مقارنة بعض الطرائق الجزائية في تحليل انموذج المؤشر الواحد شبه المعلمي مع تطبيق عملي

أ.م.د.مناف يوسف حمود/كلية الادارة والاقتصاد /جامعة بغداد
م.م.طارق عزيز صالح/كلية الادارة والاقتصاد / جامعة واسط

المستخلص

في هدف هذا البحث تم أستعمال بعض الطرائق شبه المعلميه المعتمدة على دوال جزاء مختلفة من ضمنها الطرائق المقترحة من قبل الباحث اذ تعمل هذه الطرائق على تقدير وأختيار المتغيرات المعنوية لانموذج المؤشر الواحد شبه المعلمي في آن واحد منها (طريقة-SCAD-NPLS، الطريقة المقترحة الاولى (SCAD-MAVE) والطريقة المقترحة الثانية (ALASSO-MAVE)) إذ تم أستعمال أجراء نوعين من الدراسات التطبيقية ، الأجراء الأول أستعمل فيه أسلوب المحاكاة للمقارنة بين طرائق التقدير المدروسة ولمختلف تجارب المحاكاة والتعرف على أفضل طريقة بالأعتماد على معايير المقارنة (متوسط مربعات الخطأ (MSE) ومعدل متوسط مربعات الخطأ(AMSE)).

والأجراء الثاني أستعمل فيه البيانات الحقيقية للتحقق من أداء الطرائق شبه المعلميه في الواقع العملي، وتم التوصل الى ان افضل طريقة لتقدير واختيار المتغير لانموذج المؤشر الواحد شبه العلمي هي الطريقة المقترحة الثانية (ALASSO-MAVE) لكل من تجارب المحاكاة لانموذج المؤشر الواحد شبه المعلمي الاول والبيانات الحقيقية.

المصطلحات الرئيسية للبحث/ انموذج المؤشر الواحد ،اختيار المتغير ، طريقة سكاد ، طريقة سكاد – ماف ، طريقة لاسو التكيفية - ماف .

*البحث مستل من اطروحة دكتوراه

مجلة العلوم
الاقتصادية والإدارية
العدد 90 المجلد ٢٢
الصفحات٤٠٧-٤٢٧

١. المقدمة Introduction

يعد تحليل الأنحدار من أكثر الأساليب الأحصائية التي تقوم ببناء أنموذج أحصائي و أكثر أستعمالا" لفهم العلاقة بين المتغير المعتمد (الأستجابه) Y و مجموعة المتغيرات التوضيحية X.
ففي نماذج الأنحدار الخطي المتعدد تفترض أن صيغة دالة التوقع الشرطي الأتية :-
$E[Y \mid X = x] = X^T\beta$ هي صيغة خطية . [14] وفي معظم الأحيان نستعمل طرائق التقدير المعلميه لتحليل العلاقة على فرض أن دالة التوقع الشرطي هي معلومة لعدد من المعلمات ولكن التقديرات الناتجة يمكن أن تكون مضللة للغاية اذا كان أفتراض الأنموذج المعلمي غير صحيح ، وبما أن هذا الأفتراض لايتحقق في أغلب التطبيقات العملية لانه لا يأخذ بنظر الأعتبار التأثير اللاخطي للمتغيرات التوضيحية لذا كان من المناسب أيجاد أسلوب جديد يأخذ بنظر الأعتبار هذا التأثير و هو الأنحدار اللامعلمي (Nonparametric Regression) ، وكما هو معروف لدى أغلب الباحثون أن الأنحدار اللامعلمي يعاني من مشكلة تعدد الأبعاد (The cruse of dimensionality) التي تحدث عند زيادة عدد المتغيرات التوضيحية.[٢]
ومع ذلك فأن الطرائق اللامعلميه تمتلك عيوب مؤثرة تحد من فائدتها في التطبيقات وواحدة من هذه العيوب هي أن دقة المقدر اللامعلمي تتناقص بسرعة مع تزايد عدد المتغيرات التوضيحية وتسمى هذه الظاهرة بمشكلة تعدد الأبعاد (curse of dimensionality). وبسبب تلك المشاكل تم تقديم الطرائق شبه المعلميه كحل وسط بين تلك النماذج المعلميه المقيدة والنماذج واللامعلميه المرنة للغاية ، اذ تجعل الأفتراضات حول دالة التوقع الشرطي أقوى من تلك في التقدير اللامعلمي ولكن أقل تعقيدا" من الأفتراضات للأنموذج المعلمي وهذا يتيح للطرائق شبه المعلميه أختزال البعد الفعال لمشكلة التقدير ومن ثم زيادة دقة التقدير لتلك المتوفرة في التقدير اللامعلمي ، لذا تم أستعمال أنموذج المؤشر الواحد single index model وهو أحد النماذج شبه المعلميه الشائعة اذ يعد أمتداد طبيعي للأنحدار الخطي العام والذي يتعامل مع العلاقات اللاخطية ويعرف من خلال الصيغة $g(X^T\beta)$ =$E[Y \mid X = x]$ وهو أكثر مرونة من النماذج المعلميه بسبب أنه يسمح للعلاقات غير الخطية بين مؤشر المتغير $X^T\beta$ ومتغير الأستجابة ويحتفظ بالخصائص الجيدة الى جانب قدرته على أختزال الحد من مخاطر خطأ تحديد دالة الربط ويساعد في التغلب على مشكلة تعدد الأبعاد بسبب أن المؤشر $X^T\beta$ هو مجاميع للأبعاد المرتفعه الى X.[10] .

٢. مشكلة البحث The research problem

إن دراسة عدد كبير من المتغيرات التوضيحية وحجم عينة كبير عملية معقدة اذ بزيادتها تزيد درجة التعقيد للأنموذج مما يدفع الباحثين الى أستعمال مسألة أختيار المتغير وهي واحدة من أكثر المسائل أنتشارا" عندما يراد أيجاد أنموذجا" واحدا" يمثل العلاقة بين متغير الأستجابة والمجموعة الجزئية للمتغيرات التوضيحية الممكنة بسبب أن بعض المتغيرات التوضيحية تكون غير أساسية في تاثيرها على المتغير المعتمد أو يكون تاثيرها مماثل لتأثيرات متغيرات أخرى وأن العديد من هذه المتغيرات يكون لها أرتباطا" داخليا"، كما أن الأفتراض الخطي لايتحقق في معظم التطبيقات العملية فضلا" عن ظهور مشكلة تعدد الأبعاد (The cruse of dimensionality) التي تحدث عند زيادة عدد المتغيرات التوضيحية .

٣. هدف البحث The aim of the research

يهدف هذا البحث الى استعمال بعض الطرائق شبه المعلميه في تحليل أنموذج المؤشر الواحد شبه المعلمي (semi parametric single index model) لمعالجة مشكلة الأبعاد الكبيرة في البيانات من خلال هيكل المؤشر للأنموذج فضلا" عن أستبعاد المتغيرات غير المعنوية من الأنموذج وأختيار المتغيرات المعنوية التي لديها القدرة على التنبؤ وهي خطوة مهمة جدا" في التحليل الأحصائي . ومن ثم بناء أنموذج فعال من خلال التوصل إلى أفضل طريقة بالأعتماد على معيار المقارنة ومعدل متوسط مربعات الخطأ (AMSE) يمثل العلاقة بين المتغيرات التوضيحية والمتغير المعتمد افضل تمثيل .

٤.الجانب النظري

٤.١: أنموذج المؤشر الواحد شبه المعلمي Semi parametric single index model

أن معظم النماذج التي تحتوي على كل من معلمة محدودة الأبعاد β المجهولة والتي تمثل الجزء المعلمي مع دالة ربط مجهولة $g(.)$ والتي تمثل الجزء اللامعلمي تدعى بالنماذج شبه المعلميه (Semi parametric Models) , اذ يعد هذا الأنموذج من أكثر النماذج شبه المعلميه شيوعاً و يستعمل على نطاق واسع في العلوم التطبيقية و جاءت التسمية لهذا الأنموذج لأن جميع المتغيرات التوضيحية تتلخص تحت مؤشر خطي واحد($X^T \beta$). [3]

أي أن هذا الأنموذج يبحث في تركيبة خطية واحدة لـ P من المتغيرات التوضيحية و التي تستطيع الحصول على معظم المعلومات حول العلاقة بين المتغير المعتمد (الأستجابة) Y و المتغيرات التوضيحية X و من ثم تجنب مشكلة تعدد الأبعاد (Curse of dimensionality).[16] إذ عمل الباحثون على توسيع و تعميم لأنموذج الأنحدار الخطي العام (GLM) الآتي :[15]

$$Y_i = X_i^T\beta + \in_i \;, \quad i = 1,2,\dots, n \quad \text{--------------------------------} \quad (1)$$

With $\quad E(\in_i \mid x_i) = 0$

و يمكن التعبير عن الدالة الخطية بالشكل الآتي :

$$m(x_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} = x_i^T\beta$$

اذ أن :

β : تمثل متجه المعلمات المجهولة $\qquad \beta = ({}_1 , \dots \beta_0 , \beta \quad \beta_p)^T$

x_i^T : يمثل مصفوفة المتغيرات التوضيحية $\qquad x_i^T = (1 , x_{i1} , \dots , x_{ip})$ لـ ith من المشاهدات و من درجة n*p

وتعميمه إلى أنموذج المؤشر الواحد شبه المعلمي (SSIM) ليكون أكثر مرونة من الأنموذج الخطي العام من خلال السماح للعلاقات غير الخطية بين دالة المؤشر $X^T\beta$ و المتغير المعتمد Y . إذ يتم أستبدال الأنموذج (1) الى الصيغة الآتية :-

$$Y_i = g(X_i^T\beta) + \in_i \;, \quad i = 1,2,\dots,n \quad \text{----------------------------} \quad (2)$$

إذ تشير دالة الأنحدار أحادية المؤشر إلى :

$$E[\,Y \mid X = x\,] = g\,(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} = x_i^T\beta) = g\,(m(x_i))$$

$$g(x_i^T\beta) = E[\,Y_i \mid X = x_i]$$

بأفتراض إن $m(xi)$ هي دالة مجهولة و تمثل دالة التمهيد .

اذ أن :

(x_i , y_i) : تشير إلى عينة مستقلة و متماثلة التوزيع (iid) .

$y_i \in R$: يمثل متجه المتغير المعتمد لـ ith من المشاهدات من درجة n*1 .

$x_i \in R$: يمثل مصفوفة المتغيرات التوضيحية لـ ith من المشاهدات من درجة n*p .

