

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 أن واحد منها (طريقة-SCAD-NPLS), الطريقة المقترحة الاولى (SCAD-MAVE) والطريقة المقترحة الثانية (ALASSO-MAVE) (إذ تم أستعمال إجراء 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 المعلمي 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 سكاو , طريقة سكاو - ماف , طريقة لاسو التكيفية - ماف .



1. المقدمة Introduction

يعد تحليل الأنحدار من أكثر الأساليب الأحصائية التي تقوم ببناء أنموذج أحصائي و أكثر استعمالاً لفهم العلاقة بين المتغير المعتمد (الاستجابة) Y و مجموعة المتغيرات التوضيحية X . ففي نماذج الأنحدار الخطي المتعدد تفترض أن صيغة دالة التوقع الشرطي الآتية :-

لتحليل العلاقة على فرض أن دالة التوقع الشرطي هي معلومة لعدد من المعلمات ولكن التقديرات الناتجة يمكن أن تكون مضللة للغاية إذا كان افتراض الأنموذج المعلمي غير صحيح , وبما أن هذا الافتراض لا يتحقق في أغلب التطبيقات العملية لانه لا يأخذ بنظر الاعتبار التأثير اللاخطي للمتغيرات التوضيحية لذا كان من المناسب إيجاد أسلوب جديد يأخذ بنظر الاعتبار هذا التأثير و هو الأنحدار اللامعلمي (Nonparametric Regression) , وكما هو معروف لدى أغلب الباحثون أن الأنحدار اللامعلمي يعاني من مشكلة تعدد الأبعاد (The curse of dimensionality) التي تحدث عند زيادة عدد المتغيرات التوضيحية. [2]

ومع ذلك فإن الطرائق اللامعلمية 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 المعلمية كحل وسط بين تلك النماذج المعلمية المقيدة والنماذج واللامعلمية المرنة للغاية , إذ تجعل الافتراضات حول دالة التوقع الشرطي أقوى من تلك في التقدير اللامعلمي ولكن أقل تعقيداً من الافتراضات لأنموذج المعلمي وهذا يتيح للطرائق شبه المعلمية اختزال البعد الفعال لمشكلة التقدير ومن ثم زيادة دقة التقدير لتلك المتوفرة في التقدير اللامعلمي , لذا تم استعمال أنموذج المؤشر الواحد single index model وهو أحد النماذج شبه المعلمية الشائعة إذ يعد امتداد طبيعي للأنحدار الخطي العام والذي يتعامل مع العلاقات اللاخطية ويعرف من خلال الصيغة $E[Y | X = x] = g(X^T \beta)$ وهو أكثر مرونة من النماذج المعلمية بسبب أنه يسمح للعلاقات غير الخطية بين مؤشر المتغير $X^T \beta$ ومتغير الاستجابة ويحتفظ بالخصائص الجيدة الى جانب قدرته على اختزال الحد من مخاطر خطأ تحديد دالة الربط ويساعد في التغلب على مشكلة تعدد الأبعاد بسبب أن المؤشر $X^T \beta$ هو مجاميع للأبعاد المرتفعة الى X . [10].

2. مشكلة البحث The research problem

إن دراسة عدد كبير من المتغيرات التوضيحية وحجم عينة كبير عملية معقدة إذ بزيادتها تزيد درجة التعقيد للأنموذج مما يدفع الباحثين الى استعمال مسألة اختيار المتغير وهي واحدة من أكثر المسائل انتشاراً عندما يراد إيجاد أنموذجاً واحداً يمثل العلاقة بين متغير الاستجابة والمجموعة الجزئية للمتغيرات التوضيحية الممكنة بسبب أن بعض المتغيرات التوضيحية تكون غير أساسية في تأثيرها على المتغير المعتمد أو يكون تأثيرها مماثل لتأثيرات متغيرات أخرى وأن العديد من هذه المتغيرات يكون لها ارتباطاً داخلياً , كما أن الافتراض الخطي لا يتحقق في معظم التطبيقات العملية فضلاً عن ظهور مشكلة تعدد الأبعاد (The curse of dimensionality) التي تحدث عند زيادة عدد المتغيرات التوضيحية .

3. هدف البحث The aim of the research

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

4. الجانب النظري

1.4: 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 $g(\cdot)$ والتي تمثل الجزء اللامعلمي تدعى بالنماذج شبه المعلمية (Semi parametric Models), إذ يعد هذا الأنموذج من أكثر النماذج شبه المعلمية شيوعاً ويستعمل 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 تركيبية خطية واحدة لـ P من المتغيرات التوضيحية والتي تستطيع الحصول على معظم المعلومات حول العلاقة بين المتغير المعتمد (الاستجابة) Y و المتغيرات التوضيحية X و من ثم تجنب مشكلة تعدد الأبعاد (Curse of dimensionality). [16]. إذ عمل الباحثون على توسيع و تعميم لأنموذج الانحدار الخطي العام (GLM) الآتي: [15]

$$Y_i = X_i^T \beta + \epsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

$$\text{With } E(\epsilon_i | x_i) = 0$$

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 :

$$\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$$

$$x_i^T = (1, x_{i1}, \dots, x_{ip})$$

β : تمثل متجه المعلمات المجهولة
 x_i^T : يمثل مصفوفة المتغيرات التوضيحية
و من درجة $n \times p$

و تعميمه 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) إلى الصيغة الآتية :-

$$Y_i = g(X_i^T \beta) + \epsilon_i, \quad i = 1, 2, \dots, n \quad (2)$$

إذ تشير دالة الانحدار أحادية المؤشر إلى :

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

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

بافتراض إن $m(x_i)$ هي دالة مجهولة و تمثل دالة التمهيد.
اذ أن :

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

$y_i \in \mathbb{R}$: يمثل متجه المتغير المعتمد لـ i th من المشاهدات من درجة $n \times 1$.

