

”تقدير نموذج الانحدار اللاخطي المختلط باستخدام الخوارزمية الجينية”

“Estimation of a Mixed Nonlinear Regression Model Using Genetic Algorithm”

م. عمر رمزي جاسم

كلية الإدارة والاقتصاد/ جامعة الحمدانية

omar.ramzi89@uohamdaniya.edu.iq

المستخلص

تُعدّ التوزيعات المختلطة أو المهجنة من التوزيعات العشوائية المهمة التي لها العديد من التطبيقات المختلفة منها الصناعية والاقتصاد الذي يهتم بموضوع الأوراق المالية، لذا تم في البحث تقدير معالم نموذج الانحدار اللاخطي المتعدد عندما يتبع حد الخطأ توزيعات مهجنة بين توزيعين احتماليين منها توزيع الأسّي المزدوج (لابلاس) متعدد المتغيرات وتوزيع (t) متعدد المتغيرات وبالا اعتماد على إحدى الخوارزميات التطورية والمتمثلة بالخوارزمية الجينية. وباستعمال مقدر متعدد الحدود المحلي تم تحويل دالة الانحدار اللاخطية المتعددة إلى دالة خطية وباقتراض أن مصفوفة الأوزان عبارة عن دالة النواة الطبيعية واستعمال قاعدة الإبهام لاختيار معلمة عرض الحزمة. وكجانب عملي للبحث فقد تم التنبؤ بأسعار الأسهم المالية لمختلف القطاعات بسوق العراق للأوراق المالية، ومن خلال برنامج الـ (Matlab R2020a) فقد أظهرت نتائج التنبؤ تفوق النموذج عندما يتبع الخطأ توزيع الأسّي المزدوج متعدد المتغيرات وتوزيع t متعدد المتغيرات على نموذج الانحدار الخطي المتعدد الطبيعي والشبكات العصبية الاصطناعية، وذلك بالاعتماد على معيار متوسط مربعات الخطأ ومعيار متوسط الخطأ المطلق.

الكلمات المفتاحية: نموذج الانحدار اللاخطي المتعدد، مقدر متعدد الحدود المحلي، دوال النواة، الخوارزمية الجينية، توزيع الأسّي المزدوج، الشبكات العصبية الاصطناعية.

Abstract

Mixed or hybrid distributions are considered important random distributions that have many different applications, including industrial and economic applications that are

concerned with the subject of securities. Therefore, in the research, the parameters of the multiple nonlinear regression model were estimated when the error term follows hybrid distributions between two probability distributions, including the multivariate double exponential (Laplace) distribution and the multivariate (t) distribution, and based on one of the evolutionary algorithms represented by the genetic algorithm. Using a local polynomial estimator, the multiple nonlinear regression function was converted into a linear function, assuming that the weight matrix is the normal kernel function and using the thumb rule to choose the bandwidth parameter. As a practical aspect of the research, the prices of financial stocks for various sectors in the Iraq Stock Exchange were predicted, and through the (Matlab R2020a) program, the prediction results showed the superiority of the model when the error follows the multivariate double exponential distribution and the multivariate t distribution over the normal multiple linear regression model and artificial neural networks, based on the mean square error criterion and the mean absolute error criterion.

Keywords: Multiple Nonlinear Regression Model, Local Polynomial Estimator, Kernel Function, Genetic Algorithm, Double Exponential Distribution, Artificial Neural Networks.

Introduction

(1) المقدمة

تختلف تقنيات المستخدمة في التنبؤ على نطاق واسع، وتتأثر بكثير من العوامل بما في ذلك جودة النماذج المتاحة، وأنواع الافتراضات المستخدمة وتوافر البيانات. وبشكل عام، فإن التنبؤ هو مسعى صعب، وهذا هو السبب في أن العديد من العلماء انجذبوا لدراسته.

إن أحد الجوانب الرئيسية لأي نظام لاتخاذ القرارات المالية هو فهم طبيعة العلاقات بين الأسواق المالية واقتصاد الدولة (الغرابي، 2008). فقد برزت توقعات سوق الأوراق المالية مؤخراً كواحدة من مجالات الدراسة الرئيسية. وقد تم تصميم مجال أبحاث سوق الأوراق المالية ليكون فوضوياً وغير معياري وصعباً وغير خطي (الكسو، 2005). وتهدف العديد من الأبحاث العلمية إلى زيادة دقة التنبؤ بالمؤشرات باستخدام كل من الأساليب التقليدية والذكائية. ونظراً لوجود الكثير من التداول في أسواق الأوراق المالية، فقد اكتُشف أن التحسن الطفيف في التنبؤ قد يؤدي إلى ربح كبير.

