شبة معلمية على وفق الصيغة الاتية :-

$$g(t,B) = argmin\sum_t W_{ni}(t)(Y_t\,(B)-a)^2 = \frac{\sum_t W_{ni}(t)Y_t\,(B)}{\sum_t W_{ni}(t)}$$

بما أن

$$Y_t(B) = Y_t - BX_t$$

$$g(t,B) = \frac{\sum_t W_{ni}(t)Y_t}{\sum_t W_{ni}(t)} - B\,\frac{\sum_t W_{ni}(t)X_t}{\sum_t W_{ni}(t)} = G(T) - BH(T) \qquad \dots(17)$$

حيث ان

$$H(T) = \frac{\sum_t W_{ni}(t)X_t}{\sum_t W_{ni}(t)}G(T) = \frac{\sum_t W_{ni}(t)Y_t}{\sum_t W_{ni}(t)} \quad ;$$

أذ ان $W_{ni}(t)$ تشير الى سلسلة الاوزان وان دوال الوزن هذه تكون طبيعية اذا حققت الشرط الأتي [17] :

$$n^{-1}\sum_t W_{ni}(t) = 1$$

وتكون غير سالبة اذا كانت دالة الوزن $W_{ni}(t) > 0$، وتكون دالة الوزن احتمالية اذا كانت غير سالبة وتكاملها مساوٍ للواحد .

توصف دالة الوزن $W_{ni}(t)$ بواسطة دالة كثافة مع معلمة القياس التي تعدل حجم وشكل الاوزان القريبة من (t) والشائع هو الاشارة لشكل الدالة ب K (Kernel) وتعرف دالة الوزن لتقديرات (Kernel) بالصيغة الاتية :-

$$= \frac{K(u)}{\sum K(u)} \qquad \dots(18)$$

$$W_{ni}(t) = \frac{K\left(\frac{(t-T)}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{t-T_j}{h}\right)}$$

أذ ان $k(.)$ تشير الى درجة اللب (Kernel) والتي يمكن الحصول عليها من دالة الكثافة الطبيعية القياسية أو ما يطلق عليها بـ (Gaussian Kernel) وكالاتي :-

$$k(.) = (2\pi)^{-0.5} Exp\left(-\frac{u^2}{2}\right) \qquad ...(19)$$

حيث تمثل T متغير لامعلمي اما (h) فتمثل عرض الحزمة،في هذا المقدر سوف نستخدم قاعدة حل المعادلة (Solve The Equation Rule) لتقدير عرض الحزمة.

٣-١-٣ قاعدة حل المعادلة (Solve The Equation Rule)

اقترحت هذه الطريقة من قبل $(Jones\ et\ al.)$عام ١٩٩٦م حيث هذه الطريقة تختلف عن طريقة الملئ المباشرة في الاسلوب حيث تستند هذه الطريقة الى تقليل $(AMISE)$ للحصول على المعلمة التمهيدية المثلى وكما يأتي[8, 4, 1] :-

$$h_{AMISE} = \left[\frac{R(K)}{M_k^2(K)\ \Psi_{r+k}[\gamma(\mathrm{h})]\ n}\right]^{1/(r+k+1)} \qquad ...(20)$$

حيث أن تم أستعمال دالة (Gaussian Kernel) لايجاد كل من $M_k^2(K)$ و $R(K)$ وكالاتي [4, 1] :

$$M_k^2(K) = d_K^2 = 1$$

$$R(K) = C_K = \frac{1}{2\sqrt{\pi}}$$

أن المعادلة (٢٠) تستوجب منا ايجاد $\gamma(\mathrm{h})$ ويتم حسابها كما يأتي[15, 4, 1]:

$$\gamma(\mathrm{h}) = \left[\frac{-2\ L^r(0)\ M_k^2(K)\hat{\Psi}_r(g_m)}{R(K)\ M_k(L)\ \hat{\Psi}_{r+2(\ell-1)}(g_{m+1})}\right]^{1/(r+3)} .h^{(r+1)/(r+k+1)}. \qquad (21)$$

حيث ان $m = 1$ ، وأن قيمة المعلمة التجريبية يتم حسابها كما يأتي :-

$$g = \left[\frac{-k!\ L^r(0)}{M_k(L)\ \hat{\Psi}_r^{NS}\ \mathrm{n}}\right]^{1/(r+k+1)} \qquad ...(22)$$

حيث أن دالة **Kernel** متماثلة وان $(k = 2,4, ...)$ وتمتلك r من المشتقات.

حيث يتم حساب $\hat{\Psi}_r^{NS}$ وفق الصيغة التالية :-

$$\hat{\Psi}_r^{NS} = \frac{(-1)^{\frac{r}{2}}\ r!}{(2\ \sigma)^{r+1}\ \left(\frac{r}{2}\right)!\ \sqrt{\pi}} \qquad ...(23)$$

وان $L^r(0)$ تمثل سلسلة تايلر للمتغيرات التمهيدية .