$x_i \in \mathbb{R}$: يمثل مصفوفة المتغيرات التوضيحية لـ i th من المشاهدات من درجة $n \times p$.

β : يمثل متجه المعلمات المجهولة من درجة $p \times 1$ و يجب تقديرها بحيث تحقق شروط تشخيص الأنموذج و

هي : $\beta^T \beta = 1$ or $\beta = 1$ و أول مركبة تكون موجبة ($\beta_1 > 0$)

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

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

$\epsilon_i \in \mathbb{R}$: تمثل الخطأ العشوائي ذات توزيع طبيعي مستقل و متماثل (iid) و له متوسط صفر و تباين محدد

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

T : يمثل المبدلة (Transpose).

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

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

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} g(x_{i1}^T \beta) \\ g(x_{i2}^T \beta) \\ \vdots \\ g(x_{in}^T \beta) \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in} \end{bmatrix}$$

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 فإن النموذج يدعى بأنموذج المؤشر الواحد شبه المعلمي semiparametric single index model (SSIM) والمبين في الصيغة (2) والهدف هو تقدير متجه المعلمات β المجهولة ودالة الربط $g(\cdot)$ المجهولة وأجراء التقدير للنموذج 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 $g(\cdot)$ المجهولة والحصول عليها من خلال استعمال الأنحدار اللامعلمي Y على $x_i^T \hat{\beta}$. [2]

اذ يمكن التركيز على طريقة المربعات الصغرى الاعتيادية (OLS) لتقدير متجه المعلمات β مع نفس نسبة التقارب لتلك المتحققة في الأنموذج المعلمي و من ثم فإن أنموذج المؤشر الواحد (SIM) هو أنموذج معلمي دقيق لتقدير β ذات بعد واحد, مع ملاحظه أن الحد الثابت (intercept) β_0 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_0 لعدم تأثيرها على أي متغير.

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

$$\hat{Y}_i = \hat{g}(x_i^T \hat{\beta}) \text{ ----- (3)}$$

2.4 : تقدير دالة الربط $g(\cdot)$ لأنموذج المؤشر الواحد

Estimation of link function $g(\cdot)$ for single index model

بين الباحث Horowitz عام (1998) 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 متنسق لـ $g(\cdot)$, وحل المشاكل المثلى غير الخطية للحصول على تقدير متنسق لـ $\hat{\beta}$.

وأشار الباحث Horowitz الى أن الطرائق التكرارية هي مكثفة حسابياً لأنها تتطلب تقدير متوسط الأنداد اللامعلمي في كل نقطة للبيانات لحساب دالة الهدف (Objective function) والتي قد تكون غير محدبة أو متعددة الحدود التي تؤدي الى تكرار أمثل لـ $\hat{\beta}$. [7]. فعند تقدير متجه المعلمات β فإن الخطوة التالية هي تقدير الدالة الربط $g(\cdot)$ أحادي المتغير ولتحقيق هذا الهدف يمكن أستعمال بعض مقدرات التمهيد اللبي (Kernel smoothing estimator) منها الأنداد الخطي الموضوعي (Local linear regression) والمهد اللبي ناداريا – واتسون (Nadaraya – Watson kernel smoothing). على سبيل المثال يمكن تقدير دالة التوقع الشرطي (دالة الربط) $g(\cdot)$ بأستعمال مقدر ناداريا – واتسون (NW) من خلال الصيغة الآتية :-

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

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

اذ أن :

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

. يمثل قيمة المتغير المعتمد لـ i th من المشاهدات . y_i

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

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

وأن $k_h(\cdot)$: تمثل دالة كاوس (Gaussian function) المستعملة في هذا البحث و بالشكل الآتي :

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

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][14]

3.4: طرائق التقدير شبه المعلمية Semiparametric estimation methods

ظهرت العديد من أساليب وتقنيات تقدير و اختيار المتغيرات في النماذج شبه المعلمية 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 :-

(1.3.4) طريقة المربعات الصغرى الجزائية غير المقعرة مع دالة جزاء – سجاد

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

بين الباحث Fan و آخرون في عام (2000) أن طريقة Lasso – تنتج مقدرات منحيزة لمعاملات الأنداد الكبيرة و قد دفعت هذه المشكلة الباحثان Fan & Li في عام (2001) إلى اقتراح 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 اختيار المجموعة الجزئية و أنداد الحرف لكن كذلك تنتج حلول مبعثرة (Sparsity) يضمن الأستمرارية للنماذج المختارة (أستقرار أختيار الأنموذج) و تعطي تقديرات غير منحيزة للمعاملات الكبيرة. [3]

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

و تم تطبيق طريقة – (SCAD) من قبل الباحثين Fan & Li عام (2001) على الأنموذج الخطي العام (GLM) و الحصول على تقديرات هذه الطريقة من خلال الصيغة الآتية :

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

اذ أن :

(: تمثل دالة جزاء SCAD – و تعرف بالشكل الآتي :

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

وأن :

a : تمثل معلمة ضبط ثابتة وقيمتها اكبر من (2) واقترح الباحثان Fan & Li استعمال قيمة a = 3.7 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 λ , a λ] .

