مقارنة مقدرات أنموذج شبه معلمي بأستعمال طرائق تمهيد مختلفة

أ.م.د. مناف يوسف حمود   مياسة محمد كاطع

قسم الاقتصاد/كلية الادارة والاقتصاد/جامعة بغداد

المستخلص :

في هذا البحث تم تقدير دالة الانحدار الخطي الجزئي شبة المعلمي مستعملين بذلك طرائق معلمية ومتمثلة بطريقة OLS ولامعلمية متمثلة بطريقة شرائح التمهيدsmoothing spline ومتمثلة بمقدر الشريحة التكعيبيةcubic smoothing spline ومقدر لبي kernel متمثل بمقدر Nadaraya_Watson ان هدف هذا البحث تتمثل بمقارنة تلك المقدرات مع مقدر لامعلمي لبي ومقدر خطي بأستعمال اسلوب المحاكاة وبدراسة ثلاث نماذج انحدار لامعلمي وبحجوم عينات n=40,60,100 وقيم تباين ٠.٥,١,١.٥ = σ^2 وقد اظهرت نتائج النموذج الاول أن افضل مقدر هو مقدر N.W لنموذج الانحدار الخطي الجزئي شبه المعلمي يليه مقدر C.S.S لنموذج الانحدار الخطي الجزئي شبه المعلمي ، اما نتائج النموذج الثاني والثالث فأشارت الى ان أفضل مقدرات هو مقدر شرائح التمهيد التكعيبية C.S.S لنموذج الانحدار الخطي الجزئي شبه المعلمي يليه مقدر N.W شبه المعلمي ، كما اشارت النتائج أن أقل المقدرات اداءاً وتمثيلاً للنماذج المستعملة هو المقدر المعلمي OLS الذي يفرض ان جميع المتغيرات التوضيحية في النموذج تسلك سلوك خطي (معلمي) لجميع قيم تباين البواقي وحجوم العينات المستعملة يليه المقدر اللامعلمي البي N.W الذي يفرض ان جميع المتغيرات في النموذج تسلك سلوك لامعلمي .

المصطلحات الرئيسية للبحث : **الانحدار شبه معلمي- الانحدار الخطي الجزئي- شرائح التمهيد- التمهيد اللبي- مقدر اللبي اللامعلمي .**

مجلة العلوم الاقتصادية والإدارية
المجلد 20
العدد 75
لسنة ٢٠١٤
الصفحات ٣٧٦- ٣٩٤

***ملاحظة البحث مستل من رسالة ماجستير**

1-١ المقدمة

أن الهدف الرئيس من تحليل الانحدار يتمثل باختزال البيانات المشاهدة أو تلخيصها بما يضمن عرضها للعلاقة بين كل من المتغيرات التوضيحية ومتغير الأستجابة وأن تحليل الانحدار الخطي يعطي تصورا تقريبيا لتلك العلاقة من خلال رسم أو عرض تلك العلاقة وفق اتجاه خطي تقريبي، فضلا عن هذا يعد الهدف الرئيس الثاني بأمكانية استعمال دالة الانحدار لأغراض التنبوء من خلال أعطاء قيمة للمتغيرات التوضيحية وبالاعتماد على تقدير المعلمة المجهولة يتم تقدير القيمة التوقعية للمشاهدة الجديدة Y عند النقطة X من خلال المعادلة $\hat{a} + \hat{b}X$ اذ تشير $\hat{a}$ و $\hat{b}$ الى مقدرات المعالم b،a على التوالي .

لكن في الجوانب التطبيقية يلاحظ أن أغلب البيانات تتكون من أكثر من متغير توضيحي يؤثر على متغير الاستجابة لذلك لايكون من الملائم استعمال الانحدار الخطي البسيط مما قد يعزوا هذا الى استعمال أنموذج الانحدار الخطي المتعدد علما أن هذا النوع من النماذج لها عيوب مختلفة منها المعرفة المسبقة بتوزيع البيانات المدروسة وبشكل واضح تكون المعلمات معرفه كما انه في بعض الأحيان قد لا يمثل الدالة قيد الدراسة تمثيلاَ كاملا وذلك لكون إن بعضاً من المتغيرات تسلك سلوكاً معلمياً والأخر لامعلمياً.وأن عملية أختيار المتغيرات الخاصة بالدراسة قد يتم وفق رؤى معينة حول حالة معينة ولكن عملية البناء هذه أوما تسمى بعملية النمذجة قد لاتأخذ بنظر الاعتبار مسائل أخرى كمسائل تعدد الابعاد للدراسة فضلا عن كون الانموذج المستعمل قد يعاني من مشاكل معينة كمشكلة عدم تجانس التباين مما يضطر عندها كثير من الباحثين اللجوء الى مايسمى الانحدار اللامعلمي الذي له مرونة في التعامل مع حالات عدم الخطية وكذلك وجود عدم تجانس في التباين لكنه مع هذا يعاني من مشاكل تعدد الابعاد The curse of dimensionality التي تحصل عند زيادة عدد المتغيرات ومن ثم جعل من دراسة متغيرات عديدة وأخذها بالدراسة مسألة معقدة ، هذه المشاكل قادت الباحثين الى التفكير في التعامل مع مايسمى بالنماذج شبه المعلمية والتي كانت هناك دراسات كثيرة على هذا النوع من النماذج فضلا عن النظريات التي بنيت حوله .
اما الشكل العام لهذا النوع من النماذج كما مبين في ادناه:

$$Y_i = X_i{}'\beta + g(t_i) + \varepsilon_i$$

يعد أول ظهور لمصطلح semiparametric من قبل الباحثين Santner,Cail,&Brown عام ١٩٨٠ في مجال الاحصاءات الحيوية Biometric وفي عام ١٩٨١ تم استعمال هذا المصطلح من قبل Whitehead في كتابه في مجال الرياضيات وأطلق علية مصطلح Partially parametric وفي نفس العام تم استعمال هذا المصطلح من قبل Finnas&Hoen في مجال الديمغرافي Demography..[24]
وفي عام (1988) قام الباحث (Speckman)[22] بدراسة ممهد اللب في النماذج الخطية الجزئية، في عام (1994) قام الباحثان (Green & Silverman)[9] بعرض مثال حول استعمال النماذج الخطية الجزئية وقد قارنا النتائج مع الانحدار المعلمي، وفي ١٩٩٩ قام الباحث Aydin [٤] باجراء مقارنة بين طريقتين للتمهيد هما ممهد شرائح التمهيد التكعيبية cubic smoothing spline وبين ممهد Nadaraya-Watson .وفي عام ٢٠٠٣ قام الباحثون Millimet,List,&Stengos [١٦] باجراء مقارنة بين أنموذج معلمي و شبه المعلمي عند دراستهم لمشكلة تلوث الهواء في الولايات المتحدة. في عام ٢٠١١ قام الباحث Aydin [٦] بتقديم بحثا درس فيه طرائق مختلفة لتقدير معلمة التمهيد smoothing parameters لأنموذج الانحدار الخطي الجزئي شبة المعلمي PLM عند تقدير النموذج باستعمال شرائح التمهيد التكعيبية .

٢-١ هدف البحث

تهدف البحث الى تقدير دالة الانحدار الخطي الجزئي شبة المعلمي مستعملين بذلك طرائق معلمية ومتمثلة بطريقة OLS و لامعلمية متمثلة بـ

١. طريقة شرائح التمهيدية **smoothing splines** ومتمثلة بمقدر شريحة تكعيبي **cubic smoothing spline** .
٢. طريقة اللب **kernel** والمتمثلة بمقدر **Nadaraya _Watson** .
ثم مقارنة المقدرات المذكورة انفا مع مقدر لامعلمي لبي باعتبار ان المتغيرات تسلك سلوكا لامعلميا و مقدر انحدار معلمي خطي متعدد.

٢- الجانب النظري

٢-١ أنموذج الانحدار شبه المعلمي:

يعد انموذج الانحدار الخطي الجزئي **partial linear regression model** أحد نماذج الانحدار شبه المعلميه[15] ويرمزله بـ **PLM** وهو من نماذج الانحدار التي تعتمد على متغيرات خطية **Linear** وأخرى غير خطية لامعلميه **Nonparametric** وعادة تكون متغيرات مستمرة **continuous explanatory variables** وتوثر هذه المتغيرات الخطية واللاخطية في متغير الاستجابة**Y** ، وهو تعميم لتقنيات الانحدار الخطي القياسي ويعد حالة خاصة من النماذج التجميعية **Additive models** لذلك يكون أفضل من النماذج اللامعلميه لانه يتجنب مشكلة البعدية **The curse of dimensionality** التي تحدث في النماذج اللامعلميه عند زيادة عدد المتغيرات ومن ناحية أخرى هو أكثر مرونة من نماذج الخطية القياسية لانها تقلل من الافتراضات الخطية المفروضة على هذه النماذج .[20]
ويمكن تمثيل أنموذج الانحدار شبه المعلمي بالصيغة الاتية:

$$\underline{Y} = X\underline{\beta} + g(t) + \varepsilon \qquad \ldots\ldots\ldots(1)$$

اذ يشير $\underline{Y}$ الى متجة متغير المعتمد أو متغير الاستجابة من الدرجة($n*1$)

و**X** يمثل مصفوفة المتغيرات التوضيحية (المعلمية) من الدرجة ($n*p$)
$\underline{\beta}$ متجة المعالم المجهولة من درجة($p*1$)

اذ يمثل **Xβ** الجزء المعلمي للأنموذج المدروس.
t هو متغير مستمر ويمثل عادة المتغير اللامعلمي في البيانات من درجة ($n*1$) .
g(t) تمثل دالة تمهيدية غير معرفة من الدرجة($n*1$).
ε متجة الاخطاء العشوائية من درجة ($n*1$) وتكون مستقلة بمتوسط صفر وتباين σ^2.

وللانموذج المذكور انفا عدة تسميات اذ يسمى بأنموذج الانحدار شبه المعلمي **Semiparametric regression model** [15,6]، او الانموذج الخطي الجزئي **partially linear model** [25] ، وسبب تسميته خطي كونه يتضمن جزئين جزء لامعلمي وجزء معلمي خطي تربط هذه الاجزاء علاقة تجميعية . [22]
ولكي يتم بقاء التنبؤات سليمة للأنموذج المذكور انفا يتم العمل على تقدير كل من متجة المعلمات $\underline{\beta}$

فضلا عن الدالة التمهيدية **g(t)** علما أن هذا النوع من النماذج يتم تقديرها ليس بشكل منفرد أي لكل جزء على حدة وأنما يتم بطريقة تكرارية بحيث يتم تقدير الجزء اللامعلمي أولا بعد وضع أفتراضات أولية حول قيمة المعالم المجهولة $\underline{\beta}$ ومن ثم يتم تقدير متجة المعلمات المجهولة $\underline{\beta}$ بعد تقدير الجزء اللامعلمي منه

وأن المقدر الناتج وفق هذه الطريقة يدعى بالمقدر شبه المعلمي[25] .

ويمكن تمثيل الأنموذج رقم (١) بالمصفوفات بالاسلوب الاتي:

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

وقد تم استعمال هذا الانموذج على نطاق واسع في مجالات مختلفة مثل الدراسات الاقتصادية و المالية و البيئية مثل درجات الحرارة وتلوث الهواء وسرعة الرياح وغيرها.

ولغرض توضيح هذا النوع من النماذج يتم تقدير الجزء المعلمي باستعمال أحدى الطرائق المعلمية المعروفة وهي (طريقة المربعات الصغرى الاعتيادية OLS) والتي يمكن وضعها كقيم أولية أبتدائية لقيم المعلمات المجهولة بعدها يتم تقدير الجزء الاخر اللامعلمي وفق احدى الطرائق الاتية:

١.مقداللب Nadaraya-Watson .

٢.مقدر شريحة تكعيبي cubic smoothing spline.