هناك الكثير من النماذج للتنبؤ بأسعار سوق الأوراق المالية. لاستثمار الأموال في سوق الأوراق المالية، يجب علينا أولاً تحديد ما إذا كانت أسعار الأسهم سترتفع أو تنخفض خلال الأيام القليلة القادمة. للتنبؤ بطبيعة سوق الأوراق المالية، يجب الجمع بين الكثير من تقنيات الحوسبة. لقد تطورت نظرية سوق رأس المال التقليدية، كما تطورت طرق التحليل المالي المختلفة (Heidenreich, 2013).

في أغلب الأحيان يلجأ الباحثون الى تقدير معلمات نماذج الانحدار المختلفة سواء الخطية منها أو اللاخطية عندما يتبع حد الخطأ العشوائي توزيعاً طبيعياً، ولكن هناك العديد من الحالات يكون فيها الخطأ العشوائي ينتمي الى التوزيعات العشوائية الاحتمالية المهجنة أو المختلطة اي بمعنى دمج توزيعين احتماليين، في مثل هذه الحالة تكون التوزيعات الاحتمالية المهجنة أكثر ملائمة من التوزيع الطبيعي ومن هذه التوزيعات توزيع الاسي المزدوج متعدد المتغيرات إذ إن هذا التوزيع يعتبر أكثر عمومية من نماذج الانحدار التي تتبع التوزيع الطبيعي كحالة خاصة، ولهذا التوزيع تطبيقات واسعة في العديد من المجالات الاقتصادية التي تتضمن بيانات الأسهم أو ما يتعلق بمختلف القطاعات لسوق الأوراق المالية.

يُعد نموذج الانحدار اللاخطي المتعدد من نماذج الانحدار المستعملة في تحليل البيانات الاقتصادية المختلفة؛ وذلك لمرونته الكبيرة في التعامل مع التقلبات المالية المتعددة، أذ قدم (الصفوي و نور صباح، 2011) بعض الطرائق الممهدة لتقدير دالة الانحدار اللاخطي عندما يكون الخطأ العشوائي للنموذج يتوزع طبيعياً والمتمثلة بطريقة دالة النواة و طريقة الجار الأقرب وطريقة الشريحة الممهدة، إذ استعمال طريقة العبور الشرعي لاختيار معلمة عرض الحزمة بالنسبة الى طرائق دالة النواة والشريحة الممهدة، أما بالنسبة إلى طريقة الجار الأقرب فقد تم استخدام أسلوب المسافة التقليدية لغرض اختيار معلمة عرض الحزمة، فضلاً عن استعمال أسلوب المحاكاة للمقارنة بين الطرق بنموذجين مختلفين وثلاث قيم لمعلمة التباين.

في حين قام الباحث (الشريفي، 2015)، بتقدير معلمات توزيع الاسي المزدوج باستخدام بعض الطرائق التقدير المختلفة نظراً للأهمية الكبيرة للتوزيع كونه يدخل في نمذجة العديد من الظواهر الطبيعية والاجتماعية والاقتصادية وباستخدام الطرائق الكلاسيكية والبيزية تحت دوال خسارة مختلفة، وقد تم الاستنتاج بان "دالة جامبل" (Gumble function) المقترحة تفوقت على جميع الدوال الأخرى بالاعتماد على كثير من معايير المقارنة المختلفة.

كذلك قام الباحثان: (ابنسام كريم، اسماء خميس، 2022)، بدراسة توزيع الاسي المزدوج ذي الأطراف المتماثلة، لما له من اهمية أساسية في تطوير علم الرياضيات والاحصاء وتطبيق خصائصه في المجال المالي، إذ هدف البحث على الحصول على مقدري بيز لمعلمتي الالتواء والقياس تحت دالة الخسارة التربيعية، وبافتراض أن الدالة الأسبقية أو الأولية لكل من معلمة الالتواء والقياس هي كما والاسي على التوالي، وتم التوصل الى أن مقدر بيز لمعلمتي التواء والقياس تحت دالة الخطأ التربيعية أفضل طريقة بالمقارنة مع الطرق الكلاسيكية الاخرى.