خوارزمية (Solve The Equation Rule):

ان الخطوات المتبعة لمرحلتين $(l = 2)$ لذلك نحتاج لايجاد (g_2, g_1) وبأستعمال دالة **(Gaussian)** وكما يأتي :-

١.نقوم بحساب قيمة كل من $\hat{\Psi}_6^{NS}$ و $\hat{\Psi}_8^{NS}$ و وفق قاعدة التوزيع الطبيعي وحسب المعادلة (٢٣) وتبلغ قيمتهما على التوالي كالأتي:

$$\hat{\Psi}_6^{NS} = \frac{-15}{16\sqrt{\pi}\,(\sigma)^7} \qquad \dots(24)$$

$$\hat{\Psi}_8^{NS} = \frac{105}{32\sqrt{\pi}\,(\sigma)^9} \qquad \dots(25)$$

٢.نقوم بحساب قيمة g_1 و g_2 على التوالي وفق المعادلة (22) وكما يأتي:

$$g_1 = \left[\frac{-2!\;\; L^4(0)}{M_2(L)\;\hat{\Psi}_6^{NS}\;\; n}\right]^{1/7} \qquad \dots(26)$$

حيث ان

$$L^4(0) = \frac{3}{\sqrt{2\pi}} \qquad \dots(27)$$

$$g_2 = \left[\frac{-2!\;\; L^6(0)}{M_2(L)\;\;\hat{\Psi}_8^{NS}\;\; n}\right]^{1/9} \qquad \dots(28)$$

حيث ان

$$L^6(0) = -\frac{15}{\sqrt{2\pi}} \qquad \dots(29)$$

٣.نقوم بحساب قيمة $\hat{\Psi}_4(g_1)$ و $\hat{\Psi}_6(g_2)$ على التوالي وكما يأتي:

$$\hat{\Psi}_4(g_1) = \frac{1}{n(n-1)\,g_1^5}\sum_{i=1}^{n}\sum_{j=1}^{n} L^4\left(\frac{X_i - X_j}{g_1}\right) \qquad \dots(30)$$

$$\hat{\Psi}_6(g_2) = \frac{1}{n(n-1)\,g_2^7}\sum_{i=1}^{n}\sum_{j=1}^{n} L^6\left(\frac{X_i - X_j}{g_2}\right) \qquad \dots(31)$$

٤.نقوم بحساب قيمة $\gamma(\mathrm{h})$ وفق المعادلة وكما يأتي:

$$\gamma(\mathrm{h}) = \left[\frac{-2!\;\; L^4(0)\;\; M_2^2(K)\;\; \hat{\Psi}_4(g_1)}{R(K)\;\; M_2(L)\,\hat{\Psi}_6(g_2)\;\;\; n}\right]^{1/7} .\;\; h^{5/7} \qquad \dots(32)$$

٥.نقوم بحساب قيمة المعلمة التمهيدية حسب طريقة (STE) وفق المعادلة (20):

$$\hat{h}_{\mathrm{STE}} = \left[\frac{R(K)}{M_2^2(K)\;\; \hat{\Psi}_4\big(\gamma(\mathrm{h})\big)\;\; n}\right]^{1/5} \qquad \dots(33)$$

علماً ان:

$$\widehat{\Psi}_4(\gamma(h)) = \frac{1}{n(n-1)\,(\gamma(h))^5}\sum_{i=1}^{n}\sum_{j=1}^{n} L^4\left(\frac{X_i - X_j}{(\gamma(h))}\right) \qquad \dots(34)$$

$$L^4(u) = \frac{u^4}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2} - \frac{6\,u^2}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2} + \frac{3}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$$

$$L^6(u) = \frac{u^6}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2} - \frac{15\,u^4}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2} + \frac{45\,u^2}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2} - \frac{15}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$$

$$R(K) = \frac{1}{2\sqrt{\pi}} \qquad , \quad M_2^2(K) = 1$$

٣-٤ الطريقة البيزية لتقدير أنموذج الانحدار اللوجستي شبه المعلمي [10]

اقترحت هذه الطريقة من قبل الباحث (Peter J. Lenk) عام ١٩٩٩ م والذي اعتمد على نموذج الانحدار شبه المعلمي في المعادلة رقم (١) وكما يلي :-

$$y_i = X\beta + g(t) + \varepsilon_i \qquad ,i = 1, \dots, n$$

افترض الباحث ان المركبة اللامعلمية $g(t)$ تتبع سلسله فورير $(Fourier\ series)$ وكما يأتي :

$$g(t) = \sum_{k=1}^{\infty} \theta_k \emptyset_k(t) \qquad a < t < b \qquad \dots(35)$$