والحلول لدالة جزاء – (SCAD) يمكن أن تحدد بالشكل الآتي :- [9][13]

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

وقد بينوا أن النتائج من المعاملات الصغيرة لهذه الطريقة يتم مساواتها بالصفر و عدد قليل من المعاملات الأخرى يجري تقليصها نحو الصفر مع الاحتفاظ 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 الاختيار السليم لمعلمة الضبط (Tuning parameter) أو معلمة الجزاء (Penalty parameter) (λ) والتي عادة ما يتم تحديدها بواسطة طريقة العبور الشرعي العام (GCV) Generalized cross – validation إذ تبين للباحثين Fan & Li عام (2001) أن اختيار معلمة الضبط المثالية بواسطة طريقة (GCV) تفشل في اختيار المتغير ولذلك تم اقتراح طريقة – (BIC) في عام (2011) لاختيار 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 الصيغة الآتية :

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

اذ أن : (a) تمثل معلمة ضبط ثابتة و قيمتها a = 3.7 و هي مقترحة من قبل الباحثان Fan & Li عام (2001) . [6][8]

إن الفكرة الأساسية لهذه الطريقة تتمثل باستبدال تقديرات المربعات الصغرى الاعتيادية (OLS) من خلال تقدير المربعات الصغرى الجزائية غير المقعرة (NPLS) . إذ تم اقتراح خوارزمية تكرارية لتقدير المعلمات و دالة الربط لأنموذج المؤشر الواحد و بين الباحثان أن هذه الطريقة تستطيع تقدير المعلمات و اختيار المتغيرات في آن واحد .

ولتقدير أنموذج المؤشر الواحد شبه المعلمي (SSIM) بطريقة (SCAD – NPLS) تم اتباع نفس فكرة الباحث Carroll وآخرون عام (1997) إذ أستعملوا خوارزمية تكرارية لتقدير متجه المعلمات β ودالة الربط $g(\cdot)$ في آن واحد .

ولأعطاء تقدير لـ β ودالة الربط $g(\cdot)$ يمكن التقريب موضعياً من خلال الدالة الخطية الآتية :-

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

اذ أن :

v : هو مجاور إلى u

و أن :

$$a = g(u)$$

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

هي ثوابت موضعية .

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

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

اذ أن :

$K(\cdot)$: تمثل دالة kernel المتماثلة .

h : يمثل معلمة عرض الحزمة Bandwidth

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

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

اذ أن :

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

ومع ذلك فإن دالة الربط التقديرية $g(\cdot)$ قد تكون ليست دالة خطية لحل المعادلة و هي مسألة غير خطية مثلى. و لحل المسألة غير الخطية المثلى المذكورة آنفاً , أقرح الباحثان Peng & Huang في عام (2011) فكرة التقريب الموضعي و تحديث التقدير لـ β من خلال تقليل دالة المربعات الصغرى الجزائية غير المقعرة والتي تعطي تقديرات دقيقة لـ $\beta^{(0)}$ ودالة الربط $\hat{g}(\cdot)$ بموجب المعادلة الآتية:

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

ويكون إجراء التقدير لـ β و $g(\cdot)$ لطريقة (SCAD – NPLS) لأنموذج المؤشر الواحد بموجب الخوارزمية الآتية :

الخطوة (0) : وهي خطوة التهيئة للحصول على تقدير أولى (ابتدائي) لـ β وليكن $\hat{\beta}_1$ بأستعمال طريقة المربعات الصغرى الأعتيادية (OLS) ونفرض أن :

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

اذ أن :

$\text{sign}(\hat{\beta}_{11})$: تمثل إشارة العنصر الأول إلى $\hat{\beta}_1$ وأن $\hat{\beta}_{11} > 0$ ويكون $\|\hat{\beta}\| = 1$ لغرض التشخيص.

الخطوة (1) : نحدد $\hat{\beta}$ و يتم الحصول على $\hat{a} = \hat{g}(u, \hat{\beta})$, $\hat{b} = \hat{g}(u, \hat{\beta})$ بأستعمال الأنحدار الخطي الموضعي من خلال حل المعادلة الآتية :

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

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 (2) : تحديث التقدير لـ β من خلال تقليل دالة المربعات الصغرى الجزائية غير المقعرة $\beta^{(0)}$ وتحل محلها $\hat{\beta}$ أي يتم افتراض $(\beta^{(0)} = \hat{\beta})$ وتحديد \hat{a} , \hat{b} للحصول على $\hat{\beta}$.
$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \{ \sum_{i=1}^n [y_i - \hat{g}(x_i^T \beta^{(0)}) - \hat{g}(x_i^T \beta^{(0)})(x_i^T \beta - x_i^T \beta^{(0)})]^2 + n \sum_{j=1}^p P_{\lambda}(|\beta_j|) \}$$

الخطوة (3) : يتم الأستمرار بالخطوتين (1) و (2) حتى التقارب .
الخطوة (4) : يعطي تقدير نهائي لـ $\hat{\beta}^{SCAD}$ من خلال الخطوة (3) .

ونكرر تقدير $\hat{g}(u, \hat{\beta})$ لـ $\hat{g}(u)$ من خلال حل المعادلة في الخطوة (1) وبالشكل الآتي :
$$(\hat{a}, \hat{b}) = \underset{(a,b)}{\operatorname{argmin}} \sum_{i=1}^n [y_i - a - b(x_i^T \hat{\beta}^{SCAD} - u)]^2 \cdot k_h(x_i^T \hat{\beta}^{SCAD} - u)$$

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

2.3.4 الطريقة المقترحة الاولى: طريقة تقدير التباين لأدنى معدل مع دالة الجزاء – سكا د .