β : يمثل متجه المعلمات المجهولة من درجة p*1 و يجب تقديرها بحيث تحقق شروط تشخيص الأنموذج و هي : $\|\beta^T\beta = 1$ or $\beta\| = 1 \;\|$ و أول مركبة تكون موجبة ($\beta_1 > 0$)

$g(.)$: تمثل دالة التوقع الشرطي (دالة الربط) أحادي المتغير المجهولة و يجب تقديرها و تكون من درجة n*1 .

$x_i^T\beta$: تمثل دالة معلومة للمعلمة β وتدعى بدالة المؤشر (Index) .

$\in_i \in R$: تمثل الخطأ العشوائي ذات توزيع طبيعي مستقل و متماثل (iid) و له متوسط صفر و تباين محدد σ^2 أي أن : $\in_i \sim N(0, \sigma^2)$

T : يمثل المبدلة (Transpose) .

و يمكن كتابة أنموذج المؤشر الواحد شبه المعلمي (SSIM) بصيغة المصفوفات و كما يأتي :

$Y = g (X^T\beta) + \in$

$$\begin{bmatrix} y1 \\ y2 \\ \vdots \\ yn \end{bmatrix} = \begin{bmatrix} g(x_{i1}^T\beta) \\ g(x_{i2}^T\beta) \\ \vdots \\ g(x_{in}^T\beta) \end{bmatrix} + \begin{bmatrix} \in i1 \\ \in i2 \\ \vdots \\ \in in \end{bmatrix}$$

اذ أن :

$x_i^T = (x_{i1} , x_{i2} , \dots , x_{ip})$

where $i = 1,2, \dots, n$

$j = 1,2,\dots ,p$

ولكون دالة الربط مجهولة فأن الأنموذج يدعى بأنموذج المؤشر الواحد شبه المعلمي semiparametric single index model (SSIM) والمبين في الصيغة (٢) والهدف هو تقدير متجه المعلمات β المجهولة ودالة الربط (.) g المجهولة وأجراء التقدير للأنموذج يتكون من خطوتين في الخطوة الأولى تقدير متجه المعلمات β وفي الخطوة الثانية يتم حساب قيم دالة المؤشر الخطي $x_i^T\hat{\beta}$ وأخيرا" تقدير دالة التوزيع والتي تدعى بدالة الربط (.) g المجهولة والحصول عليها من خلال أستعمال الأنحدار اللامعلمي Y على $x_i^T\hat{\beta}$. [2]

اذ يمكن التركيز على طريقة المربعات الصغرى الأعتيادية (OLS) لتقدير متجه المعلمات β مع نفس نسبة التقارب لتلك المتحققة في الأنموذج المعلمي و من ثم فأن أنموذج المؤشر الواحد (SIM) هو أنموذج معلمي دقيق لتقدير β ذات بعد واحد، مع ملاحظه أن الحد الثابت (intercept) . β هو غير قابل للتشخيص في هذا السياق لان المؤشر الخطي $x_i^T\beta$ يتضمن على المتغيرات التوضيحية المعنوية فقط لذلك لا يستعمل في تقدير الحد الثابت . β لعدم تاثيرها على أي متغير.

وكذلك يعد أنموذجا" لامعلميا" من خلال تقدير دالة الربط g(.) و هذه هي ميزة أختزال البعد لأنموذج المؤشر الواحد (SIM) ، و يتم ذلك عند تحديد دالة المؤشر($x_i^T\hat{\beta}$) فان تقدير الدالة g(.) يكون من خلال أي ممهد أنحدار لامعلمي يعتمد على البيانات ($x_i^T\hat{\beta}$، y_i) و يمكن أستعمال مقدر نادرايا- واتسون (Nadaraya – Watson estimator) أو مقدر الأنحدار الخطي الموضعي (Local Linear regression estimator(LLRE) مما ينتج : [10]

$\widehat{Y_i} = \hat{g} (x_i^T\hat{\beta})$ -- (٣)

٢.٤ : تقدير دالة الربط g(.) لانموذج المؤشر الواحد

Estimation of link function g(.) for single index model

بين الباحث Horowitz عام (١٩٩٨) أن أنموذج المؤشر الواحد (SIM) يمكن تقديره من خلال الطرائق التكرارية (iterative) أو المباشرة (Direct) غير التكرارية .

في هذا البحث سوف نستعمل بعض الطرائق التكرارية لغرض التقدير و التي يتم من خلالها تطبيق الأنحدار اللامعلمي (nonparametric regression) للحصول على تقدير متسق لـ (.) $\hat{g}$ ، وحل المشاكل المثلى غير الخطية للحصول على تقدير متسق لـ $\hat{\beta}$.

وأشار الباحث Horowitz الى أن الطرائق التكرارية هي مكثفة حسابياً لأنها تتطلب تقدير متوسط الأنحدار اللامعلمي في كل نقطة للبيانات لحساب دالة الهدف (Objective function) والتي قد تكون غير محدبة أو متعددة الحدود التي تؤدي الى تكرار أمثل لـ $\hat{\beta}$. [٧] فعند تقدير متجه المعلمات β فأن الخطوة التالية هي تقدير الدالة الربط g(.) أحادي المتغير ولتحقيق هذا الهدف يمكن أستعمال بعض مقدرات التمهيد اللبي (Kernel smoothing estimator) منها الأنحدار الخطي الموضعي (Local linear regression) والممهد اللبي ناداريا –واتسون (Nadaraya – Watson kernel smoothing) .

على سبيل المثال يمكن تقدير دالة التوقع الشرطي (دالة الربط) g(.) بأستعمال مقدر ناداريا – واتسون (NW) من خلال الصيغة الآتية :-

$$\hat{Y}_i = \hat{g}_i (x_i^T\hat{\beta} \mid \hat{\beta}) = \frac{1}{\sum_{i=1}^{n} K_h(x_i^T \beta)} . \sum_{i=1}^{n} k_h (x_i^T \beta) y_i$$

$$= \frac{1}{\sum_{i=1}^{n} K_h(x_i^T \hat{\beta} - x_j^T \hat{\beta})} . \sum_{i=1}^{n} k_h (x_i^T \hat{\beta} - x_j^T \hat{\beta}) y_i$$

اذ أن :

$= (x - X_i)^T \beta x_i^T \beta$: تمثل قيمة دالة المؤشر لـ ith من المشاهدات .

y_i : يمثل قيمة المتغير المعتمد لـ ith من المشاهدات .

$K_h (.) = k \left(\frac{\cdot}{h}\right) 1/h$: تمثل دالة kernel

$: K_h(x_i^T \beta) = k (x_i^T \mid h) 1/ h$

وأن k_h (,) : تمثل دالة كاوس (Gaussian function) المستعملة في هذا البحث و بالشكل الأتي :

$$K_h(X^T \beta) = \frac{1}{h\sqrt{2\pi}} \exp \left[\frac{-1}{2} \left(\frac{X^T \beta}{h} \right)^2 \right] \quad , \qquad -\infty < X^T \beta < \infty$$

اذ أن h : تمثل عرض الحزمة (Bandwidth) والتي يمكن تقديرها من خلال طريقة العبور الشرعي (Cross – validation (CV) و هي الطريقة المستعملة في هذا البحث .[5][١٤]

٣.٤: طرائق التقدير شبه المعلميه Semiparametric estimation methods

ظهرت العديد من أساليب وتقنيات تقدير و أختيار المتغيرات في النماذج شبه المعلميه في العقدين الماضيين والتي بالطرائق الجزائية الأمر الذي أدى إلى تحسن كبير في كل من دقة التنبؤ و الكفاءة الحسابية لأنموذج المؤشر الواحد شبه المعلمي (SSIM) و فيما يلي بعض الطرائق المستعملة في هذا البحث :-

(١.٣.٤) طريقة المربعات الصغرى الجزائية غير المقعرة مع دالة جزاء – سكاد

Nonconcave penalized least squares method with (SCAD) penalty function method .(SCAD – NPLS) method

بين الباحث Fan و آخرون في عام (٢٠٠٠) أن طريقة – Lasso تنتج مقدرات متحيزة لمعاملات الأنحدار الكبيرة و قد دفعت هذه المشكلة الباحثان Fan & Li في عام (٢٠٠١) إلى أقتراح دالة جزاء جديدة غير مقعرة (Nonparametric penalty function) تدعى (SCAD) و تعني (Smoothing clipped absolute deviation) و بينا أن هذه الطريقة تحتفظ ليس فقط بميزات جيدة لكل من أختيار المجموعة الجزئية و أنحدار الحرف لكن كذلك تنتج حلول مبعثرة (Sparsity) يضمن الأستمرارية للنماذج المختارة (أستقرار أختيار الأنموذج) و تعطي تقديرات غير متحيزة للمعاملات الكبيرة .[3]

وأظهر الباحثان Fan & Li أن دالة جزاء – (SCAD) لها ثلاث خصائص مرغوب فيها و هي عدم التحيز (Unbiasedness) ، التبعثر (Sparsity) و الأستمرارية (Continuity) .[5]

وتم تطبيق طريقة – (SCAD) من قبل الباحثين Fan & Li عام (٢٠٠١) على الانموذج الخطي العام (GLM) و الحصول على تقديرات هذه الطريقة من خلال الصيغة الاتية :

$$\underline{\hat{\beta}}^{SCAD} = \underset{\beta}{\operatorname{argmin}}\{ \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^{p} p_{\lambda}^{SCAD} (| \beta_j |) \} \qquad (12)$$

اذ أن :

$p_\lambda(\beta_j)$: تمثل دالة جزاء – SCAD و تعرف بالشكل الآتي :

$$p_\lambda^{SCAD}(|\beta_j|) = \begin{cases} \lambda\,|\beta_j| & \text{if } |\beta_j| \le \lambda \\ \frac{-|\beta_j|^2 - 2a\lambda|\beta_j| - \lambda^2}{2(a-1)} & \text{if } \lambda < |\beta_j| \le a\lambda \\ \frac{(a+1)\lambda}{2} & \text{if } |\beta| > a\lambda \end{cases} \quad \text{------ (13)}$$

وأن :

a : تمثل معلمة ضبط ثابتة وقيمتها اكبر من (٢) واقترح الباحثان Fan & Li استعمال قيمة a = ٣.٧ بسبب ان هذه القيمة تعطي اداء مرضي لمسائل اختيار المتغير المختلفة .[4][9]

λ : تمثل معلمة ضبط أخرى تكون قيمتها اكبر من الصفر و يجب تحديدها بأستعمال طريقة – (BIC) .

إذ تكون دالة جزاء – (SCAD) متماثلة ومستمرة و قابلة للتفاضل و تكون مشتقتها مساوية للصفر خارج الفترة $[-a\lambda, a\lambda]$.