٢-١-١ مقدر Nadaraya-Watson لدالة الانحدار شبه المعلمي

Nadaraya-Watson Kernel Estimator

يعد هذا المقدر أحد أقدم المقدرات اللامعلمية وأكثرها شيوعا و أستعمالا وتعود تسميته الى اسماء الباحثين اللذان اقترحا هذا المقدر Nadarayaو Watson عام 1964باعتماد على طريقة متسلسلة الاوزان . وقد درس هذا المقدر في كثير من البحوث(25,10,4,3,2,1) لكن دراسة الباحثين أقتصرت على دوال الانحدار اللامعلمية الأانه في هذا البحث سيتم دراسته كمقدر شبه معلمي ويمتاز هذا المقدر بصفات عديدة منها أمكانية أستعماله سواء أكان التصميم ثابتا أم عشوائيا فضلا عن أمتلاكه دالة محددة ومستمرة وتكاملها مساو للواحد ، حيث و يعد هذا المقدر حالة خاصة من مقدر متعدد الحدود الموضعي ذو درجة p [11] عندما p=1 ، الا أن هذا المقدر يعاني من تحيزات عند نقاط الحد وهذا ما يمكن تداركه عند أختيار القيمه المثلى للمعلمة التمهيدية .[10,2,1]

أما دالة اللب Kernel المستعملة مع هذا المقدر فتمتاز بعدة خصائص منها :

$$\int k(u)du = 1$$

$$\int uk(u)du = 0$$

أذا كانت الدالة من درجة ٢ اما اذا كانت اكبر من اثنين فالمعادلة يمكن ان تعمم الى:

$$\int u^{z}\, k(u) du = 0 \qquad , \forall \ \ z = 0,1,2 \ldots\ldots, k-1$$

حيث k تشير الى درجة دالة اللب kernel .
ويمكن تلخيص عمل مقدر N.W في دوال الانحدار شبه المعلمي وفق الصيغة الاتية :[12,11]

$$g_n(t,\beta) = \sum_{i=1}^{n} W_{ni}(t)(Y_i - X_i{}'\beta) \qquad \ldots\ldots(3)$$

اذ تشير $\{W_{ni}(t)\}_{i=1}^{n}$ الى سلسلة الاوزان وأن دوال الوزن هذه تكون طبيعية اذا حققت الشرط الاتي:

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

وتكون غير سالبة اذا كانت دالة الوزن

$$W_{ni} > 0$$

وتكون دالة الوزن احتمالية اذا كانت غير سالبة وتكاملها مساو للواحد.
اما شكل الاوزان فيتم تحديده من خلال دالة الوزن والتي تمثل ايضا دالة kernel $K\left(\frac{t-T_i}{h}\right)$

وان هذه الدالة تصل الى اعظم مايمكن عندما T_i يقترب الى T وتتناقص عندما T_i تبتعد عن T.[2]
ويمكن تعريف دالة الوزن على النحو الاتي[25]

$$W_{ni}(t) = \frac{K(\frac{t-T_i}{h_n})}{\sum_{j=1}^{n} K(\frac{t-T_j}{h_n})} = \frac{K(u)}{\sum k(u)} \ldots\ldots\ldots(4)$$

ولغرض تقدير أنموذج الانحدار شبه المعلمي للمعادلة (١) بالاعتماد على مقدر Nadaraya-Watson [١٨]
يتم وضع قيم أولية معرفة للمتجة β كي يتم تقدير الجزء اللامعلمي

$$Y_i - X_i{}'\beta = g(T_i) + \varepsilon_i \qquad \ldots\ldots(5)$$

فأن مقدر kernel $\hat{g}(t)$ لل g(t) يكون

$$\hat{g}(T_i) = \frac{\sum_{i=1}^{n} K\left(\frac{T_i - T_j}{h}\right)(Y_i - X_i\beta)}{\sum_{i=1}^{n} \quad K\left(\frac{T_i - T_j}{h}\right) + n^{-2}} \qquad \ldots\ldots\ldots(6)$$

وبفرض أن K(.) تشير الى دالة لبية والتي يمكن الحصول عليها من استعمال دالة كثافة طبيعية معيارية (Gaussian kernel) وباستعمال المقدر $\hat{g}(T_i)$ عوضا عن $g(T_i)$ في المعادلة (5) فنحصل على:

$$Y_i - X_i{'}\beta \approx \frac{\sum_{i=1}^{n} K\left(\frac{T_i - T_j}{h}\right)(Y_i - X_i\beta)}{\sum_{i=1}^{n} \quad K\left(\frac{T_i - T_j}{h}\right) + n^{-2}} \qquad \ldots\ldots(7)$$

وباستعمال التحويلات التالية نحصل على :

$$Z_i \approx U_i{'}\beta \quad , \qquad \ldots\ldots\ldots(8)$$

$$Z_i = Y_i - \frac{\sum_{i=1}^{n} Y_i K\left(\frac{T_i - T_j}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{T_i - T_j}{h}\right) + n^{-2}} \qquad \ldots\ldots\ldots(9)$$

$$U_i = X_i - \frac{\sum_{i=1}^{n} X_i K\left(\frac{T_i - T_j}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{T_i - T_j}{h}\right) + n^{-2}} \qquad \ldots\ldots(10)$$

وحسب نظرية أنموذج الانحدار الخطي فأنه بالامكان تقدير قيمة المعالم $\underline{\beta}$ كما يلي:

$$\hat{\beta}_n = (U_i{'}U_i)^{-1}(U'_i Z_i) \quad \ldots\ldots\ldots(11)$$

حيث n هو حجم العينة .
وبالتعويض في المعادلة (6) نحصل على:

$$\hat{g}_n(T_i) = \frac{\sum_{i=1}^{n} K\left(\frac{T_i - T_j}{h}\right)\left(Y_i - X_i\hat{\beta}_n\right)}{\sum_{i=1}^{n} K\left(\frac{T_i - T_j}{h}\right) + n^{-2}} \quad \ldots\ldots\ldots(12)$$

2-1-2 تمهيد الشريحة التكعيبية في تقدير دالة الانحدار شبه المعلميه [٩،٤]