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 (2002)) مع دالة جزاء – SCAD (Smoothly clipped absolute deviation) للباحثين (Fan & Li عام (2001)) . وذلك بأضافة دالة جزاء – SCAD الى دالة الخسارة لطريقة (MAVE) المتمثلة بالمعادلة الآتية :-

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

لنتنتج تقدير التبعر (A sparse estimate) و نتيجة لذلك فإن العديد من المعاملات هي أصفار و لتحديد المتغيرات مع المعاملات غير الصفريّة تلقائياً" و إلى تقدير نموذج المؤشر الواحد (SIM) أقترح الباحثان (Fan & Li) أستعمال دالة جزاء (SCAD) و يمكننا أيضاً تنفيذ أختيار المتغير و تقدير نموذج المؤشر الواحد (SIM) في أن واحد من خلال فرض جزاء (SCAD) .

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

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

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

اذ أن :
() $p_{\lambda}^{SCAD}(|\beta_k|)$: تمثل دالة جزاء – SCAD التي تقلل الأتحدار الآتي :

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

ولحساب مسألة التقليل في المعادلة (2-37) يمكن أن تحلل إلى تقليل مسألتين وفق خوارزمية (SCAD – MAVE) لأنموذج المؤشر الواحد شبه المعلمي (SSIM) .

الخطوة (0): نفرض تقدير أولى (أبتدائي) للمعلمة $\hat{\beta}^{(0)}$ بأستعمال طريقة المربعات الصغرى الأعتيادية (OLS).

الخطوة (1): يتم تثبيت $\hat{\beta}^{(0)} = \hat{\beta}$ و نحسب قيمة الحل الى (\hat{a}_j, \hat{b}_j) من خلال الصيغة الآتية :-

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

الخطوة (2): يتم تثبيت (\hat{a}_j, \hat{b}_j) و تقدير متجه المعلمات β وفق الصيغة الآتية :-

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

$$\beta: \|\beta\| = 1$$

لتبسيط الصيغه نفترض أن:

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

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

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

تصبح المسألة التي نقل بالشكل الآتي :-

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

الخطوة (3): يتم تكرار الخطوتين (1) و (2) مع $\frac{\hat{\beta}}{\|\hat{\beta}\|}$ 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 $g(\cdot)$ النهائي يكون: $\hat{a}_j = \hat{g}(u, \hat{\beta}^{SCAD-MAVE})$ (3.3.4) الطريقة المقترحة الثانية:

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 (1996) أن أسلوب (Lasso) يفتقر إلى خصائص أوراكن (Oracle properties) كما توصل الباحثان Fan & Li عام (2001) أن هذا الأسلوب لديه تحيز في تقدير المعاملات غير الصفريّة الكبيرة و أظهروا أيضاً انه لا يمتلك خصائص أوراكن مما دفع الباحث Zou عام (2006) إلى اقتراح 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 (2006) أن طريقة لاسو التكيفية (ALasso) هي طريقة وزن جزائية لدالة جزاء- L_1 لتقدير و أختيار النموذج في آن واحد و لها خصائص أوراكن المتمثلة بالطبيعي المحاذي و نسبة التقارب المثلى و الاتساق في أختيار المتغير .

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

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

وبناءً على ما تم ذكره آنفاً فإن مقدر لاسو التكميلية (Alasso) لـ $\hat{\beta}$ و لأنموذج المؤشر الواحد شبه المعلمي (SSIM) يمكن الحصول عليه من خلال الصيغة الآتية :-

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

اذن:

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

$$\text{BIC}(\lambda) = \log(\hat{\sigma}) + \text{df}(\lambda) \frac{\log(n)}{n} \text{----- (23)}$$

اذ أن :

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

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

d : تمثل البعد لـ $\hat{\beta}$

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

$\text{df}(\lambda)$: تمثل درجة حرية الأنموذج و تحدد من خلال عدد المعالم المقدر غير الصفري في $\hat{\beta}$ كما يرى الباحثون Tibshirani , Zou & Hastie عام (2007) .

n : يمثل حجم العينة

λ : تمثل معلمة الجزاء أو معلمة الضبط .

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

$w = (w_1, w_2, \dots, w_p)$: تمثل الأوزان التكميلية (Adaptive weight) و بين الباحث Zou أنه إذا كان اختيار الأوزان (w_k) بكفاءة وبطريقة تعتمد على البيانات فإن طريقة لاسو التكميلية (Alasso) يمكن أن نحقق خصائص أوراكل بحيث ينفذ كما لو كان الأنموذج الصحيح معروف و قد أقتراح استعمال الأوزان المقدره وبالشكل الآتي :-

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

من خلال استعمال تقديرات المربعات الصغرى الاعتيادية (OLS) لأختيار \widehat{W}_k اذ أن $\hat{\beta}_k$ تمثل تقديرات (OLS) و هي مقدر ابتدائي متسق لـ \sqrt{n} (يحتوي على \sqrt{n} نسبة التقارب) .

و أن γ : يمثل معلمة الأنكماش و قيمتها أكبر من الصفر و نفترض أن تكون قيمتها تساوي واحد ($\gamma = 1$) . [9]

وبناءً على فكرة الباحثان (Zho & Zang He) في عام (2011) اللذان أستعملا طريقة النوع - Lasso و طريقة - 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 - MAVE مع دالة جزاء - Adaptive LASSO وتوظيفها لتقدير و أختبار المتغير لأنموذج المؤشر الواحد في المعادلة الآتية :- [8]

$$\sum_{i=1}^n [y_i - a - b(X_i^T \beta - u)]^2 K_h(X_i^T \beta - u) \text{----- (25)}$$

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

$$\min = \sum_{j=1}^n \sum_{i=1}^n [y_i - a_j - b_j \beta^T (x_i - X_j)]^2 w_{ij} + \lambda \sum_{k=1}^p \hat{w}_k |\beta_k| \quad -- \quad (26)$$