والحلول لدالة جزاء – (SCAD) يمكن أن تحدد بالشكل الآتي :-[9][١٣]

$$\hat{\beta}_j^{SCAD} = \begin{cases} sign(\hat{\beta}_j^{OLS})\,(|\hat{\beta}_j^{OLS}| - \lambda) & \text{if } \hat{\beta}_j^{OLS} \le 2\lambda \\ \frac{[(a-1)\hat{\beta}_j^{OLS} - sign(\hat{\beta}_j^{OLS})a\lambda]}{(a-2)} & \text{if } 2\lambda < \hat{\beta}_j^{OLS} \le a\lambda \\ \hat{\beta}_j^{OLS} & \text{if } \hat{\beta}_j^{OLS} > a\lambda \end{cases} \quad \text{------ (14)}$$

وقد بينوا أن النتائج من المعاملات الصغيرة لهذه الطريقة يتم مساواتها بالصفر و عدد قليل من المعاملات الأخرى يجري تقليصها نحو الصفر مع الأحتفاظ بالمعاملات الكبيرة ، فدالة جزاء – (*SCAD*) تنتج مجموعة مبعثرة (Sparsity) من الحلول غير المتحيزة للمعاملات الكبيرة .[3]

أن طريقة – (SCAD) ليس فقط تختار المتغيرات المهمة ولكن كذلك تنتج مقدرات للمعلمة بكفاءة و كأن الأنموذج الصحيح معروف و هذه الميزة لطريقة – (SCAD) تعتمد على الأختيار السليم لمعلمة الضبط (Tuning parameter) أو معلمة الجزاء (Penalty parameter) (λ) والتي عادةً ما يتم تحديدها بواسطة طريقة العبور الشرعي العام (GCV) Generalized cross – validation إذ تبين للباحثين Fan & Li عام (٢٠٠١) أن اختيار معلمة الضبط المثالية بواسطة طريقة (GCV) تفشل في أختيار المتغير ولذلك تم أقتراح طريقة – (BIC) في عام (٢٠١١) لأختيار معلمة الضبط لطريقة – (SCAD) من قبل الباحثين Heng peng & Tao Huang وقد أثبتوا أن هذا الأجراء المقترح يحدد الأنموذج الصحيح بأستمرار. جديدة لمعيار (BIC) لطريقة دالة جزاء – SCAD بموجب الصبغة الآتية :

$$\text{BIC}(\lambda) = \sqrt{\frac{2\log(n)}{n\,(a+1)}}\,\hat{\sigma} \quad \text{--------------------------------} \quad (15)$$

اذ أن : (a) تمثل معلمة ضبط ثابتة و قيمتها a = ٣.٧ و هي مقترحة من قبل الباحثان Fan & Li عام (٢٠٠١) . [8][٦]

إن الفكرة الأساسية لهذه الطريقة تتمثل باستبدال تقديرات المربعات الصغرى الاعتيادية (OLS) من خلال تقدير المربعات الصغرى الجزائية غير المقعرة (NPLS) .

اذ تم أقتراح خوارزمية تكرارية لتقدير المعلمات و دالة الربط لأنموذج المؤشر الواحد و بين الباحثان أن هذه الطريقة تستطيع تقدير المعلمات و أختيار المتغيرات في آن واحد .

ولتقدير أنموذج المؤشر الواحد شبه المعلمي (SSIM) بطريقة (SCAD – NPLS) تم اتباع نفس فكرة الباحث Carroll وآخرون عام (١٩٩٧) إذ أستعملوا خوارزمية تكرارية لتقدير متجه المعلمات β ودالة الربط g(.) في آن واحد .

ولأعطاء تقدير لـ β و دالة الربط g(.) يمكن التقريب موضعياً من خلال الدالة الخطية الآتية :-

$$g(v) \approx g(u) + \acute{g}(u)(v-u) \equiv a + b(v-u)$$

اذ أن :

v : هو مجاور إلى u

و أن :

$$a = g(u)$$

$$b = \acute{g}(u) = \frac{\partial g(u)}{\partial u}$$

هي ثوابت موضعية .

في الجانب العملي و من خلال التقريب الخطي الموضعي مع دالة Kernel المتماثلة ، نستطيع تقدير $a = g(u)$ و $b = \acute{g}(u)$ بواسطة الأنحدار الخطي الموضعي من خلال أختيار a و b التي تقلل المعادلة الآتية :

$$\sum_{i=1}^{n} [y_i - a - b(x_i^T\beta - u)]^2 . k_h(x_i^T\beta - u) \quad \text{---------------} \quad (16)$$

اذ أن :

K(.) : تمثل دالة kernel المتماثلة .

h : يمثل معلمة عرض الحزمة Bandwidth

يحدد التقدير لدالة الربط g(.) و هي $\hat{g}$(.) و يمكن تحديث التقدير لـ β من خلال تقليل دالة المربعات الصغرى الجزائية غير المقعرة الآتية :

$$\sum_{i=1}^{n} [y_i - \hat{g}(x_i^T\beta)]^2 + n\sum_{i=1}^{p} P_\lambda(|\beta_j|) \quad \text{----------------} \quad (17)$$

اذ أن :

$P_\lambda(|\beta_j|)$: تمثل دالة جزاء – (SCAD) مع معلمة الضبط (λ) .

ومع ذلك فأن دالة الربط التقديرية g(.) قد تكون ليست دالة خطية لحل المعادلة و هي مسألة غير خطية مثلى. و لحل المسأله غير الخطية المثلى المذكورة آنفاً ، أقترح الباحثان Peng & Huang في عام (٢٠١١) فكرة التقريب الموضعي و تحديث التقدير لـ β من خلال تقليل دالة المربعات الصغرى الجزائية غير المقعرة والتي تعطي تقديرات دقيقة لـ $\beta^{(\cdot)}$ ودالة الربط $\hat{g}$(.) بموجب المعادلة الآتية:

$$\sum_{i=1}^{n} [y_i - \hat{g}(x_i^T\beta^{(0)}) - \hat{\acute{g}}(x_i^T\beta^{(0)})(x_i^T\beta - x_i^T\beta^{(0)})]^2 + n\sum_{j=1}^{p} P_\lambda(|\beta_j|) \quad \text{---} \quad (18)$$

ويكون أجراء التقدير لـ β و g(.) لطريقة (SCAD – NPLS) لأنموذج المؤشر الواحد بموجب الخوارزمية الآتية :

<u>الخطوة (٠)</u> : وهي خطوة التهيئة للحصول على تقدير أولي (أبتدائي) لـ β وليكن $\hat{\beta}$ ، بأستعمال طريقة المربعات الصغرى الأعتيادية (OLS) ونفرض أن :

$$\hat{\beta} = \frac{\hat{\beta}_1}{\|\hat{\beta}_1\|} . \text{sign}(\hat{\beta}_{11})$$

اذ أن :

sign ($\hat{\beta}_{11}$): تمثل أشارة العنصر الأول إلى $\hat{\beta}$ ، وأن $\hat{\beta}_{11} > ٠$ ويكون $\|\hat{\beta}\| = ١$ لغرض التشخيص.

<u>الخطوة (١)</u> : نحدد $\hat{\beta}$ و يتم الحصول على $\hat{b} = \hat{\acute{g}}(u \text{ ، } \hat{\beta})$ ، $\hat{a} = \hat{g}(u \text{ ، } \hat{\beta})$ بأستعمال الأنحدار الخطي الموضعي من خلال حل المعادلة الآتية :

$$(\hat{a}, \hat{b}) = \underset{a,b}{\text{argmin}} \sum_{i=1}^{n} [y_i - a - b(x_i^T\hat{\beta} - u)]^2 . k_h(x_i^T\hat{\beta} - u)$$

اذ أن :

k(.) : تمثل دالة **kernel** المتماثلة و تم استعمال دالة **Gaussian** .

h : تمثل معلمة عرض الحزمة **Bandwidth** ونختارها لتكون الأمثل بأستعمال طريقة العبور الشرعي **(CV)**.

الخطوة (٢) : تحديث التقديرلـ β من خلال تقليل دالة المربعات الصغرى الجزائية غير المقعرة $\beta^{(\cdot)}$ وتحل محلها $\widehat{\beta}$ أي يتم أفتراض ($\beta^{(\cdot)} = \widehat{\beta}$) وتحديد $\widehat{a}$ $\widehat{b}$, للحصول على $\widehat{\beta}$.

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}}\{\sum_{i=1}^{n} [y_i - \widehat{g}(x_i^T\beta^{(0)}) - \widehat{g}' (x_i^T\beta^{(0)})(x_i^T\beta - x_i^T\beta^{(0)})]^2 + n\sum_{j=1}^{p} P_{\lambda} (|\beta_j|)\}$$

الخطوة (٣) : يتم الأستمرار بالخطوتين (١) و (٢) حتى التقارب .

الخطوة (٤) : يعطي تقدير نهائي لـ $\hat{\beta}^{SCAD}$ من خلال الخطوة (٣) .

ونكرر تقدير (u ، $\widehat{\beta}$) $\widehat{g}$ لـ g(.) من خلال حل المعادلة في الخطوة (١) وبالشكل الآتي :

$$(\hat{a} , \hat{b}) = \underset{(a,b)}{\operatorname{argmin}} \sum_{i=1}^{n} [y_i - a - b (x_i^T \hat{\beta}^{SCAD} - u)]^2 . k_h (x_i^T \hat{\beta}^{SCAD} - u)$$

و يكون (u ، $\widehat{\beta}$) $\hat{a} = \widehat{g}$ هو التقدير النهائي.

٢.٣.٤ الطريقة المقترحة الاولى: طريقة تقدير التباين لأدنى معدل مع دالة الجزاء – سكاد .

Minimum average variance estimation (MAVE) Method with Smoothly clipped absolute deviation (SCAD) penalty function.

(SCAD – MAVE) Method.

الطريقة المقترحة تتضمن دمج طريقة – **MAVE** للباحث (**Xia** و آخرون عام (٢٠٠٢)) مع دالة جزاء – **SCAD (Smoothly clipped absolute deviation)** للباحثين (**Fan & Li** عام (٢٠٠١)) . وذلك بأضافة دالة جزاء – **SCAD** الى دالة الخسارة لطريقة **(MAVE)** المتمثلة بالمعادلة الآتية :-

$$\min \sum_{j=1}^{n}\sum_{i=1}^{n}\left[y_i - a_j - b_j X_{ij}^T\beta\right]^2 . w_{ij}$$

لتنتج تقدير التبعثر (**A sparse estimate**) و نتيجة لذلك فأن العديد من المعاملات هي أصفار و لتحديد المتغيرات مع المعاملات غير الصفرية تلقائيا" و إلى تقدير أنموذج المؤشر الواحد **(SIM)** أقترح الباحثان **(Fan &Li)** أستعمال دالة جزاء **(SCAD)** و يمكننا أيضاً تنفيذ أختيار المتغير و تقدير أنموذج المؤشر الواحد **(SIM)** في آن واحد من خلال فرض جزاء **(SCAD)** .