تعد طريقة شرائح التمهيد التكعيبية احدى طرائق موائمة المنحنى اللامعلمي والتي استعملت ايضا كطريقة موائمة للنماذج او الدوال شبه المعلميه وتعتمد هذه الطريقة على مجموع مربعات البواقي الجزائية Penalized residual sum of squares مضافا اليها حد الجزاء roughaness penalty كما يلي :

$$\sum_{i=1}^{n}\left(y_i - x'_i\beta - g(t_i)\right)^2 + \lambda\int_a^b (g^m(t))^2 dt \quad \ldots\ldots(13)$$

m: تشير الى المشتقة من الدرجة m^{th} للدالة g(t) ويتم ايجاد المقدر الافضل عن طريق تصغير المقدار المذكور انفا.
و يشير الحد الاول للمعادلة المذكورة انفا الى مجموع مربعات البواقي .
λ : هي معلمة التمهيد .وتشير الى عرض الحزمة لمقدر متعدد الحدود .
أما الحد الثاني من المعادلة (١٣) فيشير الى حد الجزاء ويكون موزونا بـ λ والذي يكون كبيرا عندما يكون تكامل المشتقة الثانية لدالة الانحدار g(t) كبيرا كما أن معلمة التمهيد λ تؤدي دورا مهما في تحديد قيمة الجزاء غير الممهد roughaness penalty، عندما $\lambda \to 0$ فأن مجموع مربعات البواقي سوف يوضح البيانات أي أن الحد الثاني (حد الجزاء) سيختفي من المعادلة ومقدر مجموع مربعات البواقي سيوضح البيانات أما عندما تكون قيمة λ كبيره جدا أي $\lambda \to \infty$ فأن المقدر سيطغى على مجموع مربعات البواقي وينتج منحنى ثابتا للانحدار الخطي اي يطابق الأنموذج المفسر التوضيحي مع توضيح كامل وتام للمنحني وتوضح البيانات بعدد غير منتهي من عرض الحزمة [١٠،٤،٣].

ولذا فأن معلمة التمهيد λ تعمل كمفتاح للتحكم بين حسن المطابقة The goodness of fit المتمثلة بـ $\sum_{i=1}^{n}(y_i - x_i'\beta - g(t))^2$ وممهد التقدير smoothness of estimate

مقاسا بـ $\int[g^m(t)^2]$ ويكون الشرط الضروري على دالة g أن تكون قابلة للاشتقاق مرتين ويكون

هناك امكانية اجراء تكامل لمربع المشتقة الثانية لها . وسيدعى المقدر بالشريحة التمهيدية من الدرجة m^{th} وعندما m=2 سيتم الحصول على شريحة تمهيدية تكعيبية Cubic smoothing spline .[٩]
وبصورة عامة يعد تحديد الجزاء penalty وتحديد موضع العقد مهم جدا عند دراسة شرائح التمهيد .
اذ أن الفرق بين شرائح التمهيد smoothing spline و شرائح الانحدار regression spline هو أن شرائح التمهيد تعتبر أن المشاهدات هي العقد اي ان (Knot=n) ولكن في حالة كون عدد المشاهدات كبير جدا يصبح من صعب دراسة هذا النوع من التمهيد لذلك يتم استعمال اسلوب اكثر عمومية هو شرائح الانحدار التي تعد دالة متعددة الحدود قطعية piecewise polynomial function [٧] حيث يتم اختيار العقد بشكل اختياري عن طريق حذف العقد غير الاساسية (Knots<n) [١٣,١٩] .
أن تقدير دالة الجزاء صعب جدا من ناحية البرمجة ويحتاج الى جهد وأمكانيات رياضية متقدمة ،ان طريقة حل الجزاء غير الممهد التي نوقشت من قبل الباحثان Green&Silverman [٩] في عام ١٩٩٤ اللذان أوجدا طريقة لحساب الجزاء غير الممهد كمايأتي :
نفرض أن لدينا n من المشاهدات $(t_1,........t_n)$ في الفترة $[a,b]$ ،وأن الدالة g تشير الى شريحة

تكعيبية cubic spline اذا تحقق الشرطين الاتيين :[٩]
i. في الفترة $(a,t_1),(t_1,t_2)............(t_n,b)$ فأن الدالة g هي شريحة تكعيبية متعددة الحدود polynomial cubic spline .
ii. أن متعدد الحدود القطعية polynomial pieces تكون ملائمة fitting في النقطة t_i للدالة g و g' و g'' ومستمرة في نقاط t_i ،أي أن الدالة g تكون مستمرة في الفترة $[a,b]$.

وأن الشريحة التكعيبية cubic spline تسمى شريحة تكعيبية طبيعية natural cubic spline اذا كانت المشتقة الاولى والثانية للدالة g هي صفر في النقاط a و b ، لذلك فأن دالة g تكون دالة خطية linear function في النقطتين (a,t_1) و (t_n,b) .
نفرض أن $g_i=g(t_i)$ و $\upsilon_i=g''(t_i)$ وبتعريف $\upsilon_1=\upsilon_n=0$
نفرض أن g هي متجة $(n*1)$ حيث $(g_1,g_2,.......g_n)'$ وأن υ تشير الى متجة المشتقة الثانية $(\upsilon_2,\upsilon_3,......\upsilon_{n-1})'$ من الدرجة $((n-2)*1)$ ،أن g و υ هما اللذان يحددان شكل المنحنى g .
حيث هذه المتجهات يتم تعريفها عن طريق أثنين من المصفوفات R و Q وكما يأتي
a : يشير الى متجة المسافات أو الفروق بين المشاهدتين t_{i-1},t_i وعناصرها (i^{th}) كالاتي

$$a_i = t_i - t_{i-1} \qquad i = 1,2,\dots\dots,n-1$$

Q مصفوفة بدرجة (n*(n-2)) عناصرها q_{ij} ويتم حسابها كما يأتي :

$$q_{(j-1,j)} = a_{j-1}^{-1}$$

$$q_{jj} = -a_{j-1}^{-1} - a_j^{-1}$$

$$q_{(j+1),j} = a_j^{-1}$$

$$q_{(i,j)} = 0 \quad , \forall |i-j| \geq 2$$

$$i = 1,2, \ldots\ldots n; \quad j = 1,2, \ldots\ldots n-1$$

R مصفوفة متماثلة symmetrical وبدرجة ((n-2)*(n-2)) ويتم حساب عناصرها وفق الاتي :