حيث ان

$$\emptyset_k(t) = \left(\frac{2}{b-a}\right)^{1/2} cos\left\{\pi k\left(\frac{t-a}{b-a}\right)\right\} ; k = 1, \dots, a < t < b \qquad \dots(36)$$

$$\theta_k = \int_a^b g(t)\emptyset(t)dt \qquad \dots(37)$$

ولغرض تقدير معالم ألانموذج بأسلوب بيز يتطلب تحديد التوزيع الاولي (Prior dist.) لمعالم الأنموذج، حيث ان التوزيع الاولي لـ(θ_k) يتبع التوزيع الطبيعي وكما يأتي :-

$$\theta_k \sim N(0, \tau^2 \exp(-\gamma c_k))$$

حيث ان : τ^2 تمثل معلمة عدم اليقين لداله (g) والتي تحدد عن طريق المفاضله بين التوزيع السابق والامكان الاعظم .

γ: تحدد عن طريق اضمحلال او تفكك معاملات فورير (*Fourier coefficients*)

c_k : تمثل مؤشر لسلسلة فورير (Fourier series) والتي تساوي :-

$$c_k = K \quad ,K = 1, 2, \dots, n-1 \qquad \dots(38)$$

في هذه الطريقة سيتم الاعتماد على اربعة معالم والتي يجب تحديد التوزيع الاولي لهم وهي (β , σ^2 ، τ^2 , γ) وان التوزيع الاولي للمعالم الاربعة هو :-

$$\beta \sim \mathrm{N}\,(\mathrm{bo}, \beta \mathrm{o})$$
$$\sigma^2 \sim \mathrm{beta}\,(\mathrm{a}, \mathrm{b})$$
$$\tau^2 \sim \mathrm{IG}\,(\mathrm{u0}/2, \mathrm{vo}/2)$$
$$\gamma \sim EXP\,(\mathrm{Wo})$$

وبعد فرض التوزيع الاولي للمعالم تم حساب التوزيع اللاحق وكما يأتي :-

$$X_{\mathrm{k}} = \gamma^{\mathrm{i}} + \frac{1}{\mathrm{c_k}} Ln\left[1 - 2\left(\frac{\tau}{\theta_{\mathrm{k}}}\right)^2 \exp\left(-\mathrm{c_k}\gamma^{\mathrm{i}}\right)\ln(u_k)\right] \qquad \ldots(39)$$

حيث ان:-

γ تم فرض قيمه اوليه ومساويه للواحد الصحيح وتمثل الجزء التكراري للمعادلة المذكورة آنفاً.

$\mathrm{c_k}$: يمثل مؤشر لسلسله فورير في المعادله (٣٨) .

τ^2 : قيمة سحبت عشوائياً عن طريق توزيع كاما المعكوس بمعالم (١ ، ١٠).

u_k : متجه عشوائي سحب عشوائياً عن طريق توزيع المنتظم القياسي(**Standard uniform**).

θ_{k} : يتم حسابها عن طريق المعادله الاتية :-

$$\theta_{\mathrm{k}} = n^{-1}\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{y_i} - \mathrm{X}'\beta)\emptyset_k(t) \qquad \ldots(40)$$

بعد حساب X_{k} يتم حساب γ التكرارية من خلال طريقة اسلوب (**Markov Chain Monte Carlo**) وكما يأتي :-

$$\gamma^{\mathrm{i}+1} = -\frac{1}{\mathrm{w_k}} Ln[\exp(-w_k c) - u\{\exp(-w_k c) - exp(-w_k X)\}] \qquad \ldots(41)$$

حيث أن

$w_k = wo - 0.5\sum_{i=1}^{n} c_k$ وان wo تمثل معلمة التوزيع الاسي والتي تمثل متوسط البيانات لامعلمية

X: تمثل $X = \min(X_k)$

u: تمثل قيمة واحدة سحبت عشوائياً عن طريق توزيع القياسي المنتظم .