و فكرة الطريقة المقترحة **(SCAD – MAVE)** هي الحاجة إلى تقدير المؤشر الواحد **Single – index** من خلال الصيغة الآتية:-

$$\underline{\hat{\beta}}^{SCAD} = \operatorname{argmin} \sum_{j=1}^{n}\sum_{i=1}^{n}\left[y_i - a_j - b_j X_{ij}^T\beta\right]^2 w_{ij} + n\sum_{k=1}^{p} p_{\lambda}^{SCAD} (|\beta_k|) \quad --- \quad (19)$$

$$\beta : \|\beta\| = 1$$
$$a_j, b_j, j = 1,2,\ldots n$$

اذ أن :

(p_{λ}^{SCAD} $|\beta_k|$) : تمثل دالة جزاء – **SCAD** التي تقلل الأنحدار الآتي :

$$\sum_{i=1}^{n}[y_i - g(X_i^T\beta)]^2 + n\sum_{k=1}^{p} p_{\lambda}^{SCAD} (|\beta_k|) \quad ------------------- \quad (20)$$

ولحساب مسألة التقليل في المعادلة (٢-٣٧) يمكن أن تحلل إلى تقليل مسألتين وفق خوارزمية (**SCAD – MAVE**) لأنموذج المؤشر الواحد شبه المعلمي **(SSIM)** .

<u>الخطوة (٠)</u> : نفرض تقدير أولي (أبتدائي) للمعلمة $\widehat{\beta}^{(0)}$ بأستعمال طريقة المربعات الصغرى الأعتيادية (OLS).

<u>الخطوة (١)</u> : يتم تثبيت $\hat{\beta}^{(0)} = \hat{\beta}$ و نحسب قيمة الحل الى $(\hat{a}_j , \hat{b}_j)$ من خلال الصيغة الآتية :-

$$\begin{pmatrix} \hat{a}_j \\ \hat{b}_j \end{pmatrix} = \left\{ \sum_{i,j}^{n} w_{ij}^{\hat{\beta}^{(0)}} \begin{pmatrix} 1 \\ X_{ij}^T \hat{\beta} \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij}^T \hat{\beta} \end{pmatrix}^T \right\}^{-1} . \sum_{i,j}^{n} w_{ij}^{\hat{\beta}^{(0)}} \begin{pmatrix} 1 \\ X_{ij}^T \hat{\beta} \end{pmatrix} Y_i$$

<u>الخطوة (٢)</u> : يتم تثبيت $(\hat{a}_j , \hat{b}_j)$ و تقدير متجه المعلمات β وفق الصيغة الآتية :-

$$\hat{\beta}^{SCAD-MAVE} = \underset{\beta: \|\beta\| = 1}{\text{argmin}} \{ \sum_{j=1}^{n} \sum_{i=1}^{n} \{ y_i - \hat{a}_j - \hat{b}_j X_{ij}^T \beta \}^2 w_{ij}^{\hat{\beta}^{(0)}} + n \sum_{k=1}^{p} p_{\lambda}^{SCAD} (|\beta_k|) $$

لتبسيط الصيغه نفترض أن:

$$a_j, b_j \; j = 1,2, \dots, n$$

$$Y_{ij}^* = Y_i \, (w_{ij}^{\hat{\beta}^{(0)}})^{1/2} - \hat{a}_j (w_{ij}^{\hat{\beta}^{(0)}})^{1/2}$$

$$X_{ij}^* = \hat{b}_j \, X_{ij}^T (w_{ij}^{\hat{\beta}^{(0)}})^{1/2}$$

تصبح المسأله التي تقلل بالشكل الآتي :-

$$\hat{\beta}^{SCAD-MAVE} = \underset{\beta}{\text{argmin}} \sum_{i=1}^{n} [Y_{ij}^* - X_{ij}^{*T} \beta]^2 + n \sum_{k=1}^{p} p_{\lambda}^{SCAD} (|\beta_k|)$$

<u>الخطوة (٣)</u> : يتم تكرار الخطوتين (١) و (٢) مع $\frac{\hat{\beta}}{\|\hat{\beta}\|} = \hat{\beta}^{(0)}$ حتى التقارب و المتجه الأخير هو متجه مقدر (SCAD – MAVE) لـ $\hat{\beta}^{(0)}$ و يعرف من خلال $\hat{\beta}^{SCAD-MAVE}$

و تقدير دالة الربط g(.) النهائي يكون: $\hat{a}_j = \hat{g}\,(u\,,\, \hat{\beta}^{SCAD-MAVE})$.

<u>(٣.٣.٤) الطريقة المقترحة الثانية :</u>

طريقة تقدير التباين لأدنى معدل مع دالة جزاء – (Adaptive LASSO).

Minimum average variance estimation (MAVE) method With Adaptive least absolute shrinkage and selection operator (lasso) penalty function.

(ALASSO – MAVE) Method

بين الباحث Tibshirani في عام (١٩٩٦) أن أسلوب (Lasso) يفتقر إلى خصائص أوراكل (Oracle properties) كما توصل الباحثان Fan &Li عام (٢٠٠١) أن هذا الأسلوب لديه تحيز في تقدير المعاملات غير الصفرية الكبيرة و أظهروا أيضاً انه لا يمتلك خصائص أوراكل مما دفع الباحث Zou عام (٢٠٠٦) إلى أقتراح أسلوب جديد يدعى لاسو التكيفية (Adaptive Lasso) للنموذج الخطي العام إذ أن فكرة هذا الأسلوب يعمل على تعيين أوزان تكيفية مختلفة للمعاملات المختلفة الجزائية في دالة جزاء – L_1 مما يؤدي
إلى زيادة الجزاء للمعاملات التي تقترب من الصفر و من ثم أختزال التحيز في تقدير الدالة و تحسين دقة أختيار المتغير .[12]

وبين الباحث Zou عام (٢٠٠٦) أن طريقة لاسو التكيفية (ALasso) هي طريقة وزن جزائية لدالة جزاء- L_1 لتقدير وأختيار الأنموذج في آن واحد و لها خصائص أوراكل المتمثلة بالطبيعي المحاذى و نسبة التقارب المثلى و الأتساق في أختيار المتغير .

وصيغة مقدر (Alasso) لمتجه المعلمات β لأنموذج الخطي العام (GLM) من خلال تقليل المعادلة الآتية [9][6]:

$$\hat{\beta}^{ALASSO} = \underset{\beta}{\operatorname{argmin}} \left\{\sum_{i=1}^{n}(y_i - X_i^T\beta)^2 + \lambda\sum_{k=1}^{p}\widehat{W}_k|\beta_k|\right\} \text{---------- } (21)$$

وبناءا" على ما تم ذكره آنفا" فأن مقدر لاسو التكيفية (Alasso) لـ $\hat{\beta}$ و لأنموذج المؤشر الواحد شبه المعلمي (SSIM) يمكن الحصول عليه من خلال الصيغة الآتية :-

$$\hat{\beta}^{ALASSO} = \operatorname{argmin}\left\{\sum_{i=1}^{n}(y_i - g(X_i^T\beta))^2 + \lambda\sum_{k=1}^{p}\widehat{W}_k|\beta_k|\right\} \text{------ } (22)$$

اذأن:

$\lambda\sum_{k=1}^{p}\widehat{W}_k|\beta_k|$: تمثل دالة جزاء - (Alasso) مع معلمة الضبط (λ):وتحسب من خلال الصيغة الاتية:

$$BIC(\lambda) = \log(\hat{\sigma}) + df(\lambda)\frac{\log(n)}{n} \text{ ----------------- } (٢٣)$$

اذ أن :

$\hat{\sigma}$: تمثل القيمة التقديرية للأنحراف المعياري للخطأ العشوائي و يحسب وفق الصيغة الآتية :

$$\hat{\sigma} = \sqrt{\frac{1}{n-d}\sum_{i=1}^{n}[yi - \hat{g}(x_i^T\hat{\beta})]^2} \text{ -------------------------- } (24)$$

d : تمثل البعد لـ $\hat{\beta}$

$\hat{g}(x_i^T\hat{\beta})$: تمثل القيمة التقديرية لدالة الربط .

df (λ) : تمثل درجة حرية الأنموذج و تحدد من خلال عدد المعالم المقدرة غير الصفرية في $\hat{\beta}$ كما يرى الباحثون Tibshirani , Zou & Hastie عام (٢٠٠٧) .

n : يمثل حجم العينة

λ : تمثل معلمة الجزاء أو معلمة الضبط .

ونتائج معلمة الجزاء أو معلمة الضبط المثلى تعرف من خلال $\hat{\lambda}_{BIC}$.

$w = (w_1, w_2, \ldots, w_p)$: تمثل الأوزان التكيفية (Adaptive weight) و بين الباحث Zou أنه إذا كان أختيار الأوزان (w_k) بكفاءة وبطريقة تعتمد على البيانات فأن طريقة لاسو التكيفية (Alasso) يمكن أن نحقق خصائص أوراكل بحيث ينفذ كما لو كان الأنموذج الصحيح معروف و قد أقترح أستعمال الأوزان المقدرة وبالشكل الآتي :-

$$\widehat{W}_k = \frac{1}{|\hat{\beta}_k|^{\gamma}} \quad , \qquad k = 1,2,\ldots,p$$

من خلال أستعمال تقديرات المربعات الصغرى الأعتيادية (OLS) لأختيار $\widehat{W}_k$ اذأن $\hat{\beta}_k$ تمثل تقديرات (OLS) و هي مقدر أبتدائي متسق ـــ $\sqrt{n}$ لـ β (يحتوي على $\sqrt{n}$ نسبة التقارب) .