$$r_{(i,i)} = \frac{1}{3}\left[a_{(i-1)} + a_{(i)}\right] \qquad ; i = 1,3 \ldots\ldots, n-1$$

$$r_{(i,i+1)} = r_{(i+1,i)} = \frac{1}{6} a_i \qquad ; i = 1, \ldots\ldots\ldots, n-2$$

$$r_{(i,j)} = 0 \qquad , \forall \ |i-j| \geq 2$$

$$i = 1,2, \ldots .n-1 \quad ; \ j = 1,2, \ldots .n-1$$

وبما أن المصفوفة R هي مصفوفة diagonal حيث أن :

$$r_{ii} > \sum_{i \neq j} |r_{ij}|$$

وبما أن المصفوفة R هي مصفوفة موجبة ومعرفة positive –definite فأن R^{-1} موجودة لذلك يمكن تعريف المصفوفة K كما يلي

$$K = QR^{-1}Q'$$

وأثبت الباحثان [9] Green and Silverman أن المتجهات g و ʊ هما شرائح تكعيب طبيعية أذا وفقط اذا تحقق الشرط الاتي:

$$Q'g = R\upsilon$$

وعند توفر الشرط المذكور انفا فان الجزاء غير الممهد يحسب وفق الشكل الاتي

$$\int_a^b [g''(t)]^2 dt = \upsilon' R \upsilon$$

$$= g'kg \dots\dots\dots\dots\dots\dots\dots (14)$$

وبالاستفادة من النظرية المذكور انفا يمكن حساب الجزاء غير الممهد بسهولة عن طريق ضرب المصفوفات و المتجهات .[16,9]

ولغرض تقدير أنموذج الانحدار شبه المعلمي للمعادلة (١) بالاعتماد على شريحه التمهيد التكعيبية والذي نوقش من قبل Speckman [22,8,7,6,5] في عام ١٩٨٨ الذي افترض ان $g=(g(t_1),\dots\dots.g(t_n))$ يشير الى متجة القيم للدالة g عند نقاط العقد التي تمثل المشاهدات $t_1,t_2,\dots\dots t_n$ أما $\hat{g}_\lambda$ تشير الى مقدر

شريحة تكعيبية cubic spline estimator للمتجة وأن

$$y = [y_1 \dots\dots y_n]'$$

اذ أن

$$\hat{g} = [\hat{g}_\lambda(t_1) \dots\dots\dots \hat{g}_\lambda(t_n)]' = (S_\lambda)(y_1 \dots\dots y_n)'$$

وبالاعتماد على صيغة المصفوفات فيمكن تمثيل المقدر وفق الصيغة الاتية:

$$\hat{g}_\lambda = S_\lambda \underline{y} \dots\dots\dots (15)$$

اذ تمثل λ معلمة تمهيد للشريحة

S_λ تشير الى مصفوفة تمهيد معرفة موجبة ومربعة من درجة $(n*n)$ positive – definite (symmetrical) smoothing matrix وتحسب من الصيغة الاتية

$$S_\lambda = (I + \lambda K)^{-1} \dots\dots\dots (16)$$

اذ تعتمد هذه المصفوفة عند حسابها حسب الصيغة (١٦) على λ (معلمة التمهيد) وعلى نقاط العقد $(t_1,\ldots\ldots t_n)$.
ونفرض أن

$$\tilde{X} = (I - S_\lambda)X \ldots\ldots\ldots (17)$$

$$\tilde{Y} = (I - S_\lambda)Y \ldots\ldots\ldots (18)$$

فيكون تقدير $\hat{\beta}$ لأنموذج PLM كما يأتي

$$\hat{\beta} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y} \ \ldots\ldots\ldots (19)$$

وبتعويض المعادلة المذكورة انفا في أنموذج PLM فنحصل على

$$Y_i^* = Y_i - X'\hat{\beta}$$

$$Y_i^* = g(t_i) + \varepsilon_i \ldots\ldots\ldots. (20)$$

ثم نحصل على تقدير $\hat{g}$ بطريقة css للقيمة Y_i^* للجزء اللامعلمي للأنموذج plm فيكون المقدر

$$\hat{g} = S_\lambda\left(Y - X\hat{\beta}\right) \ldots\ldots\ldots\ldots (21)$$

$$= S_\lambda Y_i^*$$

اذ يتم تقديرمتجة المتوسطات $\hat{\mu}$ للدالة شبة المعلمية يأتي [21]

$$\hat{\mu} = \hat{g} + X\hat{\beta} = H_\lambda Y \ldots\ldots. (22)$$

ومن الجدير بالاشارة الى وجود عديد من الدراسات و البحوث على ايجاد المقدر الافضل للمعلمة التمهيدية الا ان في هذا البحث سوف يتم التطرق الى أحدى طرائق تقدير المعلمة التمهيدية والتي تدعى بطريقة قاعدة الابهام Rot عند استعمال مقدرN.W وطريقة العبور الشرعي CV عند استعمال مقدر CSS.

3-2 الانموذج اللامعلمي اللبي Nonparametric Kernel model(23,18,14)

في هذا البحث تم أدخال مقدر لامعلمي لدالة الانحدار متعددة المتغيرات وأن الهدف من هذا المقدر يستند على أفتراض أن المتغيرات في النموذج تسلك سلوكا لامعلميا ولاخطيا مع متغير الاستجابة مما يتطلب تقدير تلك العلاقة بالاعتماد على مقدر لامعلمي وأن صيغة هذا النوع من المقدرات تكون بالاعتماد على أنموذج الانحدار الاتي :

$$Y = g(X,T) + \varepsilon \qquad \ldots\ldots\ldots(23)$$

اذ تشير g الى دالة انحدار لامعلمية لكن للمتغيرات X,T وأن المقدر لتلك الدالة يكون وفق الصيغة الاتية:

$$\hat{g}_n(x,t) = \frac{\sum_{i=1}^{n} Y_i K_1\left(\frac{x-x_i}{h_x}\right) K_2\left(\frac{T-T_i}{h_T}\right)}{\sum_{i=1}^{n} K_1\left(\frac{x-x_i}{h_x}\right) K_2\left(\frac{T-T_i}{h_T}\right) + n^{-2}} \qquad \ldots\ldots(24)$$

علما ان هذا المقدر يكون بأفتراض وجود متغيرين وبالامكان تعميم هذا النوع من المقدرات بتعميم تلك الصيغة وفق عدة متغيرات مستعملة.