ولغرض الوصول إلى تقدير حسابي أمثل لمعلمات توزيع الاسي المزدوج متعدد المتغيرات سيتم اللجوء إلى استخدام التقنيات الذكائية. إذ تعتمد الدراسات الحديثة على التوليف أو التداخل بين الاتجاهات المختلفة للعلوم، فالتداخل ما بين هذه العلوم يعمل على جمع العديد من المزايا المختلفة مع بعضها والابتعاد أو محاولة تجاوز العيوب الموجودة في بعض الطرائق عبر استخدام الحلول المتوفرة في الاتجاه الموظف فيه. إلا أن موضوع التوظيف في مختلف العلوم شائك ومعقد يحتاج إلى دراية كافية في كلا الاتجاهين المستخدمين لكي تتحقق النتائج المطلوبة، وعلى هذا الأساس، اعتمد هذا البحث على توظيف إحدى الخوارزميات التطورية والمتمثلة بالخوارزمية الجينية في تقدير معلمات توزيع الاسي

المزدوج، واستخدام نموذج الانحدار اللاخطي المتعدد كبديل ناجح للانحدار الخطي المتعدد الطبيعي والشبكات العصبية الاصطناعية، وذلك لغرض التنبؤ بأسعار الأسهم في أسواق العراق للأوراق المالية.

تم تنظيم البحث على جانبين؛ إذ اشتمل الجانب الأول على الجانب النظري الخاص بالبحث الذي تضمن وصف نموذج الانحدار اللاخطي المتعدد عندما يتبع حد الخطأ التوزيع الاسي المزدوج متعدد المتغيرات، وعملية تحويل النموذج الى نموذج خطي بالاعتماد على مقدر متعدد الحدود المحلي، فضلاً عن وصف الخوارزمية الجينية وكيفية استخدامها في تقدير معالم النموذج. في حين تكون الجانب الثاني للبحث من تطبيق ما تم ذكره في الجانب النظري على بيانات حقيقية متمثلة بسوق العراق للأوراق المالية للمدة من (2017/1/1) لغاية (2021/1/1). وأخيراً في هذا البحث عرض أهم الاستنتاجات والتوصيات التي توصل اليها.

Local polynomial estimator

(2) مقدر متعدد الحدود المحلي:

يُعد هذا المقدر من المقدرات المهمة والجيدة المستعملة في المقدرات اللاخطية ويفضل على مقدرات النواة كما انه يستعمل في العديد من التصاميم منها التصميم الثابت او العشوائي لتحليل العديد من التجارب، ويمكن تمثيل نموذج الانحدار اللاخطي المتعدد بالصيغة الآتية: (You et. al. , 2013)

$$Y_i = m(T_i) + e_i \quad \dots (1)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} m(T_1) \\ m(T_2) \\ \vdots \\ m(T_n) \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \quad \dots (2)$$

إذ إن $m(T_i)$ تمثل متجه دالة مجهولة ذو (T) من المتغيرات التوضيحية اللاخطية وعادة ما يكون متغير مستمر، ونلاحظ من المعادلة (1) بان انموذج الانحدار اللاخطي المتعدد يكون غير خطي وعليه تم تحويله الى نموذج خطي بالاعتماد على توسيع تايلر (Taylor Expansion) وكالاتي:

نفترض بان الدالة (m) تمتلك مشتقة من الرتبة (u) عند النقطة (t) والنقاط (T_i) تقع في جوار النقطة (t) وعليه فان توسيع تايلر للدالة (m) يكون كالاتي:

$$m(T_i) = \beta_0 + \beta_1(t - T_i) + \beta_2(t - T_i)^2 + \dots + \beta_u(t - T_i)^u \quad \dots (3)$$

$$m(T_i) = \sum_{j=0}^u \beta_j(t - T_i)^j \quad , \quad i = 1, 2, \dots, n \quad \dots (4)$$

ومن المعادلة (4) نلاحظ بأن دالة الانحدار اللاخطية اصبحت خطية، وعليه يمكن كتابة النموذج الخطي بصيغة المصفوفات وكالاتي: (You et. al. , 2013)

$$\underline{Y} = T\underline{\beta} + \underline{e} \quad \dots (5)$$

أذن أن:

$$T = \begin{bmatrix} (t - T_1) & (t - T_1)^2 & (t - T_1)^3 & \dots & (t - T_1)^u \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (t - T_n) & (t - T_n)^2 & (t - T_n)^3 & \dots & (t - T_n)^u \end{bmatrix}_{n \times u}, \quad \underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_u \end{bmatrix}_{u \times 1}$$