المعادلات (41) و (39) استعملت لتمهيد البيانات، وبعد تمهيد البيانات سيتم تقدير داله اللامعلمية عن طريق داله (Kernel) وكما يأتي :-

$$\tilde{g}(\mathrm{t}) = \mathrm{n}^{-1}\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{y_i} - \mathrm{X}_{\mathrm{i}}'\beta)\mathrm{C_{k,n}}(x_{\mathrm{i}}, \mathrm{t}) \qquad \ldots(42)$$

حيث أن $\mathrm{C_{k,n}}$ تمثل دالة ($kernel$) وتساوي :-

$$\mathrm{C_{k,n}}(t_{\mathrm{i}}, \mathrm{t}) = \sum_{\mathrm{k}=1}^{\mathrm{K}} w_{n,k}\emptyset_k(u)\emptyset_k(t) \qquad \ldots(43)$$

$$w_{n,k}=\frac{n(b-a)\tau^2\exp(-\gamma c_k)}{(b-a)\sigma^2+n\tau^2\exp(-\gamma c_k)} \quad \ldots(44)$$

بعد حساب الدالة اللامعلمية في معادله (42) يتم بعدها حساب معالم الجزء المعلمي (β) كون قيم المعالم (β) التي حسابها قبل تقدير دالة الجزء اللامعلمي كانت معالم أولية وحسبت من خلال مصفوفة الاوزان للمتغير اللامعلمي، في المعادلة ادناه سيتم حساب معالم الجزء المعلمي بعد استبعاد اثر المتغير اللامعلمي وكما يأتي :-

$$\hat{\beta}=(x'x)^{-1}x'\left(y-g(t)\right) \quad \ldots(45)$$

٤ الجانب التجريبي

٤-١ المقدمة (Introduction) :-

يعد التحليل بأستعمال المحاكاة وسيلة رياضية لحل الكثير من المعادلات والعلاقات الرياضية ، اذ إن هناك العديد من الحالات التي لا يمكن تمثيلها رياضيا اما بسبب الطبيعة العشوائية للمسألة المدروسة، أو بسبب تعقيد صياغتها،أو نظرا للتفاعلات اللازمة لوصف المسألة قيد الدراسة وصفا دقيقا.وفي جميع الحالات التي تستعصي على الصياغة الرياضية،تعد المحاكاة الأداة الوحيدة التي يمكن استعملها للحصول على إجابات مناسبة.

ويمكن تعريف المحاكاة على أنها تقنية عددية تستعمل للقيام باختبارات على حاسوب عددي، و تتضمن علاقات منطقية و رياضية تتفاعل فيما بينها لتصف سلوك و بنية منظومة معقدة في العالم الحقيقي على امتداد فترة من الزمن.

على الرغم من أنه ينظر في بعض الأحيان إلى المحاكاة على أنها الطريقة التي غالبا ما تستخدم عند فشل اية الأساليب الأخرى ، فإن التقدم الذي حدث مؤخرا في أساليب المحاكاة و توافر البرمجيات و التطورات التقنية قد جعلت من المحاكاة أحد أكثر الأدوات المقبولة و المستخدمة بشكل واسع في تحليل النظم .

تم استعمال بعض الدوال الجاهزة والدوال البرمجية في برنامج الـ(**Matlab 2014a**) في توليد البيانات وبناء نماذج المحاكاة لغرض المقارنة بين الطرائق بأختلاف احجام العينات والتباينات .

٤-٢ توليد المتغيرات العشوائية Generating Random Variables

تم تنفيذ تجارب المحاكاة باستعمال ثلاثة حجوم للعينات$(n=30,60,90)$ وبتكرارات(**Replicates** = **100**) لكل تجربة وكما يأتي :-

١- يتم توليد المتغيرات التوضيحية X_i على وفق التوزيع الطبيعي والمتغيرt_i على وفق التوزيع المنتظم وباستعمال الدوال الجاهزة والموجودة في برنامج $Matlab$ وكما يأتي :-

$$X_1=normrnd(\mu_1,\sigma_1)$$
$$X_2=normrnd(\mu_2,\sigma_2)$$
$$t=unifrnd(\min(t),\max(t))$$

٢- الاخطاء العشوائية تتوزع بمتوسط صفر وتباين (pq) حيث يتم حساب قيمة (p) وفق المعادلة (٤) ، وقد تم أفتراض ثلاث قيم لتباين الخطأ وهي$(\sigma=0.2\,,0.5\,,0.8)$

٣- المتغير المعتمد (Y) سيتم توليدها باستعمال دالة الانحدار شبه المعلمي بدلالة المتغيرات التوضيحية التي تم توليدها من الفقرة (1) مضافاً اليها حد الخطأ الذي تم توليدة في الفقرة (2) .

٣-٤ الأنموذج المستعمل في المحاكاة Simulation Model

تم استعمال أنموذج شبه المعلمي الجزئي (PLM) لصياغة المركبات اللامعلمية وكما يأتي :-

$$Y = logit\left(\frac{p_i}{1-p_i}\right) = XB + g(t_i) + \varepsilon$$

حيث أن :-

$$X = (1, X1, X2)\ ;\ \beta = (\beta_0, \beta_1, \beta_2)\ \ ; g(t_i) = \left(\ g(t_1), g(t_2), \dots, g(t_n)\right)$$

وتكون المركبة اللامعلمية على وفق مقدرات الطرائق التي ذكرت في المعادلات (17) و (42) .