ويشير كلا من $K_1(.)$ و$k_2(.)$ الى دوال لبية قد تكون متساوية او مختلفة أي قد تكون $K_1(.)=K_2(.)$ أو قد يكون $K_1(.)\neq K_2(.)$ ،

في حين أن معلمة التمهيد h في المقدر يتم أختيارها وفق أحدى طرائق تقدير معلمة التمهيدية .

في هذا البحث توجد ثلاث متغيرات توضيحية لذلك تكون الصيغة للمقدر كما يأتي:

$$\hat{g}_n(x_i,t) = \frac{\sum_{i=1}^{n} Y_i K_1\left(\frac{x_1-x_{i1}}{h_{x1}}\right) K_2\left(\frac{x_2-x_{i2}}{h_{x2}}\right) K_3\left(\frac{T-T_i}{h_T}\right)}{\sum_{i=1}^{n} K_1\left(\frac{x_1-x_{i1}}{h_{x1}}\right) K_2\left(\frac{x_2-x_{i2}}{h_{x2}}\right) K_3\left(\frac{T-T_i}{h_T}\right) + n^{-2}} \qquad \ldots\ldots(25)$$

3- الجانب التجريبي

تم تنفيذ تجارب المحاكاة باستعمال ثلاث حجوم للعينات (n=40,60,100) وبتكرار مقداره ٥٠٠ (replicatec=500) لكل تجربة وكما يأتي:

I. المتغيرات التوضيحي X_i تتوزع توزيعاً طبيعياً بوسط (٠) وتباين (١)

$$X \sim N(0,1)$$

II. المتغير التوضيحي اللامعلمي t_i يتوزع توزيعاً طبيعياً بوسط (٠) وتباين (١)

$$t \sim N(0,1)$$

III. الاخطاء العشوائية تتوزع توزيعاً طبيعياً بوسط صفر وتباين σ^2 $\in \sim N(0,\sigma^2)$ وقد تم أفتراض ثلاث قيم لتباين الخطأ هي (٠.٥، ١، ١.٥) .

IV. المتغير المعتمد Y_i يتم توليده من خلال النماذج المستعملة في تجارب المحاكاة وذلك باستعمال دوال الانحدار بدلالة المتغيرات التوضيحية في الفقرة (I وII) مضافا اليها الخطأ .

3-١ النماذج المستعملة في المحاكاة

ان التنوع الكبير في النماذج ناتج من تنوع الظواهر التي تمثلها وقد تم اختيار بعض النماذج التي تناسب الطرائق المستعملة في هذه الرسالة من بحوث منشورة فعلا :

1. $g(t) = 3.2t^2 - 1$ [١4,١8,٢3]

2. $g(t) = 0.5\sin(2\pi t_i)$ [6]

3. $g(t) = t_i - 3t_i^2 + 3t_i^3$ [21]

أماالدالة اللبية kernel المستخدمة هي دالة كثافة طبيعية قياسية Gaussion kernel(standard normal density function)

$$K(.) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$$

علما ان تفسير الرموز في الجداول كانت :

أنموذج الانحدار الخطي الجزئي	Partial linear regression model	PLM
مقدر نادارايا-واتسن اللبي	Nadaraya-Watson Kernel Estimator	N.W
مقدر شريحة تكعيبي	Cubic Smoothing Spline	C.S.S
طريقة المربعات الصغرى الاعتيادية	Ordinary Least Square Method	OLS
مقدر لامعلمي لبي	Non parametric Kernel Estimator	Non

جدول (1)
يبين معدل متوسط مربعات الخطأMASE لمقدرات الأنموذج الاول

n	Method σ^2	OLS	PLM N.W	Non N.W	PLM C.S.S
40	0.5	10.4116	3.63934	14.8621	3.62162
	1	11.5974	6.43624	7.20495	4.30896
	1.5	12.0816	5.29560	10.2010	5.49632
60	0.5	8.237962	3.60412	5.78784	4.13289
	1	7.47182	3.10898	4.17165	3.35120
	1.5	6.84087	3.85861	5.40412	4.64405
100	0.5	4.855693	2.00853	2.04367	2.12387
	1	4.953918	2.29423	2.44068	2.32607
	1.5	4.916306	2.79637	2.98121	2.83749

جدول (٢)
يبين معدل متوسط مربعات الخطأMASE لمقدرات الأنموذج الثاني

n	Method σ^2	OLS	PLM N.W	Non N.W	PLM C.S.S
40	0.5	5.94484	4.08702	6.04601	3.83454
	1	5.72923	4.29249	4.59150	3.83094
	1.5	5.48759	4.13410	5.7871	4.05679
60	0.5	4.14287	2.80149	3.33604	2.76552
	1	4.18623	3.27735	3.44381	3.35422
	1.5	3.17872	2.57968	2.79718	2.52679
100	0.5	3.735144	2.76311	2.93002	2.53378
	1	4.758539	3.59775	3.46203	3.41524
	1.5	2.755581	1.57852	1.63750	1.42079

جدول (٣)

يبين معدل متوسط مربعات الخطأMASE لمقدرات الأنموذج الثالث

n	Method σ^2	OLS	PLM N.W	Non N.W	PLM C.S.S
40	0.5	4.41998	3.54137	3.61793	3.58210
	1	4.49690	2.53261	3.12436	2.66588
	1.5	4.34512	2.55785	2.91472	2.26700
60	0.5	4.070481	3.66207	3.71155	3.65779
	1	4.077361	3.30806	3.17146	3.25752
	1.5	4.0691522	3.17220	3.19618	3.13980
100	0.5	3.0325965	2.56032	2.57373	2.52185
	1	3.0331830	1.86856	2.05117	1.83725
	1.5	3.0311844	1.87347	1.87022	1.83574

٣-٢ تفسير نتائج الجانب التجريبي:

تفسيرات الجدول رقم (1) الخاص بقيم (MASE) لكل من المقدرات ولجميع أحجام العينات والتباينات المستعملة للأنموذج الاول :

1.أظهرت النتائج ان مقدر N.W لنموذج PLM هو المقدر الافضل لجميع قيم تباين البواقي وحجوم العينات المستعملة يليه مقدر C.S.S لنموذج PLM ، ماعدا عند حجم عينة n=40 وتباين σ^2 =0.5,1 اذ اشارت النتائج ان مقدر الانحدار شبه المعلمي باستعمال مقدر الشريحة التكعيبية C.S.S هو افضل من المقدر شبه المعلمي بأستعمال N.W .