(3) وصف نموذج الانحدار اللاخطي المتعدد المختلط:

Description of the Mixed Multiple Nonlinear Regression Model:

من خلال نموذج الانحدار اللاخطي المتعدد الذي تم الإشارة إليه في المعادلة رقم (5)، وبافتراض أن متجه الأخطاء العشوائية يتبع التوزيع الأسّي المزدوج متعدد المتغيرات، حيث يمكن إيجاد دالة الكثافة الاحتمالية بالاستعانة بمفهوم التوزيعات المختلطة من التوزيع الأسّي والطبيعي متعدد المتغيرات وكالآتي:

$$\underline{e}|X \sim N_n(0, X \sigma^2 \varphi^{-1}) \quad , \quad X \sim \exp(\sigma^2)$$

إذن إن:

X : متغير عشوائي يتبع التوزيع الأسّي.

σ^2 : معلمة التباين.

φ : مصفوفة قطرية تمثل أوزان دالة النواة من الدرجة $(n \times n)$.

$$\varphi = \text{diag}(\ker_h(t - T_i)) \quad ; \quad \ker_h(.) = \frac{1}{h} \ker\left(\frac{\cdot}{h}\right)$$

$\ker_h(.)$: دالة النواة وتكون محددة ومستمرة ومتماثلة وتكاملها مساوي للواحد الصحيح، وفي هذا البحث، تم استعمال دالة النواة الطبيعية. (Heidenreich , et al. 2013)

h : معلمة عرض الحزمة وتكون موجبة، وأكثر أهمية من دالة النواة، ومن الممكن اختيارها حسب خبرة الباحث، وكانت كالآتي: (Heidenreich , et al. 2013)

$$h = 1.06 * \hat{\sigma} * n^{-\frac{1}{5}} \quad \dots (6)$$

وعليه، فإن دالة الكثافة الاحتمال لمتجه الأخطاء العشوائية المشروطة بالمتغير العشوائي (X) هي كالآتي:

$$f(\underline{e}|X) = \frac{|\boldsymbol{\varphi}|^{\frac{1}{2}}}{(2\pi X)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2 X} (\underline{e})^T \boldsymbol{\varphi}(\underline{e})} \quad \dots (7)$$

وان دالة الكثافة الاحتمال للمتغير العشوائي (X) تكون كالآتي: (الغرابي، 2008)

$$P(X) = \frac{1}{\sigma^2} \exp\left[-\frac{X}{\sigma^2}\right] \quad ; X > 0 \quad \dots (8)$$

وبالاعتماد على نظرية التوزيعات الخليطة، فإن دالة الكثافة الاحتمال لمتجه الاخطاء غير الشرطية بالمتغير العشوائي (X) تكون كالآتي:

$$f(\underline{e}) = \int_0^\infty f(\underline{e}|X) P(X) dX$$

$$f(\underline{e}) = \frac{2 |\boldsymbol{\varphi}|^{\frac{1}{2}} \left(\frac{(\underline{e})^T \boldsymbol{\varphi}(\underline{e})}{2} \right)^{\frac{2-n}{2}}}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} * K_{\frac{2-n}{2}} \left(\sqrt{2 (\underline{e})^T \boldsymbol{\varphi}(\underline{e})} \right) \quad \dots (9)$$

المعادلة (9) تمثل دالة الكثافة الاحتمال للتوزيع الأسّي المزدوج متعدد المتغيرات. (Eltoft et al., 2006)

أذ أن: (Lemonte and Cordeiro, 2011)

$K_{\frac{2-n}{2}}(.)$: تمثل دالة بسل المعدلة من النوع الثاني ذات الرتبة $\left(\frac{2-n}{2}\right)$ التي تأخذ الصيغة الآتية:

$$K_{\frac{2-n}{2}}(x) = \frac{1}{2} \int_0^\infty y^{-\frac{2-n}{2}-1} \exp\left(-\frac{1}{2} x(y + y^{-1})\right) dy \quad x > 0 \quad \dots (10)$$