٤-٤ تحليل نتائج المحاكاة :

من الجدول المرقم (٣-١) نلاحظ ان

عند حجم عينة ٣٠ اظهرت النتائج ان المقدر (LLE) هو الافضل يليه المقدر البيزي بتباين ٠.٢ و ٠.٥ ، اما عند تباين ٠.٨ تبين ان المقدر البيزي هو الافضل يليه المقدر (LLE).

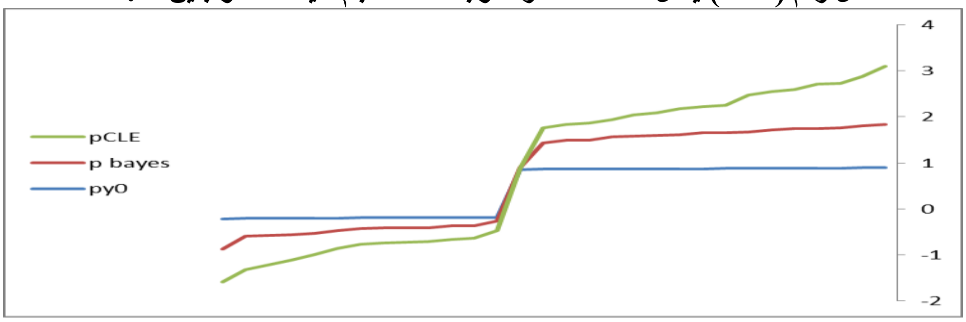
عند حجم عينة ٦٠ اظهرت النتائج ان المقدر (LLE) هو الافضل يليه المقدر البيزي بتباين ٠.٢ و٠.٥ و ٠.٨.

عند حجم عينة ٩٠ اظهرت النتائج ان المقدر (LLE) هو الافضل يليه المقدر البيزي بتباين ٠.٢ و٠.٥ و٠.٨.

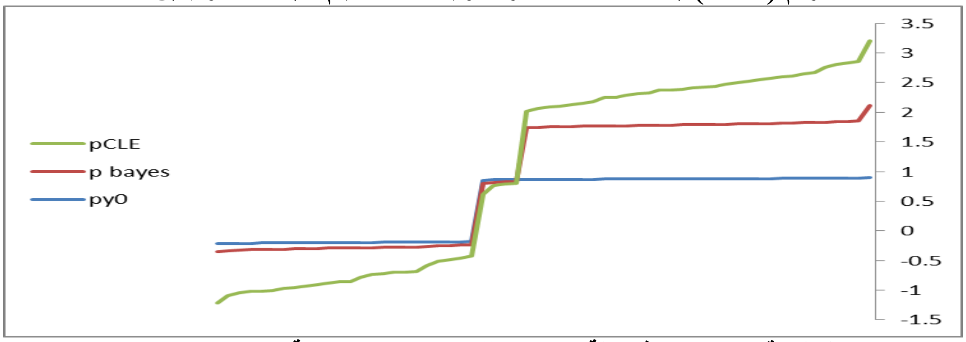
الجدول رقم (٣-١) الخاص بقيم MSE للطرائق التقدير

n الطرائق	σ	٣٠	٦٠	٩٠
Bayes	٠.٢	٠.١٥١٢	٠.١٨١٠	٠.١٣٦٤
	٠.٥	٠.١٦١٠	٠.١٦٥٦	٠.١٣٧٩
	٠.٨	٠.١٥١٣	٠.١٥٩٦	٠.١٣٦٤
LLE	٠.٢	٠.١٤٤٨	٠.١٣٨٨	٠.١٣٢٥
	٠.٥	٠.١٦١٠	٠.١٣٩٧	٠.١٣٦٦
	٠.٨	٠.١٥٥٤	٠.١٣٢٧	٠.١٣٢٥

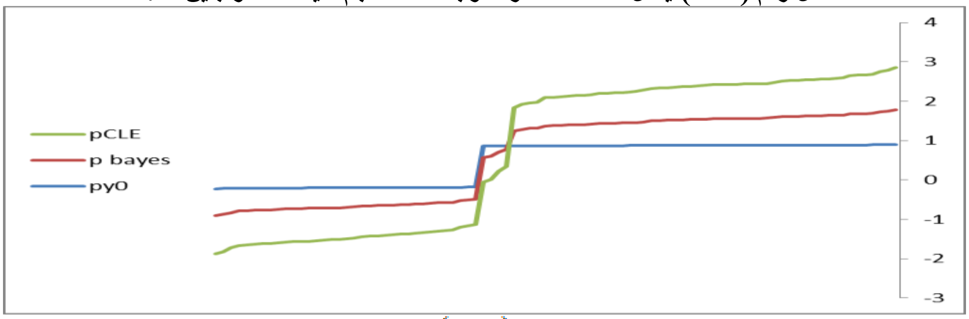
شكل رقم (٣-١) يمثل دالة الانحدار اللوجت عند حجم عينة ٣٠ وتباين ٠.٢ *