و أن γ : يمثل معلمة الأنكماش و قيمتها أكبر من الصفر و نفترض أن تكون قيمتها تساوي واحد (γ = ١) . [٩]

وبناءاً على فكرة الباحثان (Zho & Zang He) في عام (٢٠١١) اللذان أستعمالا طريقة النوع ـ Lasso و طريقة ـــ SCAD لتقدير وأختبار المتغير لأنموذج المؤشر الواحد شبه المعلمي (SSIM) تم اقتراح خوارزمية من قبل الباحث لطريقة (Adaptive LASSO- MAVE) من خلال دمج دالة الخسارة لطريقة ـ MAVE مع دالة جزاء - Adaptive LASSO وتوظيفها لتقدير و أختبار المتغير لأنموذج المؤشر الواحد في المعادلة الآتية :-[8]

$$\sum_{i=1}^{n}[y_i - a - b(X_i^T\beta - u)]^2 K_h(X_i^T\beta - u) \text{ ------------- } (٢٥)$$

وتم أستعمال طريقة المربعات الصغرى الجزائية لدالة جزاء (ALASSO) لمسألة تقليل المعادلة الآتية و التي تعرف بدالة الهدف :-

$$\min_{a,b,\beta,\|\beta\|=1} = \sum_{j=1}^{n}\sum_{i=1}^{n}\left[y_i - a_j - b_j\beta^T(x_i - X_j\right]^2 w_{ij} + \lambda\sum_{k=1}^{p}\ \widehat{w}_k\,|\beta_k| \quad \text{--} \qquad (26)$$

نلاحظ أن الجزء الأول للمعادلة (٢٦) هو دالة الخسارة إلى طريقة ــ (MAVE) لتقدير قيمة المعلمات β ولها جمع داخلي هو :-

$$\sum_{i=1}^{n}\left[y_i - a_j - b_j\beta^T(x_i - X_j)\right]^2 w_{ij}$$

والجزء الثاني للمعادلة (٢٦) يمثل دالة جزاء لاسوالتكيفية (Adaptive lasso penalty function) وهذا الجزء يجعل $\hat{\beta}$ متناثرة (Sparsity) و من ثم يؤدي إلى أختيار المتغير (Variable selection) .

و يمكن تلخيص الخوارزمية المقترحة لطريقة (ALASSO – MAVE) لأنموذج المؤشر الواحد شبه المعلمي (SSIM) و التي تعمل على تقدير و أختيار المتغير في آن واحد بالخطوات الآتية :-

<u>الخطوة (٠)</u> : نحصل على تقدير أولي (أبتدائي) لـ β وليكن $\hat{\beta}^{(0)}$ بطريقة المربعات الصغرى الأعتيادية (OLS) و نفرض أن :

$$\hat{\beta} = \frac{\hat{\beta}^{(0)}}{\|\hat{\beta}^{(0)}\|}\ \text{sign}\ (\hat{\beta}_1^{(0)})$$

اذ أن :

sign ($\hat{\beta}_1^{(0)}$): يمثل أشارة العنصر الأول لـ $\hat{\beta}^{(0)}$ و قيمته أكبر من الصفر و أن $\|\hat{\beta}\| = ١$ لغرض التشخيص.

<u>الخطوة (١)</u> : نحدد $\hat{\beta}$ و يتم الحصول على $\{\hat{a}_j, \hat{b}_j\}_{j=1}^{n}$ من خلال حل المعادلة الآتية :-

$$(\hat{a}, \hat{b}) = \underset{a_j, b_j}{\text{argmin}}\ \sum_{i=1}^{n}\left[y_i - a_j - b_j(x_i - X_j)^T\ \hat{\beta}\right] w_{ij}$$

إذ أن :

$$W_{ij} = \frac{k\left(\frac{x_i^T\hat{\beta} - x_j^T\hat{\beta}}{h}\right)}{\sum k\left(\frac{x_i^T\hat{\beta} - x_j^T\hat{\beta}}{h}\right)}$$

k(.) : تمثل دالة kernel المتماثلة و تم أستعمال دالة (Gaussian) .

h : تمثل معلمة عرض الحزمة (Bandwidth) و نختارها لتكون الأمثل و تحسب من خلال طريقة العبور الشرعي (CV) .

<u>الخطوة (٢)</u> : نحدد $\{\hat{a}_j, \hat{b}_j\}_{j=1}^{n}$ للحصول على $\hat{\beta}^{ALASSO-MAVE}$ من خلال حل المعادلة الآتية :-

$$\hat{\beta}^{ALASSO-MAVE} = \underset{\beta}{\text{argmin}}\left\{\sum_{j=1}^{n}\sum_{i=1}^{n}\left[y_i - \hat{a}_j - \hat{b}_j(X_i - X_j)^T\beta\right]^2 W_{ij} + \lambda\sum_{k=1}^{p}\widehat{W}_k|\beta_k|\right\}$$

ويمكن تبسيطها بالشكل الاتي:-

$$\hat{\beta}^{ALASSO-MAVE} = \underset{\beta}{\text{argmin}}\left\{\sum_{j=1}^{n}\sum_{i=1}^{n}\left[Y_{ij}^* - X_{ij}^{*T}\beta\right]^2 W_{ij}^* + \lambda\sum_{k=1}^{p}\widehat{W}_k|\beta_k|\right\}$$

اذ ان :

$$Y_{ij}^{*} = y_i - \hat{a}_j \ \ \text{and} \ \ X_{ij}^{*} = \hat{b}_j \left(X_i - X_j\right)^T$$

في الخطوة الثانية يتم تقدير β للمشاهدات $\{Y_{ij}^{*}, X_{ij}^{*}\}_{i,j=1}^{n}$ مع الأوزان $\{W_{ij}^{*}\}_{j=1}^{n}$.

الخطوة (٣) :يتم الأستمرار بتكرار الخطوتين (١) و (٢) حتى التقارب و يكون التقدير النهائي لـ $\hat{\beta}^{ALASSO-MAVE}$.

و التقدير النهائي لـ g(.) هو ($\hat{\beta}^{ALASSO-MAVE}$، u) $\hat{a} = \hat{g}$ ويتم الحصول عليه من خلال حل المعادلة الآتية :-

$$(\hat{a}, \hat{b}) = \underset{a,b}{\text{argmin}} \left\{ \sum_{i=1}^{n} \left[y_i - a - b(X_i^T \hat{\beta}^{ALASSO-MAVE} - u) \right]^2 k \left(\frac{X_i^T \hat{\beta}^{ALASSO-MAVE} - u}{h} \right) \right\}$$

٥ . معايير المقارنة:- [12] [١]

هناك العديد من المعايير تقيس مقدار الكفاءة في تقدير دالة الأنحدار التي تم تناولها نظرياً لأنموذج المؤشر الواحد شبه المعلمي , إذ تم أستعمال المعيار الآتي :-

1.5: معيار (AMSE)

و يمثل معدل متوسط مربعات الخطأ (Average Mean squares error)

$$\text{AMSE} = \frac{1}{n} E \sum_{i=1}^{n} \left(y_i - \hat{g}(X_i^T \hat{\beta}) \right)^2 \quad \text{-----------------------} \quad (27)$$

اذ ان

$\hat{g}(.)$: تمثل القيم التقديرية لدالة الربط g(.) .

n : يمثل عدد المشاهدات .

ويكون الأنموذج الأفضل الذي يعطي أقل قيمة لـ (AMSE) .

٦.الجانب التجريبي:

تم في هذا المبحث استعمال الاسلوب التجريبي (المحاكاة) في مقارنة طرائق التقدير شبه المعلميه لانموذج المؤشر الواحد لبيان افضل الطرائق المستعملة والتي تمثل البيانات تمثيلا" سليما" .

لقد تضمنت تجارب المحاكاة الحالات الافتراضية المبينة في الجدول رقم (١) لغرض تقدير وأختيار المتغير لأنموذج المؤشر الواحد شبه المعلمي (SSIM) من خلال تطبيق طرائق التقدير شبه المعلميه المستعملة في هذا البحث التي يتم من خلالها تحقيق هدف البحث . وتم الأعتماد على برنامج (R-package للحصول على النتائج) .

جدول رقم (١)

يمثل القيم الافتراضية لتجارب المحاكاة

Experiment	P	σ =1			σ = 0.5			=0.1 σ		
		n			n			n		
I	٧	٥٠	100	200	٥٠	100	200	٥٠	100	200

اذ أن كل تجربة من التجارب الثلاث بحثت عند ثلاثة مستويات للتباين (σ =1,0.5,0.1) وبحجوم عينات مختلفة وهي (n = 50 , 100, 200) وعدد متغيرات توضيحية(p= ٧) وكررت كل تجربة (٤٠٠) مرة.ولكل دالة من الدوال المدروسة الآتية :

الدالة الاولى [٧]
$$g(X^T\beta) = \sin(\frac{\pi X^T\beta}{10})$$

الدالة الثانية [1]
$$g(X^T\beta) = \sin(2\pi (X^T\beta)^3)$$

مع القيم الافتراضية لمتجه المعلمات الاتية

$$\hat{\beta}^{(0)} = (-0.001, -0.002, 0.000, 0.000, 0.310, 0.310, -0.210)^T$$

اذ تم الحصول عليها من البيانات الحقيقية باستعمال طريقة المربعات الصغرى الاعتيادية (OLS).

مناقشة نتائج تجارب المحاكاة:

في هذا الجزء سيتم عرض وتحليل نتائج تجارب المحاكاة في تقدير وأختيار المتغير للأنموذج قيد البحث ولمختلف دوال الأنحدار المستعملة . ومن الجدير بالذكر انه تم أستعمال دالة **Gaussian kernel** لأختيار الدالة اللبية **kernel** وبالاعتماد على برنامج **R** وتم الحصول على النتائج لكل دالة من الدوال المستعملة بالاعتماد على برنامج **R** والموضحة من جدول رقم (٢) الى جدول رقم (٤) والتي سيتم تحليلها لاحقا".

جدول رقم (٢) يمثل مقارنة طرائق التقدير شبه المعلميه لانموذج المؤشر الواحد باستعمال معيار المقارنة (AMSE) $g(X^T\beta) = \sin(\pi X^T\beta/10)$. لحالة الدالة الاولى

methods	n	P=7		
		$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.1$
SCAD-NPLS	50	٠.٧٦٦٧٠	٠.٢٢٨٣٤	٠.٠٧٣٧٣
	100	١.١٣٨٠٦	٠.٤٠١٧٩	٠.٠٨٢١٩
	200	١.٢٥٨٩٦	٠.٤٢٥١٩	٠.٢٥١٠٦
SCAD-MAVE Suggest-1	50	٠.٧٦٧٦٨	٠.٢٢٨٢٤	٠.٠٧٣٩٣
	100	١.١٣٨٠٥	٠.٤٠٠٩٩	٠.٠٨١٨٢
	200	١.٢٥٨٩٥	٠.٤٢٥٢١	٠.٢٥١٠٩
ALASSO-MAVE Suggest-2	50	٠.٧٧٢٠٨	٠.٢٣٧٢٠	٠.٠٩٥٦٦
	100	١.١١٥٠٠	٠.٣٤٩٨٩	٠.٠٩٠٩٧
	200	١.١٧٥٦٩	٠.٣٣٢٩١	٠.٢٠١٧٦

جدول رقم (٣) يمثل مقارنة طرائق التقدير شبه المعلميه لانموذج المؤشر الواحد باستعمال معيار المقارنة (AMSE)

methods	n	P=7		
		$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.1$
SCAD-NPLS	50	١.٤٣٧٦٥	٠.٧٤٠٩٠	٠.٧٢٠٠٧
	100	١.٥٤٠١٤	٠.٩٢٦٦٢٨٦٩	٠.٥٠٩٢٨٦
	200	١.٤٥١٠٣	٠.٧٤١١٤	٠.٦٢٨٠٣
SCAD-MAVE Suggest-1	50	١.٤٣٥٨٤	٠.٧٤٠٩١	٠.٧٢٠٠٦
	100	١.٥٣٨٨٨	٠.٩٢٦٦٢٨٦١	٠.٥٠٩٢٨٠
	200	١.٤٥١٠٤	٠.٧٣٧٠١	٠.٦٢٨٠٢
ALASSO-MAVE Suggest-2	50	١.٤٩١٧٠	٠.٧٥٠٦٦	٠.٦١٨٩٥
	100	١.٥٤٢١٦	٠.٩٢٥٠٦٩	٠.٦١٨٩٦
	200	١.٣٩٨٠٥	٠.٧٥٣٠٦٧	٠.٦٣٠٨٧

لحالة الدالة الثانية $g(X^T\beta) = \sin(2\pi(X^T\beta)^3)$.