$a, b, \beta, \|\beta\|=1$

نلاحظ أن الجزء الأول للمعادلة (26) هو دالة الخسارة إلى طريقة – (MAVE) لتقدير قيمة المعلمات β ولها جمع داخلي هو :-

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

والجزء الثاني للمعادلة (26) يمثل دالة جزاء لاسوالتكيفية (Adaptive lasso penalty function) وهذا

الجزء يجعل $\hat{\beta}$ متناثرة (Sparsity) و من ثم يؤدي إلى اختيار المتغير (Variable selection).
و يمكن تلخيص الخوارزمية المقترحة لطريقة (ALASSO – MAVE) للنموذج المؤشر الواحد شبه المعلمي (SSIM) و التي تعمل على تقدير و اختيار المتغير في آن واحد بالخطوات الآتية :-

الخطوة (0) : نحصل على تقدير أولي (أبتدائي) لـ β وليكن $\hat{\beta}^{(0)}$ بطريقة المربعات الصغرى الاعتيادية (OLS) و نفرض أن :

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

اذ أن :

$\text{sign}(\hat{\beta}_1^{(0)})$: يمثل إشارة العنصر الأول لـ $\hat{\beta}^{(0)}$ و قيمته أكبر من الصفر و أن $\|\hat{\beta}\| = 1$ لغرض التشخيص.

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

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

إذ أن :

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

$k(\cdot)$: تمثل دالة kernel المتماثلة و تم استعمال دالة (Gaussian).
 h : تمثل معلمة عرض الحزمة (Bandwidth) و نختارها لتكون الأمثل و تحسب من خلال طريقة العبور الشرعي (CV).

الخطوة (2) : نحدد $\{\hat{a}_j, \hat{b}_j\}_{j=1}^n$ للحصول على $\hat{\beta}^{ALASSO-MAVE}$ من خلال حل المعادلة الآتية :-

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

ويمكن تبسيطها بالشكل الآتي :-

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

اذ ان :

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

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

الخطوة (3): يتم الاستمرار بتكرار الخطوتين (1) و (2) حتى التقارب و يكون التقدير النهائي لـ $\hat{\beta}^{ALASSO-MAVE}$.

و التقدير النهائي لـ $g(\cdot)$ هو $(\hat{a}, \hat{b})^{ALASSO-MAVE}$ ويتم الحصول عليه من خلال حل المعادلة الآتية :-

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

5. معايير المقارنة:- [1][12]

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

1.5: معيار (AMSE)

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

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

اذ ان

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

n : يمثل عدد المشاهدات .

و يكون الأنموذج الأفضل الذي يعطي أقل قيمة لـ (AMSE) .

6. الجانب التجريبي:

تم في هذا المبحث استعمال الاسلوب التجريبي (المحاكاة) في مقارنة طرائق التقدير شبه المعلمية لأنموذج المؤشر الواحد لبيان افضل الطرائق المستعملة والتي تمثل البيانات تمثيلاً " سليماً" .
لقد تضمنت تجارب المحاكاة الحالات الافتراضية المبينة في الجدول رقم (1) لغرض تقدير وأختيار المتغير لأنموذج المؤشر الواحد شبه المعلمي (SSIM) من خلال تطبيق طرائق التقدير شبه المعلمية المستعملة في هذا البحث التي يتم من خلالها تحقيق هدف البحث . وتم الاعتماد على برنامج (R-package) للحصول على النتائج .

جدول رقم (1)

يمثل القيم الافتراضية لتجارب المحاكاة

Experiment	P	$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.1$
		n	n	n
I	7	50 100 200	50 100 200	50 100 200

اذ أن كل تجربة من التجارب الثلاث بحثت عند ثلاثة مستويات للتباين $(\sigma = 1, 0.5, 0.1)$ وبحجوم عينات

مختلفة وهي $(n = 50, 100, 200)$ وعدد متغيرات توضيحية $(p = 7)$ وكررت كل تجربة (400) مرة. ولكل دالة من الدوال المدروسة الآتية :

$$g(X^T \beta) = \sin\left(\frac{\pi X^T \beta}{10}\right) \quad [7] \text{ الدالة الاولى}$$

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

مع القيم الافتراضية لمتجه المعلمات الآتية

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

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 واختيار المتغير للنموذج قيد البحث ولمختلف دوال الأتحاد المستعملة. ومن الجدير بالذكر انه تم استعمال دالة Gaussian kernel لأختيار الدالة اللبية kernel وبالاتماد على برنامج R وتم الحصول على النتائج لكل دالة من الدوال المستعملة بالاتماد على برنامج R والموضحة من جدول رقم (2) الى جدول رقم (4) والتي سيتم تحليلها لاحقاً.