شكل رقم (٣-٢) يمثل دالة الانحدار اللوجت عند حجم عينة ٦٠ وتباين ٠.٢

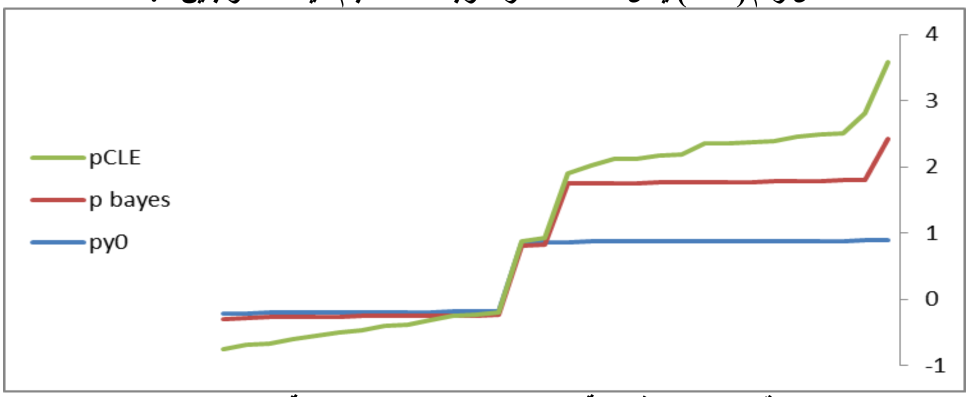


شكل رقم (٣-٣) يمثل دالة الانحدار اللوجت عند حجم عينة ٩٠ وتباين ٠.٢

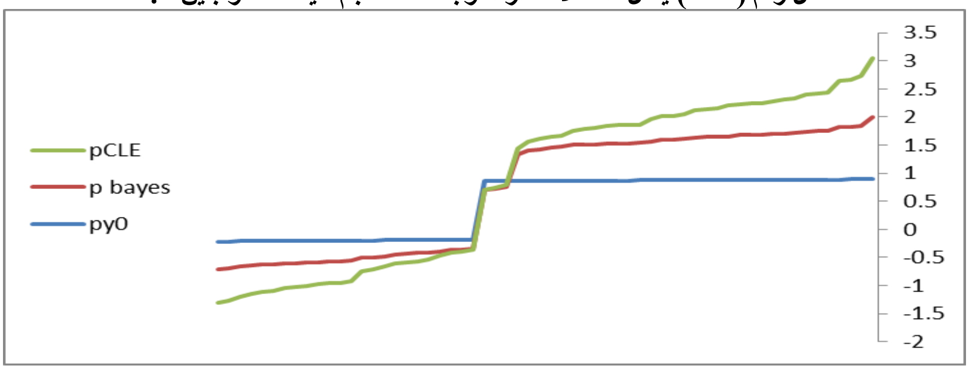


- هنا تم رسم دالة الانحدار اللوجت والتي تساوي $logit\left(\frac{p_i}{1-p_i}\right)$

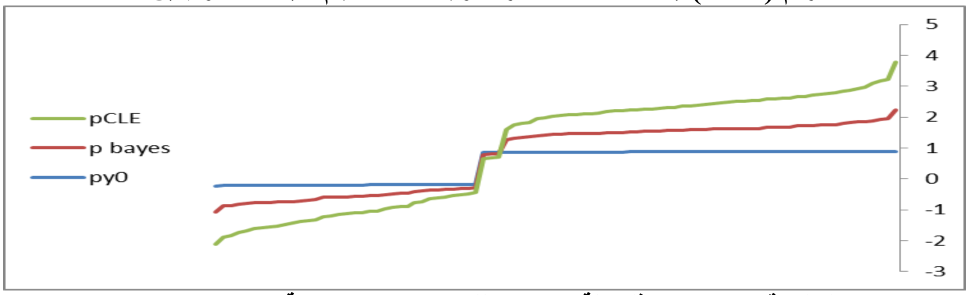
شكل رقم (٣-٤) يمثل دالة الانحدار اللوجت عند حجم عينة ٣٠ وتباين ٠.٥