2.كما أشارت النتائج أن اقل المقدرات اداءاً وتمثيلاً للأنموذج المفترض هو مقدر المعلمي OLS لجميع قيم تباين البواقي واحجام العينات المستعملة، ماعدا عند حجم عينة n=40 وتباين σ^2 =0.5 اذ كان المقدر اللامعلمي اللبي هو الاقل اداءاً وتمثيلاً للأنموذج المفروض.

3.تقل قيم MASE بتزايد حجوم العينات و لجميع المقدرات المستعملة .

4.تزداد قيم MASE لجميع المقدرات مع تزايد قيم تباين البواقي ما عدا عند حجم عينة n=60 اذ نلاحظ تذبذب في قيمة MASE لجميع المقدرات ،وكذلك عند حجم عينة n=40 اذ نلاحظ تزايد قيمتها عند مقدر OLS فقط تفسيرات الجدول رقم (2) الخاص بقيم (MASE) لكل من المقدرات ولجميع أحجام العينات والتباينات المستعملة للأنموذج الثاني :

١.أظهرت النتائج ان مقدر C.S.S للأنموذج الخطي الجزئي PLM هو المقدر الافضل لجميع قيم تباين البواقي وحجوم العينات المستعملة يليه مقدر N.W، ماعدا عند حجم عينة n=60 وتباين σ^2 =1 نلاحظ ان مقدر N.W هو الافضل يليه مقدر C.S.S لنموذج PLM .

٢.أظهرت النتائج ان أسوء مقدر هو المقدر OLSولجميع قيم تباين البواقي واحجام العينات المستعملة ما عدا عند حجم عينة n=40 وتباين σ^2 =0.5,1 اذ اظهرت النتائج عدم قدرة المقدر اللامعلمي على تمثيل الانموذج بشكل امثل .

٣.تقل قيمة MASE عند زيادة حجوم العينات و لجميع المقدرات ماعدا عند تباين σ^2 = 1 نلاحظ تذبذب قليل في قيمة MASE لجميع المقدرات المستعملة حيث قيمتها عند حجم عينة n=100 اعلى بقليل من قيمتها عند حجم عينة n=60 .

٤.نلاحظ تذبذب قيمة MASE لجميع المقدرات مع تزايد قيم تباين البواقي ، ماعدا عند حجم عينة n=40 لمقدر C.S.S لنموذج PLM تزداد قيمة MASE مع تزايد قيم تباين البواقي.

تفسيرات الجدول رقم (3) الخاص بقيم (MASE) لكل من المقدرات ولجميع أحجام العينات والتباينات المستعملة للأنموذج الثالث:

١.أظهرت النتائج ان أفضل مقدر هو مقدر C.S.S لنموذج PLM لجميع قيم تباين الخطأ و حجوم العينات المستخدمة يليه مقدر N.W لنموذج PLM ماعدا عند حجم عينة n=40وتباين $\sigma^2 = 0.5$ فأن مقدر N.W لنموذج PLM هو افضل مقدر يليه مقدر C.S.S لنموذج PLM ، وعند حجم عينة n=40,60 وتباين $\sigma^2 = 1$ فأن المقدر اللامعلمي اللبي هو الافضل يليه N.W لنموذج PLM.
٢.كما أشارت النتائج أن اقل المقدرات اداءاً وتمثيلاً للأنموذج المفترض هو مقدر OLS لجميع قيم تباين الخطأ و حجوم العينات المستخدمة.
٣.نلاحظ تذبذب قيمة MASE لجميع المقدرات مع تزايد حجم العينة لجميع المقدرات المستعملة ماعدا مقدر OLS تقل قيمة MASE مع تزايد حجم العينة.
٤.نلاحظ تذبذب قيمة MASE لجميع المقدرات مع تزايد قيم تباين البواقي.

٤- الاستنتاجـــات:

بناءا على ما تم تحليله من نتائج المحاكاة وفقا لكل أنموذج من نماذج الانحدار اللامعلمي ولجميع الحالات الأخرى تم التوصل إلى الاستنتاجات الآتية :
١.كحالة عامة إن قيم MASE تتناسب تناسبا عكسيا مع إحجام العينات، إي أنها تتناقص عند تزايد حجوم العينات .
٢.ان مقدر الانحدار شبه المعلمي باستعمال مقدر الشريحة التكعيبية C.S.S ومقدر N.W هما المقدرات الافضل للنماذج المستعملة .
٣.كما أشارت النتائج أن اقل المقدرات اداءاً وتمثيلاً للنماذج المستعملة هو مقدر المعلمي OLS لجميع قيم تباين البواقي وحجوم العينات المستعملة.

٥-الدراسات المستقبلية

تفتح هذه الدراسة افاقا رحبة أمام الاخرين في المستقبل لدراسات عديدة من أهم هذه الدراسات المقترحه مايأتي:
١. تقدير أنموذج الانحدار شبه المعلمي PLM بأستعمال طريقة تمهيد Wavelet .
٢. دراسة نماذج أخرى شبه معلميه مثل أنموذج احادي المؤشرsingle index model وأنموذج احادي المؤشر الخطي الجزئي partially linear single index model في حالة بيانات تامة وغير تامة.
٣. تقدير نماذج لامعلمية وشبة معلمية بأستعمال انواع اخرى من شرائح التمهيد مثل شرائح تربيعية B-Spline وشرائح الانحدار regression spline وتقدير الجزء المعلمي منه بأستعمال طرائق تقدير اخرى مثل بيز وطريقة الامكان الاعظم MLE وغيرها.