جدول رقم (2) يمثل مقارنة طرائق التقدير شبه المعلمية لانموذج المؤشر الواحد باستعمال معيار المقارنة

$$g(X^T \beta) = \sin(\pi X^T \beta / 10) \text{ (AMSE)} \text{ . لحالة الدالة الاولى}$$

methods	n	P=7		
		$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.1$
SCAD-NPLS	50	0.76670	0.22834	0.07373
	100	1.13806	0.40179	0.08219
	200	1.25896	0.42519	0.25106
SCAD-MAVE Suggest-1	50	0.76768	0.22824	0.07393
	100	1.13805	0.40099	0.08182
	200	1.25895	0.42521	0.25109
ALASSO-MAVE Suggest-2	50	0.77208	0.23720	0.09566
	100	1.11500	0.34989	0.09097
	200	1.17569	0.33291	0.20176

جدول رقم (3) يمثل مقارنة طرائق التقدير شبه المعلمية لانموذج المؤشر الواحد باستعمال معيار المقارنة

$$(AMSE)$$

methods	n	P=7		
		$\sigma = 1$	$\sigma = 0.5$	$\sigma = 0.1$
SCAD-NPLS	50	1.43765	0.74090	0.72007
	100	1.54014	0.9262869	0.509286
	200	1.45103	0.74114	0.62803
SCAD-MAVE Suggest-1	50	1.43584	0.74091	0.72006
	100	1.53888	0.9262861	0.509280
	200	1.45104	0.73701	0.62802
ALASSO-MAVE Suggest-2	50	1.49170	0.75066	0.61895
	100	1.54216	0.925069	0.61896
	200	1.39805	0.753067	0.63087

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

جدول رقم (4) يوضح افضلية طرائق التقدير شبه المعلميه لامتودج 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-1 (SCAD-MAVE)
	200	Suggest-2 (ALASSO-MAVE)	Suggest-2 (ALASSO-MAVE)	Suggest-2 (ALASSO-MAVE)
الدالة الثانية	50	Suggest-1 (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 :

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

من خلال الجدولين رقم (2) و(4) نبين الاتي :-

1. في حالة عدد المتغيرات التوضيحية (P=7) عند حجم العينة الصغيرة (50) اظهرت النتائج ان طريقة (SCAD-NPLS) ولقيم تباينات الاخطاء(0.1,1) هي افضل طريقة بينما حققت الطريقة المقترحة الاولى الافضلية عند تباين الخطأ (0.5) .
2. حققت الطريقة المقترحة الثانية الافضلية لحجوم العينات المتوسطة والكبيرة (100,200) ولكافة قيم تباينات الاخطاء عدا حالة حجم العينة المتوسطة(100) عند قيمة تباين الخطأ(0.1) فكانت الافضلية للطريقة المقترحة الاولى .

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

من خلال الجدولين رقم (3) و(4) نبين الاتي :-

1. في حالة عدد المتغيرات التوضيحية(P=7) اظهرت النتائج ان طريقة (SCAD-NPLS) هي افضل طريقة في حالة العينة الصغيرة (50) ولقيمة تباين الخطأ (0.5) وحققت الطريقة المقترحة الاولى الافضلية عند قيمة تباين الخطأ(1) بينما حققت الطريقة المقترحة الثانية الافضلية عند قيمة تباين الخطأ (0.1) .
2. كما اظهرت النتائج ان الطريقة المقترحة الاولى هي الافضل عند حجم العينة المتوسطة (100) لقيم تباينات الاخطاء (0.1,1) في حين حققت الطريقة المقترحة الثانية افضلية عند قيمة تباين الخطأ (0.5) .
3. اظهرت النتائج ان الطريقة المقترحة الاولى هي افضل طريقة عند حجم العينة الكبيرة (200) لقيم تباينات الاخطاء (0.1,0.5) بينما حققت الطريقة المقترحة الثانية عند قيمة تباين (1) .

7. الجانب العملي :

في هذا الإجراء 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 المعلمية لأنموذج المؤشر الواحد الذي يمثل الدراسة التطبيقية فمن خلال معرفة العوامل المؤثرة ومراقبتها سيساعد المستثمر في اتخاذ قرار البيع أو قرار الشراء . ويركز هذا الجانب على حساب المركبة المعلمية (الجزء المعلمي) والمركبة اللامعلمية (الجزء اللامعلمي) لأنموذج المؤشر الواحد شبه المعلمي (SSIM) اذ تتضمن المركبة المعلمية فقط المعلمات المعنوية غير الصفيرية لمتجه المعلمات β ذات بعد $(P \times 1)$ والذي يعكس تأثير المتغيرات التوضيحية على التغير في المتغير المعتمد أما المركبة اللامعلمية فتمثله بالجزء $g(.)$ وتمثل متجه ذات رتبة $(n \times 1)$ ومن اجل اظهار دور تأثير المتغيرات التوضيحية على المتغير المعتمد فقد تم أخذ بيانات سنوية لعائد السهم السوقي في العراق ويتكون مجتمع الدراسة من جميع الشركات التجارية العراقية المدرجة في سوق العراق للاوراق المالية خلال الفترة من 2008م وحتى 2011م والتي تتكون من (73) شركة نظامية اذ بياناتها تطرح من خلال هذا السوق بانتظام خلال سنوات الدراسة وموزعة على (8) قطاعات مختلفة أما حجم عينة الدراسة النهائي فتتكون من (60) شركة اذ تم اختيارها 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 (5)

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

الرقم	أسم القطاع	عدد الشركات	عينة الدراسة
1	المصارف	21	21
2	التأمين	4	4
3	الاستثمار	2	2
4	الخدمات	9	7
5	الصناعة	21	15
6	الفنادق والسياحة	10	6
7	الزراعة	5	5
8	الاتصالات	1	-
	-	73	60
	الجموع		

والمتغيرات المستعملة في الجانب التطبيقي يمكن توضيحها بالشكل الآتي :

بالنسبة للمتغير المعتمد Y والذي يمثل عائد السهم السوقي ويحسب من خلال (صافي ربح السنة مقسوماً على رأس المال المدفوع والذي يمثل رأس المال المستلم فعلاً من قبل الشركة).
أما بالنسبة للمتغيرات التوضيحية فالمتغير $X1$ يمثل نسبة دوران السهم ويحسب من خلال (عدد الاسهم المتداولة لعام كامل مقسوماً على عدد الاسهم الصادرة (رأس المال الاسمي) $100X$). اذ ان رأس المال الاسمي هو رأس المال الذي يحق للشركة إصداره ومسجل في عقد تأسيس الشركة. والمتغير $X2$ يمثل نسبة حقوق الملكية وتحسب من خلال (أجمالي حقوق المساهمين مقسوماً على مجموع الموجودات $x 100$).
والمتغير $X3$ يمثل مكرر الأرباح ويحسب من خلال (سعر الأغلاق للسهم مقسوماً على عائد السهم).