جدول رقم (٤) يوضح افضلية طرائق التقدير شبه المعلميه لانموذج المؤشر الواحد عند اختلاف الدوال
وحجوم العينات وباختلاف قيم تباينات الاخطاء

Functions	n	P=7		
		$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.1$
الدالة الاولى	50	SCAD-NPLS	Suggest-1 (SCAD-MAVE)	SCAD-NPLS
	100	Suggest-2 (ALASSO-MAVE)	Suggest-2 (ALASSO-MAVE)	Suggest-١ (SCAD-MAVE)
	200	Suggest-2 (ALASSO-MAVE)	Suggest-2 (ALASSO-MAVE)	Suggest-2 (ALASSO-MAVE)
الدالة الثانية	50	Suggest-١ (SCAD- MAVE)	SCAD-NPLS	Suggest-2 -MAVE) (ALASSO
	100	Suggest-1 (SCAD- MAVE)	Suggest-2 -MAVE) (ALASSO	Suggest-1 (SCAD- MAVE)
	200	Suggest-2 -MAVE) (ALASSO	Suggest-1 (SCAD-MAVE)	Suggest-1 (SCAD- MAVE)

لغرض اعطاء صورة واضحة لنتائج تجارب المحاكاة تم تفسير النتائج المبينة في الجداول من ولجميع طرائق التقدير شبه المعلميه نجد الاتي :

اولا": انموذج المؤشر الواحد شبه المعلمي الاول (الدالة الاولى)

من خلال الجدولين رقم (٢) و(٤) نبين الاتي :-

١. في حالة عدد المتغيرات التوضيحية (P=٧) عند حجم العينة الصغيرة (٥٠) اظهرت النتائج ان طريقة (SCAD-NPLS) ولقيم تباينات الاخطاء(٠.١,١) هي افضل طريقة بينما حققت الطريقة المقترحة الاولى الافضلية عند تباين الخطأ (٠.٥) .
٢. حققت الطريقة المقترحة الثانية الافضلية لحجوم العينات المتوسطة والكبيرة (١٠٠,٢٠٠) ولكافة قيم تباينات الاخطاء عدا حالة حجم العينة المتوسطة(١٠٠) عند قيمة تباين الخطأ(٠.١) فكانت الافضلية للطريقة المقترحة الاولى .

ثانيا": انموذج المؤشر الواحد شبه المعلمي الثاني (الدالة الثانية)

من خلال الجدولين رقم (٣) و(٤) نبين الاتي :-

١. في حالة عدد المتغيرات التوضيحية(P=٧) اظهرت النتائج ان طريقة (SCAD-NPLS) هي افضل طريقة في حالة العينة الصغيرة (٥٠) ولقيمة تباين الخطأ (٠.٥) وحققت الطريقة المقترحة الاولى الافضلية عند قيمة تباين الخطأ(١) بينما حققت الطريقة المقترحة الثانية الافضلية عند قيمة تباين الخطأ (٠.١) .
٢. كما اظهرت النتائج ان الطريقة المقترحة الاولى هي الافضل عند حجم العينة المتوسطة (١٠٠) لقيم تباينات الاخطاء (٠.١,١) في حين حققت الطريقة المقترحة الثانية افضلية عند قيمة تباين الخطأ (٠.٥) .
٣. اظهرت النتائج ان الطريقة المقترحة الاولى هي افضل طريقة عند حجم العينة الكبيرة (٢٠٠) لقيم تباينات الاخطاء (٠.١,٠.٥) بينما حققت الطريقة المقترحة الثانية عند قيمة تباين (١) .

٧. الجانب العملي :

في هذا الاجراء يتم عمل دراسة تطبيقية للعوامل المؤثرة على عائد السهم السوقي لجميع الشركات المدرجة في سوق العراق للاوراق المالية اذ تهدف هذه الدراسة الى التعرف على أهم العوامل المؤثرة على عائد السهم من خلال تطبيق طرائق التقدير شبه المعلميه لأنموذج المؤشر الواحد الذي يمثل الدراسة التطبيقية فمن خلال معرفة العوامل المؤثرة ومراقبتها سيساعد المستثمر في أتخاذ قرار البيع أو قرار الشراء .

ويركز هذا الجانب على حساب المركبة المعلميه (الجزء المعلمي) والمركبة اللامعلميه (الجزء اللامعلمي) لانموذج المؤشر الواحد شبه المعلمي (SSIM) اذ تتضمن المركبة المعلميه فقط المعلمات المعنوية غير الصفرية لمتجه المعلمات β ذات بعد ($P\times 1$) والذي يعكس تأثير المتغيرات التوضيحية على التغير في المتغير المعتمد أما المركبة اللامعلمية فمتمثلة بالجزء $g(.)$ وتمثل متجه ذات رتبة($n\times 1$) ومن اجل اظهار دور تاثير المتغيرات التوضيحية على المتغير المعتمد فقد تم أخذ بيانات سنوية لعائد السهم السوقي في العراق.

ويتكون مجتمع الدراسة من جميع الشركات التجارية العراقية المدرجه في سوق العراق للاوراق المالية خلال الفترة من ٢٠٠٨م وحتى ٢٠١١م والتي تتكون من (73) شركة نظامية اذ بياناتها تطرح من خلال هذا السوق بأنتظام خلال سنوات الدراسة وموزعة على(٨) قطاعات مختلفة أما حجم عينة الدراسة النهائي فتتكون من (٦٠) شركة اذ تم أختيارها وفقاً الى الشروط التالية وهي ان تكون مدرجة ويتم تداول اسهمها خلال الفترة اعلاه وامكانية الحصول على البيانات أو المعلومات اللازمة لاتمام الدراسة والجدول ادناه يوضح تصنيف حجم العينة حسب القطاعات.

جدول رقم (٥)

يوضح تصنيف حجم العينه للشركات المدرجة في سوق العراق للاوراق المالية حسب القطاعات

الرقم	أسم القطاع	عدد الشركات	عينةالدراسة
١	المصارف	٢١	٢١
٢	التامين	٤	٤
٣	الأستثمار	٢	٢
٤	الخدمات	٩	٧
٥	الصناعة	٢١	١٥
٦	الفنادق والسياحة	١٠	٦
٧	الزراعة	٥	٥
٨	الأتصالات	١	-
الجموع	-	٧٣	٦٠

والمتغيرات المستعملة في الجانب التطبيقي يمكن توضيحها بالشكل الآتي :

بالنسبة للمتغير المعتمد Y والذي يمثل عائد السهم السوقي ويحسب من خلال (صافي ربح السنة مقسوما" على رأس المال المدفوع والذي يمثل رأس المال المستلم فعلا" من قبل الشركة).

أما بالنسبة للمتغيرات التوضيحية فالمتغير$X1$ يمثل نسبة دوران السهم ويحسب من خلال (عدد الاسهم المتداولة لعام كامل مقسوما" على عدد الاسهم الصادرة(رأس المال الاسمي)X١٠٠) . اذ ان رأس المال الأسمي هو رأس المال الذي يحق للشركة أصداره ومسجل في عقد تأسيس الشركة. والمتغير $X2$ يمثل نسبة حقوق الملكية وتحسب من خلال (أجمالي حقوق المساهمين مقسوما"على مجموع الموجودات x ١٠٠) . والمتغير $X3$ يمثل مكرر الأرباح ويحسب من خلال (سعر الأغلاق للسهم مقسوما" على عائد السهم) .

والمتغير X4 يمثل نسبة التداول وتحسب من خلال (الموجودات المتداولة مقسوما" على مصادر التمويل قصيرة الأجل) اذ أن الموجودات المتداولة تتكون من النقديةاو المخزون (بالكلفة) بالأضافة الى المدينون ، أما بالنسبة لمصادر التمويل قصيرة الآجل فهي تشمل الحسابات الجارية والودائع والتخصيصات والدائنون .

والمتغير X5 يمثل القيمة الدفترية للسهم ويحسب من خلال (إجمالي حقوق المساهمين مقسوما" على عدد الاسهم المصدرة(رأس المال الاسمي) . والمتغير X6 يمثل سعر الإغلاق السنوي للسهم ، ويقصد به آخر سعر تداول نفذ على سهم الشركة خلال عام . أما بالنسبة للمتغير X7 فيمثل معدل السعر السنوي للسهم ويحسب من خلال (حجم التداول مقسوما" على عدد الأسهم المتداولة لعام كامل) .

١.٧ جمع البيانات:

تم جمع البيانات من القوائم المالية للشركات التجارية العراقية وبالاعتماد على دليل الشركات الصادر من سوق العراق للأوراق المالية للفترة من العام ٢٠٠٨م وحتى العام ٢٠١١م والخاصة بعائد السهم السوقي Y والعوامل المؤثرة عليه لكافة الشركات.

ان الخطوة الاولى قبل تقدير الانموذج هو فحص العلاقة بين المتغيرات التوضيحية (المستقلة) والمتغير المعتمد بيانيا" للتاكد من كون العلاقة خطية او غير خطية اذ تم الاعتماد على برنامج SPSS لاختبار البيانات من خلال رسم الانتشار لها وتم التوصل الى ان جميع العلاقات بين المتغيرات التوضيحية والمتغير المعتمد هي غير خطية.