المصادر

1. مصادر العربية

1.حمزة،سعد كاظم،(٢٠٠٩)،" مقارنة بعض الطرائق أللبية في تقدير نماذج الانحدار اللامعلمية بوجود بيانات تامة وغير تامة" رسالة ماجستير في الإحصاء كلية الادارة والاقتصاد، جامعة بغداد.
2.حمود، مناف يوسف،(٢٠٠٠)،"مقارنة مقدرات kernel اللامعلمية لتقدير دوال ألانحدار رسالة ماجستير في الإحصاء كلية الادارة والاقتصاد، جامعة بغداد
3.متي،نور صباح و الصفاوي ، صفاء يونس، (٢٠١١)،"تقدير دوال الانحدار اللامعلمي باستخدام بعض اساليب التمهيد" المجلة العراقية للعلوم الاحصائية(٢٠)،عدد خاص بوقائع المؤتمر العلمي الرابع لكلية علوم الحاسبات والرياضيات.ص٣٧٣-٣٩٢.

٢. المصادر الاجنبية

4. Aydin,D.(1999)”Acomparison of the nonparametric regression model using smoothing spline and kernel regression”international journal of mathematical.physical and engineering scinces –vol .2 ,N0.2, pp.75-79.
5.Aydin,D.(2007)” Estimation of GDP in Turkey by nonparametric regression models” Proceedings of the 6th WSEAS International Conference on Instrumentation, Measurement, Circuits & Systems, Hangzhou, China,pp.221-225.
6. Aydin,D.(2011)’’ Partially Linear Models Based on Smoothing Spline Estimated by Different Selection Methods: A Simulation Study”
Department of Statistics, Faculty of Arts and Sciences, Muğla University. www.interstat.statjournals.net/year2011/articles/1105006.
7. Aydin,D.and Tuzemen,M.(2010)’’ Estimation in Semi-parametric and Aditive Regression using Smoothing and Regression Spline’’ Second International Conference on Computer Research and Development, computer society
8.Eubank,R.L,Kambour.E,Klippe.K,Reese,C. and Schimek.M.(1998) ”Estimation in partially linear models” computational statistics &data analysis 29.pp27-34.
9. Green, P.J. and Silverman , B.W., (1994), “Nonparametric regression and generalized linear models : A roughness penalty approach”,Chapman and Hall, London.
10. Härdle, W. (1990) ” Applied Nonparametric regression “ Cambridge, Cambridge University press.
11. Härdle,W.,Liang,H.,Gao.J.(2000)” Partially Linear Models “ http://cdn.preterhuman.net.

12.Härdle,W., Mori,Y.& Vieu,Ph.,(2007)" Statistical Methods for Biostatistics and Related Fields ".Springer-Verlag Berlin Heidelberg .
13.Hens,N.,(2005)" Non- and Semi-parametric Techniques for Handling Missing Data"
http://ibiostat.be/publications/phd/nielhens.pdf
14. Kartiko,S.H.,(2012)" Semiparametric Regression with Missing Values" Math Dept, UGM, Yogyakarta, Indonesia.
http://mfile.narotama.ac.id/files/Umum/JURNAL%20UGM/SEMIPARAMETRIC.
15."Kernel Regressing",(2013)
http://en.wikipedia.org/wiki/Nadaraya-watson-estimator
16. Millimet, D.,and List, J., Stengos, T., (2003) "The environmental Kuznets curve: Real progress or misspecified models? Rev Econ Stat (85)4 .pp1038–1047
17. Noor A. Ibrahim,N.A. and Suliadi,S (2010)" GEE-Smoothing spline in semiparametric model with correlated nominal Data: Estimation and simulation study" pubilished in: proceeding ASM'10 proceedings of the u th international coference on applied mathematics,simulation,modeling.
pp19-26.
18. Qin,Y., Zhang,Sh.,and Zhu ,X.,(2006)" Semi-parametric optimization for missing data imputation" Springer, Appl Intell
DOI 10.1007/s10489-006-0032-0.
19. Ruppert ,D.,(2002)" Selecting the Number of Knots for Penalized Splines" Journal of Computational and Graphical Statistics, Vol.11, N0. 4, PP735–757.
20.Ruppert,D.,Wand,M.P.andCarroll,R.J.,(2003)"semiparametric regression"Cambridge university press.
21. Schimek,M.G.,(2000)," Estimation and inference in partially linear models with smoothing splines" Journal of Statistical Planning and Inference(91) ,PP 525_540.
22. Speckman,P.,(1988), " Kernel Smoothing in Partial Linear Models"
J. R. Statist. Soc. 50, No. 3, pp. 413-436.
23. Wang,Q.,Linton,O.,and Härdle,W.,(2004)," Semiparametric Regression Analysis with Missing Response at Random " the institute for fiscal studies department of economics, UCL ,cemmap working paper CWP11/0.
24.Wellner,J.,Klaassen,J.,and Ritov,Y.,(2005),"semiparametric models:a Review of progress since BKRW(1993)" Wspc/Trim size:ain6ix for review volume.
25.Yatchew,A.,(2003),"semiparametric regression for the applied .econometrician"Cambridge university press

A comparison of the Semiparametric Estimators model using different smoothing methods

Abstract:

In this paper, we made comparison among different parametric ,nonparametric and semiparametric estimators for partial linear regression model users parametric represented by ols and nonparametric methods represented by cubic smoothing spline estimator and Nadaraya-Watson estimator, we study three nonparametric regression models and samples sizes n=40,60,100,variances used σ^2=0.5,1,1.5 the results for the first model show that N.W estimator for partial linear regression model(PLM) is the best followed the cubic smoothing spline estimator for (PLM),and the results of the second and the third model show that the best estimator is C.S.S.followed by N.W estimator for (PLM) ,the results also indicated that the lowest estimator and representation of the models used is the parametric estimator OLS followed by nanparametric estimator N.W.

Key words: partial linear regression model, cubic smoothing spline, kernel smoothing Nadaraya-Watson estimator.