والمتمغير X4 يمثل نسبة التداول وتحسب من خلال (الموجودات المتداولة مقسوماً على مصادر التمويل قصيرة الأجل) إذ أن الموجودات المتداولة تتكون من النقدية أو المخزون (بالكلفة) بالإضافة إلى المدينون، أما بالنسبة لمصادر التمويل قصيرة الأجل فهي تشمل الحسابات الجارية والودائع والتخصيصات والدائنون.

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

1.7 جمع البيانات:

تم جمع البيانات من القوائم المالية للشركات التجارية العراقية وبالإعتماد على دليل الشركات الصادر من سوق العراق للأوراق المالية للفترة من العام 2008م وحتى العام 2011م والخاصة بعائد السهم السوقي Y والعوامل المؤثرة عليه لكافة الشركات.

إن الخطوة الأولى قبل تقدير النموذج هو فحص العلاقة بين المتغيرات التوضيحية (المستقلة) والمتغير المعتمد بيانياً للتأكد من كون العلاقة خطية أو غير خطية إذ تم الإعتماد على برنامج SPSS لاختبار البيانات من خلال رسم الانتشار لها وتم التوصل إلى أن جميع العلاقات بين المتغيرات التوضيحية والمتغير المعتمد هي غير خطية.

2.7 مناقشة نتائج البيانات الحقيقية:

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

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

ومن خلال البيانات الحقيقية واعتماد الانموذج اعلاه يمكن الحصول على تقديرات متجه المعلمات ودالة الربط (دالة عائد السهم) لمختلف طرائق التقدير شبه المعلمية وكما مبينة في الجدولين رقم (6) و(7) ادناه:

جدول رقم (6) يمثل القيم التقديرية لمتجه المعلمات β (الجزء المعلمي) لانموذج المؤشر الواحد شبه المعلمي لجميع الطرائق للبيانات الحقيقية

parameters	methods		
	SCAD-NPLS	SCAD-MAVE Suggest-1	ALASSO-MAVE Suggest-2
	Beta	Beta	Beta
$\hat{\beta}_1$	0.006969	0.006985	0.000000
$\hat{\beta}_2$	-0.380558	-0.380563	0.000000
$\hat{\beta}_3$	0.245487	0.245583	0.000000
$\hat{\beta}_4$	0.548585	0.548587	0.000000
$\hat{\beta}_5$	0.545814	0.545839	0.000000
$\hat{\beta}_6$	0.000000	0.000000	0.000000
$\hat{\beta}_7$	0.442722	0.442644	1.000000



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

جدول رقم (7)

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

methods					
SCAD-NPLS		SCAD-MAVE Suggest-1		ALASSO-MAVE Suggest-2	
$\hat{g}(xb)$	$\hat{g}(xb)$	$(xb)\hat{g}$	$\hat{g}(xb)$	$(xb)\hat{g}$	$\hat{g}(xb)$
0.221916	0.181623	0.221955	0.181573	0.099556	0.155052
0.172070	0.196333	0.172113	0.196373	0.134763	0.318154
0.255377	0.261657	0.255436	0.261666	0.095228	0.875822
0.173751	0.186507	0.173786	0.186540	0.113327	0.391879
0.213648	0.220613	0.213691	0.220677	0.096817	0.121264
0.176170	0.216059	0.176225	0.216003	0.095228	0.164153
0.177760	0.213468	0.177797	0.213580	0.166123	0.095028
0.240393	0.181498	0.240510	0.181449	0.189409	0.247892
0.183973	0.144050	0.183983	0.144007	0.095111	0.097525
0.207949	0.072137	0.208001	0.072186	0.095922	0.097861
0.199288	0.177803	0.199336	0.177732	0.109086	0.131445
0.197028	0.207899	0.197069	0.207837	0.096896	0.165633
0.206709	0.253619	0.206758	0.253693	0.096844	0.143999
0.226231	0.231118	0.226175	0.231111	0.095105	0.127168
0.189231	0.151496	0.189266	0.151979	0.136203	0.154919
0.195822	0.226122	0.195861	0.226061	0.108464	0.127220
0.231199	0.225347	0.231150	0.225369	0.095694	0.171870
0.180301	0.228166	0.180334	0.228107	0.104487	0.104999
0.219485	0.204632	0.219518	0.204702	0.100499	0.184753
0.226120	0.183153	0.226138	0.183235	0.097321	0.586633
0.229599	0.126945	0.229597	0.127466	0.099910	0.556168
0.149010	0.179802	0.148971	0.179806	0.097556	0.574984
0.175290	0.260295	0.175325	0.260294	0.150759	0.262170
0.104508	0.184383	0.104513	0.184428	0.096102	0.827417
0.230007	0.177755	0.229942	0.177783	0.097617	0.786079
0.172413	0.232248	0.172459	0.232219	0.104902	0.098137
0.175919	0.220634	0.175853	0.220687	0.095478	0.472383
0.122947	0.172120	0.122947	0.172168	0.509251	0.114842
0.232218	0.204689	0.232179	0.204760	0.854817	0.190362
0.089374	0.131796	0.089887	0.131777	0.131621	0.157891