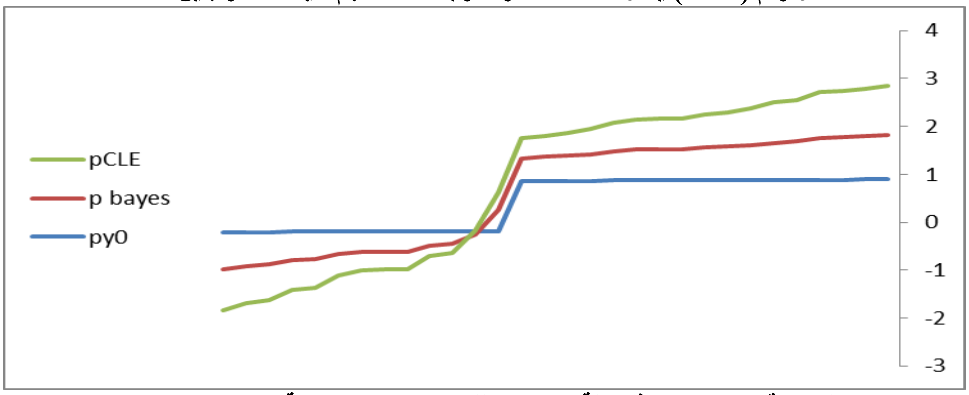
شكل رقم (٣-٥) يمثل دالة الانحدار اللوجت عند حجم عينة ٦٠ وتباين ٠.٥



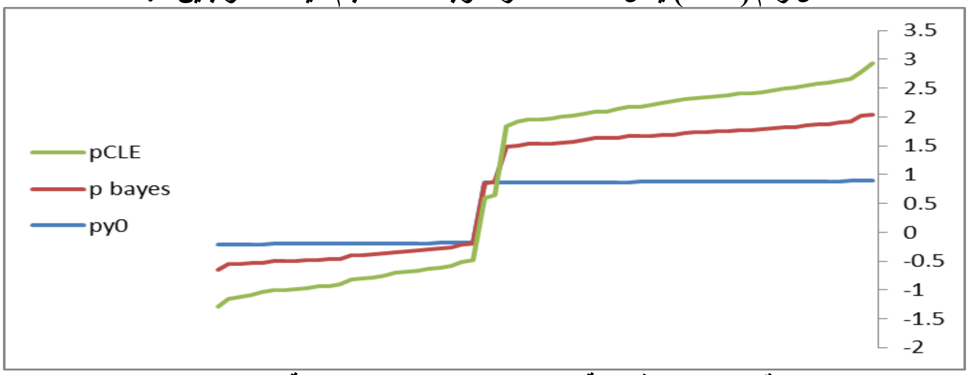
شكل رقم (٣-٦) يمثل دالة الانحدار اللوجت عند حجم عينة ٩٠ وتباين ٠.٥



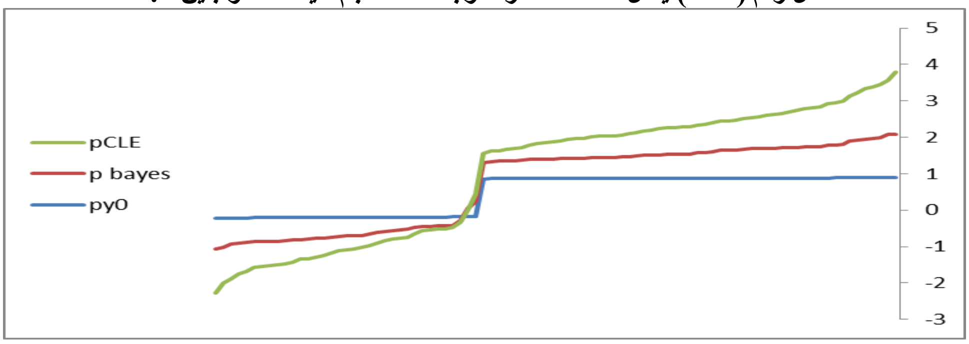
شكل رقم (٣-٧) يمثل دالة الانحدار اللوجت عند حجم عينة ٣٠ وتباين ٠.٨



شكل رقم (٣-٨) يمثل دالة الانحدار اللوجت عند حجم عينة ٦٠ وتباين ٠.٨



شكل رقم (٣-٩) يمثل دالة الانحدار اللوجت عند حجم عينة ٩٠ وتباين ٠.٨



٥- الاستنتاجات و التوصيات

٥-١ الاستنتاجات

بعد تنفيذ تجارب المحاكاة وما تم عرضة من نتائج وتحليل في الجانب التجريبي أستنتج الباحث مايأتي :-

١- في اغلب تجارب المحاكاة وبأختلاف احجام العينات وباختلاف التباينات نلاحظ ان المقدر(LLE) هو الافضل .

٢- أشارت نتائج المحاكاة ان المقدر البيزي يكون أفضل في حالة أحجام العينات الصغيرة .