٢.٧ مناقشة نتائج البيانات الحقيقية:

تم الحصول على نتائج البيانات الحقيقية باستعمال برنامج R لمختلف طرائق التقدير شبه المعلميه لانموذج المؤشر الواحد المستعمل للتنبؤ بعائد السهم في سوق العراق للاوراق المالية المبين في ادناه :

$$g= iY(x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + x_{i4}\beta_4 + x_{i5}\beta_5 + x_{i6}\beta_6 + x_{i7}\beta_7) + \epsilon_i \qquad ----(28)$$

ومن خلال البيانات الحقيقية واعتماد الانموذج اعلاه يمكن الحصول على تقديرات متجه المعلمات ودالة الربط (دالة عائد السهم) لمختلف طرائق التقدير شبه المعلميه وكما مبينة في الجدولين رقم (٦) و(٧) ادناه :

جدول رقم (٦) يمثل القيم التقديرية لمتجه المعلمات β (الجزء المعلمي) لانموذج المؤشرالواحد شبه المعلمي لجميع الطرائق للبيانات الحقيقية

parameters	methods		
	SCAD-NPLS	SCAD-MAVE Suggest-1	ALASSO-MAVE Suggest-2
	Beta	Beta	Beta
$\widehat{\beta}_1$	٠.٠٠٦٩٦٩	٠.٠٠٦٩٨٥	٠.٠٠٠٠٠٠٠
$\widehat{\beta}_2$	-٠.٣٨٠٥٥٨	-٠.٣٨٠٥٦٣	٠.٠٠٠٠٠٠٠
$\widehat{\beta}_3$	٠.٢٤٥٤٨٧	٠.٢٤٥٥٨٣	٠.٠٠٠٠٠٠٠
$\widehat{\beta}_4$	٠.٥٤٨٥٨٥	٠.٥٤٨٥٨٧	٠.٠٠٠٠٠٠٠
$\widehat{\beta}_5$	٠.٥٤٥٨١٤	٠.٥٤٥٨٣٩	٠.٠٠٠٠٠٠٠
$\widehat{\beta}_6$	٠.٠٠٠٠٠٠٠	٠.٠٠٠٠٠٠٠	٠.٠٠٠٠٠٠٠
$\widehat{\beta}_7$	٠.٤٤٢٧٢٢	٠.٤٤٢٦٤٤	١.٠٠٠٠٠٠٠

جدول رقم (٧)

يمثل القيم التقديرية لدالة الربط $(X)g\beta$ (الجزء اللامعلمي) دالة عائد السهم في سوق العراق للاوراق المالية لانموذج المؤشرالواحد شبه المعلمي لجميع الطرائق للبيانات الحقيقية

methods					
SCAD-NPLS		SCAD-MAVE Suggest-1		ALASSO-MAVE Suggest-2	
$\hat{g}$(xb)	$\hat{g}$(xb)	(xb)$\hat{g}$	$\hat{g}$(xb)	(xb)$\hat{g}$	$\hat{g}$(xb)
٠.٢٢١٩١٦	٠.١٨١٦٢٣	٠.٢٢١٩٥٥	٠.١٨١٥٧٣	٠.٠٩٩٥٥٦	٠.١٥٥٠٥٢
٠.١٧٢٠٧٠	٠.١٩٦٣٣٣	٠.١٧٢١١٣	٠.١٩٦٣٧٣	٠.١٣٤٧٦٣	٠.٣١٨١٥٤
٠.٢٥٥٣٧٧	٠.٢٦١٦٥٧	٠.٢٥٥٤٣٦	٠.٢٦١٦٦٦	٠.٠٩٥٢٢٨	٠.٨٧٥٨٢٢
٠.١٧٣٧٥١	٠.١٨٦٥٠٧	٠.١٧٣٧٨٦	٠.١٨٦٥٤٠	٠.١١٣٣٢٧	٠.٣٩١٨٧٩
٠.٢١٣٦٤٨	٠.٢٢٠٦١٣	٠.٢١٣٦٩١	٠.٢٢٠٦٧٧	٠.٠٩٦٨١٧	٠.١٢١٢٦٤
٠.١٧٦١٧٠	٠.٢١٦٠٥٩	٠.١٧٦٢٢٥	٠.٢١٦٠٠٣	٠.٠٩٥٢٢٨	٠.١٦٤١٥٣
٠.١٧٧٧٦٠	٠.٢١٣٤٦٨	٠.١٧٧٧٩٧	٠.٢١٣٥٨٠	٠.١٦٦١٢٣	٠.٠٩٥٠٢٨
٠.٢٤٠٣٩٣	٠.١٨١٤٩٨	٠.٢٤٠٥١٠	٠.١٨١٤٤٩	٠.١٨٩٤٠٩	٠.٢٤٧٨٩٢
٠.١٨٣٩٧٣	٠.١٤٤٠٥٠	٠.١٨٣٩٨٣	٠.١٤٤٠٠٧	٠.٠٩٥١١١	٠.٠٩٧٥٢٥
٠.٢٠٧٩٤٩	٠.٠٧٢١٣٧	٠.٢٠٨٠٠١	٠.٠٧٢١٨٦	٠.٠٩٥٩٢٢	٠.٠٩٧٨٦١
٠.١٩٩٢٨٨	٠.١٧٧٨٠٣	٠.١٩٩٣٣٦	٠.١٧٧٧٣٢	٠.١٠٩٠٨٦	٠.١٣١٤٤٥
٠.١٩٧٠٢٨	٠.٢٠٧٨٩٩	٠.١٩٧٠٦٩	٠.٢٠٧٨٣٧	٠.٠٩٦٨٩٦	٠.١٦٥٦٣٣
٠.٢٠٦٧٠٩	٠.٢٥٣٦١٩	٠.٢٠٦٧٥٨	٠.٢٥٣٦٩٣	٠.٠٩٦٨٤٤	٠.١٤٣٩٩٩
٠.٢٢٦٢٣١	٠.٢٣١١١٨	٠.٢٢٦١٧٥	٠.٢٣١١١١	٠.٠٩٥١٠٥	٠.١٢٧١٦٨
٠.١٨٩٢٣١	٠.١٥١٤٩٦	٠.١٨٩٢٦٦	٠.١٥١٩٧٩	٠.١٣٦٢٠٣	٠.١٥٤٩١٩
٠.١٩٥٨٢٢	٠.٢٢٦١٢٢	٠.١٩٥٨٦١	٠.٢٢٦٠٦١	٠.١٠٨٤٦٤	٠.١٢٧٢٢٠
٠.٢٣١١٩٩	٠.٢٢٥٣٤٧	٠.٢٣١١٥٠	٠.٢٢٥٣٦٩	٠.٠٩٥٦٩٤	٠.١٧١٨٧٠
٠.١٨٠٣٠١	٠.٢٢٨١٦٦	٠.١٨٠٣٣٤	٠.٢٢٨١٠٧	٠.١٠٤٤٨٧	٠.١٠٤٩٩٩
٠.٢١٩٤٨٥	٠.٢٠٤٦٣٢	٠.٢١٩٥١٨	٠.٢٠٤٧٠٢	٠.١٠٠٤٩٩	٠.١٨٤٧٥٣
٠.٢٢٦١٢٠	٠.١٨٣١٥٣	٠.٢٢٦١٣٨	٠.١٨٣٢٣٥	٠.٠٩٧٣٢١	٠.٥٨٦٦٣٣
٠.٢٢٩٥٩٩	٠.١٢٦٩٤٥	٠.٢٢٩٥٩٧	٠.١٢٧٤٦٦	٠.٠٩٩٩١٠	٠.٥٥٦١٦٨
٠.١٤٩٠١٠	٠.١٧٩٨٠٢	٠.١٤٨٩٧١	٠.١٧٩٨٠٦	٠.٠٩٧٥٥٦	٠.٥٧٤٩٨٤
٠.١٧٥٢٩٠	٠.٢٦٠٢٩٥	٠.١٧٥٣٢٥	٠.٢٦٠٢٩٤	٠.١٥٠٧٥٩	٠.٢٦٢١٧٠
٠.١٠٤٥٠٨	٠.١٨٤٣٨٣	٠.١٠٤٥١٣	٠.١٨٤٤٢٨	٠.٠٩٦١٠٢	٠.٨٢٧٤١٧
٠.٢٣٠٠٠٧	٠.١٧٧٧٥٥	٠.٢٢٩٩٤٢	٠.١٧٧٧٨٣	٠.٠٩٧٦١٧	٠.٧٨٦٠٧٩
٠.١٧٢٤١٣	٠.٢٣٢٢٤٨	٠.١٧٢٤٥٩	٠.٢٣٢٢١٩	٠.١٠٤٩٠٢	٠.٠٩٨١٣٧
٠.١٧٥٩١٩	٠.٢٢٠٦٣٤	٠.١٧٥٨٥٣	٠.٢٢٠٦٨٧	٠.٠٩٥٤٧٨	٠.٤٧٢٣٨٣
٠.١٢٢٩٤٧	٠.١٧٢١٢٠	٠.١٢٢٩٤٧	٠.١٧٢١٦٨	٠.٥٠٩٢٥١	٠.١١٤٨٤٢
٠.٢٣٢٢١٨	٠.٢٠٤٦٨٩	٠.٢٣٢١٧٩	٠.٢٠٤٧٦٠	٠.٨٥٤٨١٧	٠.١٩٠٣٦٢
٠.٠٨٩٣٧٤	٠.١٣١٧٩٦	٠.٠٨٩٨٨٧	٠.١٣١٧٧٧	٠.١٣١٦٢١	٠.١٥٧٨٩١

نلاحظ من خلال النتائج المبينة في الجدول رقم (٦) ان افضل معادلة تنبؤ لانموذج المؤشر الواحد شبه المعلمي بطريقة (SCAD–NPLS) هي مبينة كما في ادناه :

$$\hat{y_i} = \hat{g}_6(\text{٠.٠٠٦٩٦٩}x_{i1} - \text{٠.٣٨٠٥٥٨}x_{i2} + \text{٠.٢٤٥٤٨٧}x_{i3} + \text{٠.٥٤٨٥٨٥}x_{i4} + \text{٠.٥٤٥٨١٤}\,x_{i5} + \text{٠.٤٤٢٧٢٢}\,x_{i7}) \quad (\text{٢٩})$$

وافضل معادلة تنبؤ للطريقة المقترحة الاولى (SCAD-MAVE) هي كما في ادناه :

$$\hat{y_i} = \hat{g}_6(\text{٠.٠٠٦٩٨٥}x_{i1} - \text{٠.٣٨٠٥٦٣}x_{i2} + \text{٠.٢٤٥٥٨٣}x_{i3} + \text{٠.٥٤٨٥٨٧}x_{i4} + \text{٠.٥٤٥٨٣٩}\,x_{i5} + \text{٠.٤٤٢٦٤٤}\,x_{i7}) \quad (\text{٣٠})$$

وافضل معادلة تنبؤ للطريقة المقترحة الثانية (ALASSO-MAVE) هي كما في ادناه :

$$\widehat{y_i} = \widehat{g_1}(\ x_{i7}\ ) \quad \text{-------------------------} \quad (31)$$

والجدول الاتي يوضح نتائج قيم متوسط مربعات الخطأ لدالة عائد السهم للبيانات الحقيقية ولمختلف طرائق التقدير .