نلاحظ من خلال النتائج المبينة في الجدول رقم (6) ان افضل معادلة تنبؤ لانموذج المؤشر الواحد شبه المعلمي بطريقة (SCAD-NPLS) هي مبينة كما في ادناه :

$$\hat{y}_i = \hat{g}_e(0.006969x_{i1} - 0.380558x_{i2} + 0.245487x_{i3} + 0.548585x_{i4} + 0.545814x_{i5} + 0.442722x_{i7}) \quad (29)$$

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

$$\hat{y}_i = \hat{g}_e(0.006985x_{i1} - 0.380563x_{i2} + 0.245583x_{i3} + 0.548587x_{i4} + 0.545839x_{i5} + 0.442644x_{i7}) \quad (30)$$

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

$$\widehat{y}_i = \widehat{g}_1(x_{i7}) \text{ ----- (31)}$$

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

جدول رقم (8) يمثل قيم متوسط مربعات الخطأ (MSE) لجميع طرائق التقدير شبه المعلمية لانموذج المؤشر الواحد للبيانات الحقيقية

Method	MSE
SCAD-NPLS	0.084177
SCAD-MAVE(suggest-1)	0.084184
ALASSO-MAVE(suggest-2)	0.061635

من خلال دراسة نتائج البيانات الحقيقية المبينة في الجداول رقم (6) و(7) و(8) نجد الاتي :
1. 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 .

2. كما نلاحظ من خلال النتائج المبينه في الجدول رقم (8) ان طرائق التقدير شبه المعلمية (NPLS-SCAD), والطريقة المقترحة الاولى (SCAD-MAVE) اعطت نتائج متقاربة لقيم متوسط مربعات الخطأ (MSE) .

3. كما اظهرت النتائج المبينة في الجدول رقم (6) ان الطريقة المقترحة الاولى (SCAD-MAVE) اعطت نتائج متقاربة الى طريقة (SCAD-NPLS) لمتجه المعلمات .

4. نلاحظ من خلال الجدول رقم (7) ان القيم التقديرية لدالة الربط (دالة عائد السهم) لجميع طرائق التقدير شبه المعلمية كانت قريبة من القيم الحقيقية للمتغير المعتمد (عائد السهم) .

8. الاستنتاجات والتوصيات:

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

8.1 الاستنتاجات:

8.1.1 الجانب التجريبي:

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

اولاً: انموذج المؤشر الواحد شبه المعلمي الأول:

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

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

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

ثانياً: نموذج المؤشر الواحد شبه المعلمي الثاني:

1. كحالة عامة اظهرت النتائج ان الطريقة المقترحة الاولى (SCAD-MAVE) هي افضل طريقة في تقدير واختيار المتغير في ان واحد لمعظم حالات النموذج .
2. يلاحظ وبشكل عام ان هناك تذبذباً في قيم (AMSE) مع تزايد حجم العينة فبعضها يزداد بزيادة حجم العينة وبعضها يتناقص بزيادة حجم العينة لمختلف قيم تباينات الأخطاء.
3. كما بينت النتائج ان هناك تزايداً في قيم (AMSE) مع تزايد قيمة تباين الخطأ لكافة حجومات العينات لجميع طرائق التقدير شبه المعلمية .

2.1.8 الجانب العملي:

1. 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 .
2. 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 الزيادة في عائد السهم بمقدار قيمة معاملته .
3. اعطت طرائق التقدير الشبه المعلمية الاتية (SCAD-NPLS والطريقة المقترحة الاولى (MAVE-SCAD)) نتائج متقاربة لقيم متوسط مربعات الخطأ (MSE) .

2.8 التوصيات:

- في ضوء الجانب النظري وبناءً على ماتم التوصل اليه من استنتاجات ادناه اهم التوصيات
1. اوصي باستعمال الطريقة المقترحة الثانية (ALASSO-MAVE) في تحليل النموذج المؤشر الواحد شبه المعلمي كطريقة تقدير واختيار المتغير في آن واحد لما ابدته من كفاءة عالية مقارنة بطرائق التقدير شبه المعلمية الاخرى .
 2. 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 تحذب (concave penalty function minimax).
 3. يمكن استعمال طرائق اخرى لايجاد معلمة عرض الحزمة (hbandwid) مثل طريقة قاعدة الابهام.
 4. اوصي باستعمال صيغ اخرى من الدوال اللبية (kernel) مثل (function Epanechnikov) في طرائق التقدير شبه المعلمية المستعملة في هذا البحث في تحليل النموذج المؤشر الواحد .
 5. بالامكان جعل هذه الدراسة اساساً لدراسات مستقبلية وذلك من خلال توظيفها على نماذج شبه معلمية اخرى مثل نموذج متعدد المؤشرات (Multi-index model).

9. المصادر:

- [1] يوسف , خلود يوسف خمو . (2004) , " مقارنة أساليب بيز مع طرائق أخرى لتقدير منحني الانحدار اللامعلمي " , اطروحة دكتوراه فلسفة في الاحصاء , كلية الإدارة والاقتصاد , جامعة بغداد .
- [2] Akkus ,O . (2011) , " Xplore package for the popular parametric and semi-parametric single index models " . Journal 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 programming " . IEEE xplore , 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] Kopytoun , E., and Santalova , D . (2007) , “ Application of the single index model for forecasting of the inland conveyances “, Recent advances in stochastic modeling and data analysis . Singupre , world scientific publishing 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 strategies in linear and partially linear models “ , A thesis submitted to the faculty of Graduate studies through the department of mathematics and statistics in partial fulfillment 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 Wisconsin , 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 fulfillment 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