٥-٢ التوصيات

على ضوء الاستتناجات التي توصلنا اليها من خلال البحث يمكن اجمال التوصيات الاتية :-

١- استعمال مقدر البيزي عند التحليل في أنموذج الانحدار اللوجستي شبه المعلمي في حالة احجام العينات الصغيرة ، لما تبديه من كفاءة وسرعة ومرونة عالية في التطبيق

٢- استعمال مقدر(LLE) في حالة احجام العينات الكبيرة .

٣- تقدير نماذج شبه المعلمية بأستعمال شرائح التمهيد (Smoothing Spline) في أنموذج الانحدار اللوجستي شبه المعلمي .

٦- المصادر

١- أبراهيم، مروة خليل ٢٠١٣م ،"تقويم بيانات العمر والجنس للتعدادات السكانية مع تطبيق عملي لبيانات التعداد العام للسكان لسنة ١٩٩٧م في العراق" رسالة ماجستير في الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد .

٢- البلداوي، تسنيم حسن ١٩٩٦م "مقارنة تحليلية بين انموذج الانحدار اللوجستي ونماذج الدوال التميزية" أطروحة دكتوراه في الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد.

٣- حسين، شيرين علي ٢٠٠٩م ،"مقدرات الإمكان الأعظم الموزونة الحصينة ومقارنتها مع طرائق أخرى لانموذج اللوجستك مع تطبيق عملي" رسالة ماجستير في الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد.

٤- حمود، مناف يوسف، ٢٠٠٥م،"مقارنة المقدرات اللامعلمية لتقدير دوال الكثافة الاحتمالية" اطروحة دكتوراه فلسفة في الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد .

٥- عيسى، أسيل مسلم، ٢٠١١م ،" مقارنة بعض المقدرات شبه المعلمية لتقدير دالة استهلاك الطاقة الكهربائية لمدينة بغداد " رسالة ماجستير في الاحصاء، كلية الادارة والاقتصاد ، جامعة بغداد

٦- كاطع، مياسة محمد، ٢٠١٤م، " مقارنة النماذج اللامعلمية وشبه المعلمية بوجود القيم المفقودة مع تطبيق عملي للناتج المحلي الاجمالي العراقي للمدة (١٩٧١-٢٠١٠)م" رسالة ماجستير في الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد .

7- **Hardle, W. ,Laing ,H. ,Gao.J. ,2000 , "Partially Linear Models" http:/cdn.preterhuman.net.**
8- **Molanes, Lopez , E.M. ,Cao ,R. ,2008 , "Plug-in Bandwidth Selector for the Kernel Relative Density Estimator " Annals of the Institute of Statistical Mathematics AISM , page 273-300 .**
9- **Murphy,S.A. Van Der Vaart , A.W. 2002 " On Profile Likelihood " Journal of the American Statistical Association 95 , Chapter 6 , page 43-57.**
10- **Peter J.Lenk, 1999, " Bayesian Inference For semiparametric Regression Using a Fourier Representation " J.R. Statist Soc. Part 4 page 863-879.**
11- **P. M .Robinson , (1988) ," Root-N consistent semiparametric regression" , Econometrical 56 , No. 4 , 931_954.**
12- **Shaefer, R.L 1979, "Multicollinearty and logistic regression", ph.D. dissertation , university of Michigan , USA.**
13- **Shen, J & Gao, S, 2008 , "A Solution to Separation and Multicollinearity in Multiple Logistic Regression" , Indiana University School of Medicine , Journal of Data Science 6 , page 515-531.**
14- **Speckman,P., 1988, "Kernel Smoothing in Partial Linear Models" J.R.Statist. Soc.50 , No 3 , page 413-436.**
15- **Tenreiro, C.,2011, "ombining Cross-validation and Plug-in Methods for Kernel Density Bandwidth Selection" CMUC and DMUC, University of Coimbra.**
16- **Xiangjin Shen, Shiliang Li, Hiroki Tsurumi, 2013, "Comparison of Parametric and Semiparametric Binary Response Model " Jstor.**
17- **Yatchew,A., 2003, "Semiparametric Regression for the applied econometrician " Cambridge University Press.**

A Comparison Between Classic Local Least Estimatop And Bayesian Methoid For Estimating Semiparametric Logistic Regression Model

Abstract :

Semi-parametric models analysis is one of the most interesting subjects in recent studies due to give an efficient model estimation. The problem when the response variable has one of two values either 0 (no response) or one – with response which is called the logistic regression model.

We compare two methods Bayesian and local least estimator. Then the results were compared using MSe criteria.

A simulation had been used to study the empirical behavior for the Logistic model , with different sample sizes and variances. The results using represent that the Bayesian method is better than the local least estimator at small samples sizes.

Key word /logistic Regression - semiparametric model- (LLE)- Bayesian - Kernel function- bandwidth.