جدول رقم (٨) يمثل قيم متوسط مربعات الخطأ (MSE) لجميع طرائق التقدير شبه المعلميه لانموذج المؤشر الواحد للبيانات الحقيقية

Method	MSE
SCAD-NPLS	٠.٠٨٤١٧٧
SCAD-MAVE(suggest-1)	٠.٠٨٤١٨٤
ALASSO- MAVE(suggest-2)	٠.٠٦١٦٣٥

من خلال دراسة نتائج البيانات الحقيقية المبينة في الجداول رقم (٦) و(٧) و(٨) نجد الاتي :

١. اظهرت النتائج ان الطريقة المقترحة الثانية (ALASSO-MAVE) هي افضل طريقة في تقدير متجه المعلمات (الجزء المعلمي) ودالة الربط (دالة عائد السهم) والتي تمثل الجزء اللامعلمي لانموذج المؤشر الواحد شبه المعلمي وهذا يعني ان هذه الطريقة قادره على تمثيل دالة عائد السهم في سوق العراق للاوراق المالية .

٢. كما نلاحظ من خلال النتائج المبينه في الجدول رقم (٨) ان طرائق التقدير شبه المعلميه (SCAD-NPLS)، والطريقة المقترحة الاولى(SCAD-MAVE) اعطت نتائج متقاربة لقيم متوسط مربعات الخطأ (MSE) .

٣. كما اظهرت النتائج المبينة في الجدول رقم (٦) ان الطريقة المقترحة الاولى (SCAD-MAVE) اعطت نتائج متقاربة الى طريقة (SCAD-NPLS) لمتجه المعلمات .

٤. نلاحظ من خلال الجدول رقم (٧) ان القيم التقديرية لدالة الربط (دالة عائد السهم) لجميع طرائق التقدير شبه المعلميه كانت قريبة من القيم الحقيقية للمتغير المعتمد (عائد السهم) .

٨. الاستنتاجات والتوصيات:

في ضوء الجانب النظري وبناء" على نتائج تجارب المحاكاة والبيانات الحقيقية تم التوصل الى مجموعة من الاستنتاجات والتوصيات الاتية :-

٨ . 1 الاستنتاجات:

١.١.٨ الجانب التجريبي:

من خلال نتائج تجارب المحاكاة تم التوصل الى اهم الاستنتاجات الخاصة لكل انموذج و كالاتي:

اولا" : انموذج المؤشر الواحد شبه المعلمي الأول:

١. كحالة عامة تم التوصل الى ان الطريقة المقترحة الثانية (LASSOA-MAVE) هي افضل طريقة في تقدير متجه المعلمات ودالة الربط واختيار المتغير في آن واحد لمعظم حجوم العينات وباختلاف قيم تباينات الاخطاء.

٢. يلاحظ ان هناك تذبذبا" في قيم (AMSE) لجميع الطرائق مع تزايد حجم العينة فبعضها يزداد بزيادة حجم العينة وبعضها يتناقص بزيادة حجم العينة بسبب سلوك الدالة المستعملة.

٣.اظهرت النتائج ان معظم قيم (AMSE) تتزايد بزيادة قيمة تباين الخطأ لكافة حجوم العينات لجميع طرائق التقدير .

ثانيا" : انموذج المؤشر الواحد شبه المعلمي الثاني:

١ . كحالة عامة اظهرت النتائج ان الطريقة المقترحة الاولى (SCAD-MAVE) هي افضل طريقة في تقدير واختيار المتغير في ان واحد لمعظم حالات الانموذج .

٢. يلاحظ وبشكل عام ان هناك تذبذبا" في قيم (AMSE) مع تزايد حجم العينة فبعضها يزداد بزيادة حجم العينة وبعضها يتناقص بزيادة حجم العينة لمختلف قيم تباينات الاخطاء.

٣. كما بينت النتائج ان هناك تزايدا" في قيم (AMSE) مع تزايد قيمة تباين الخطأ لكافة حجوم العينات لجميع طرائق التقدير شبه المعلميه .

٢.١.٨ الجانب العملي:

١.اظهرت نتائج البيانات الحقيقية ان الطريقة المقترحة الثانية (ALASSO-MAVE) هي افضل طريقة مقارنة بطرائق التقدير الاخرى لكونها تنتج اقل (MSE) في تقدير الجزء المعلمي (متجه المعلمات) والجزء اللامعلمي (دالة الربط المتمثلة بدالة عائد السهم في سوق العراق للاوراق المالية) واختيار المتغير لانموذج المؤشر الواحد شبه المعلمي وهذا يعني ان هذه الطريقة قادرة على تمثيل دالة عائد السهم .

٢. كما نلاحظ من خلال النتائج ان افضل معادلة تنبؤ لانموذج المؤشر الواحد شبه المعلمي تم الحصول عليها بواسطة استعمال الطريقة المقترحة الثانية ومن هذه المعادلة نستنتج وجود علاقة معنوية بين المتغير التوضيحي المتمثل بمعدل السعر السنوي للسهم وعائد السهم وهذا يعني ان اي زيادة في معدل السعر سيؤدي الى الزياد في عائد السهم بمقدار قيمة معامله .

٣.اعطت طرائق التقدير الشبه المعلميه الاتية (SCAD-NPLS) والطريقة المقترحة الاولى (MAVE-SCAD)) نتائج متقاربة لقيم متوسط مربعات الخطأ (MSE) .

2.8 التوصيـــــات:

في ضوء الجانب النظري وبناء" على ماتم التوصل اليه من استنتاجات ادناه اهم التوصيات

١.اوصي باستعمال الطريقة المقترحة الثانية (ALASSO-MAVE) في تحليل انموذج المؤشر الواحد شبه المعلمي كطريقة تقدير واختيار المتغير في آن واحد لما ابدته من كفاءة عالية مقارنة بطرائق التقدير شبه المعلميه الاخرى .

٢.يمكن دمج اي دالة جزاء اخرى غير الواردة في هذا البحث مع طريقة (MAVE) للحصول على طريقة مقترحة جديدة تعمل على تقدير واختيار المتغير لانموذج المؤشر الواحد شبه المعلمي مثل دالة جزاء اقل اعلى تحدب(concave penalty function minimax).

٣.يمكن استعمال طرائق اخرى لايجاد معلمة عرض الحزمة (hbandwidt) مثل طريقة قاعدة الابهام.

٤.اوصي باستعمال صيغ اخرى من الدوال اللبية (kernel) مثل (function Epanechnikov) في طرائق التقدير شبه المعلمية المستعملة في هذا البحث في تحليل انموذج المؤشر الواحد .

٥. بالامكان جعل هذه الدراسة اساس لدراسات مستقبلية وذلك من خلال توظيفها على نماذج شبه معلميه اخرى مثل انموذج متعدد المؤشرات(Multi-index model).

9.المصادر:

[١] يوسف ، خلود يوسف خمو . (٢٠٠٤) ، " مقارنة أساليب بيز مع طرائق أخرى لتقدير منحنى الأنحدار اللامعلمي " ، اطروحة دكتوراه فلسفة في الاحصاء ، كلية الأدارة والأقتصاد ، جامعة بغداد .

[2] Akkus ,O . (2011) , “ Xplore package for the popular parametric and semi-parametric single index models “ .Journel of science ,vol.24 , No.4, pp. 753-762 .

[3] Gasso , G., Rakotomemonjy , A., and Canu , S . (2008) , “ solving non- convex lasso-type problems with DC programing “ . IEEE xplor , pp. 450-455 .

[4] Huang ,J., and Xie , H .(2007) , “ Asymptotic oracle properties of SCAD –

penalized least squares estimators " . university of Iowa , vol.55 , pp. 149- 166 .

[5] Kopytou , E., and Santalova , D . (2007) , " Application of the single index model for forcasting of the inland conveyances ", Recent advances in stochastic modeling and data analysis . Singupre , world scientific poblishing co pte Ltd , pp. 268-276 .

[6] Li , R., and Liang , H .(2008) , " variable selection in semi-parametric regression modeling " . The Annals statistic , vol.36 , No, pp. 261-286 .

[7] Naik , P.A., and Tsai, C.L. (2001) , " single index model selections " . Biometrika 88 , pp. 821-832 .

[8] Peng , H., and Huang ,T.(2011) , " penalized least squares for single index models " . Journal of statistical planning and inference 141 , pp. 1362-1379 .

[9] Raheem , S.M.E.(2012) , " Absolute penalty and shrinkage stretegies in linear and partially linear models " , A thesis submitted to the faculty of Graduate studies through the department of mathematics and statistics in partial fulfillement of the requirements for the degree of doctdr of philosophy at the university of Windsor .

[10] Shi , J,. (2014) , " variable selection methods via penalized likelihood and comparison of their properties " . university of California , shi @stat . vcsb . edu .

[11] Simonoff , J.S., and Tsia , C.L. (2002) , " score tests for single index model " . Technometrics , vol.44, No.2 , pp. 142-151 .

[12] Su , L., and Zhang , Y . (2013) , " variable selection in non-parametric and semi-parametric regression model " . school of Economics , Singapore Management university .

[13] Tanaka , H . (2009) , " semi-parametric least squares Estimation of A single index model under monotonicity " . Department of Economics , university of Wisconsian , Madison , USA

[14] Thomas , J.F. (2006)," Simulation study for single index model" . submitted to the Department of Mathematical sciences of Clemson university , in partial fulfilliment for The requirements for The degree of Master of science in Mathematical sciences .

[15] Tibshirani , R . (1996) , " Regression shrinkage and selection via The lasso " . Journal of The Royal statistical society , series B , 58 , PP. 267-288 .

[16] Weng , Y ., Zhao, Y., Tang , G., and Liu , Z .(2013) , " prediction of The mechanical properties of Hot-rolled C-Mn steels by single index model " . computer science , Education (ICCSE) , IEEE , PP. 275-280 .

"compared some of penalized methods in analysis the semi-parametric single index model with practical application"

ABSTRACT

In this research been to use some of the semi-parametric methods the based on the different function penalty as well as the methods proposed by the researcher because these methods work to estimate and variable selection of significant at once for single index model including (SCAD-NPLS method , the first proposal SCAD-MAVE method , the second proposal ALASSO-MAVE method) .As it has been using a method simulation time to compare between the semi-parametric estimation method studied , and various simulation experiments to identify the best method based on the comparison criteria (mean squares error(MSE) and average mean squares error (AMSE)).

And the use of real data again to verify the performance of semi-parametric methods indeed the practical , was reached the best method for

estimate and variable selection of semi parametric single index model is the second method proposed (ALASSO-MAVE) for each of the simulation experiments of the first semi -parametric single index model and real data .

Keywords: single index model , variable selection, SCAD method ,SCAD-MAVE method , ALASSO-MAVE method