وعليه، فإن التوزيع الاحتمالي العشوائي لمتجه المشاهدات (Y) هو توزيع الأسّي المزدوج متعدد المتغيرات؛ لأنه تركيبة خطية بدلالة متجه الاخطاء العشوائية ، وأن دالة الكثافة الاحتماليه له تكون كالآتي:

$$f(\underline{Y}) = \frac{2 |\boldsymbol{\Phi}|^{\frac{1}{2}} \left(\frac{(\underline{Y} - T\boldsymbol{\beta})^T \boldsymbol{\Phi} (\underline{Y} - T\boldsymbol{\beta})}{2} \right)^{\frac{2-n}{2}}}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} \times K_{\frac{2-n}{2}} \left(\sqrt{2 (\underline{Y} - T\boldsymbol{\beta})^T \boldsymbol{\Phi} (\underline{Y} - T\boldsymbol{\beta})} \right) \quad \dots (11)$$

ويُمكن التعبير عن هذا التوزيع وصفيًا بالآتي:

$$\underline{Y} \sim DE(L)_{(n)}(T\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Phi}^{-1})$$

Genetic Algorithm :

(4) الخوارزمية الجينية:

تكمن فكرة الاساسية للخوارزمية الجينية من خلال توليد بعض الحلول للمشكلة عشوائياً، ثم يتم فحص هذه الحلول ومقارنتها لبعض المعايير التي يضعها مصمم الخوارزمية، وأفضل الحلول هي التي تبقى، أما الحلول الأقل كفاءة فيتم إهمالها عملاً بالقاعدة البيولوجية (البقاء للأصلح)، والخطوة اللاحقة هي مزوجة الحلول المتبقية (الحلول الأكثر كفاءة) لإنتاج حلول جديدة على غرار ما يحصل في الكائنات الحية؛ وذلك بمزج مورثاتها (جيناتها)؛ إذ إن الكائن الجديد الناتج ستكون صفاته عبارة عن مزيج من صفات والديه وهذه الحلول الناتجة من التزاوج تدخل أيضاً تحت الفحص والتنقيح لمعرفة مدى كفاءتها واقتربها من الحل الأمثل، وبهذا تتم عمليات التزاوج والانتقاء حتى تصل العملية إما لعدد معين من التكرارات (يقدره مصمم الخوارزمية) أو تصل الحلول الناتجة أو إحداها إلى نسبة كفاءة كبيرة. وهكذا فالخوارزميات الجينية تولد بنجاح الفرضيات بوساطة تكرار الطفرة وإعادة تجميع أفضل أجزاء الفرضيات الموجودة.

إن الخوارزمية الجينية يمكن أن تعالج أية دالة هدف مع القيود أو دونها سواء أكانت خطية أم لا خطية، وفضاء الحل قد يكون ببعد واحد أو أكثر (Goldberg 1989) وبصورة عامة فإن خطوات عمل الخوارزمية الجينية توصف كالآتي :

1- [البداية Start]: تمثل توليد مجتمع عشوائي أولي من الكروموسومات، أي بمعنى إيجاد حلول مناسبة أولية للمسألة.

2- [اللياقة Fitness]: هي تحويل دالة الهدف (Objective Function) إلى دالة مناسبة للحل في الخوارزمية الجينية.

3- [مجتمع جديد New Population]: توليد جيل جديد من الأفراد بتكرار الخطوات الآتية إلى أن يكتمل الجيل،

وتتضمن:

- الاختيار (Selection): يتم اختيار اثنين من الكروموسومات من المجتمع الابتدائي كوالدين استناداً إلى دالة اللياقة (أفضل القيم التي لها فرص أكبر للاختيار).

- التداخل الإبدالي (Crossover): يعني إجراء إحدى عمليات التعابر للحصول على الذرية (Offspring) ويكون بين كروموسومين يتم تحديد هم سابقاً.
 - الطفرة (Mutation): مع احتمالية وجود الطفرة يتم عمل الطفرة للسلف الجديد بموقع معين في الكروموسوم، وتجري بين الجينات في الكروموسوم الواحد.
 - الاستبدال (Replacement): تستبدل الابناء في المجتمع الجديد ليصبحوا آباءً في المجتمع اللاحق.
 - الاختبار (Test): عند تحقق شرط التوقف، فإن الخوارزمية الجينية تتوقف وتعيد الحل الجيد من آخر جيل متكون.
 - الدورة (Loop): يتم الرجوع إلى الخطوة 2.
- إن كل تكرار (دورة) لهذه العملية يسمى بالجيل (Generation)، وبعد نهاية التنفيذ يقوم الباحث بتقديم تقرير عن الحقائق التي تم التوصل إليها (Fogel 1998).

(5) استخدام الخوارزمية الجينية لتقدير معلمات نموذج الانحدار اللاخطي المتعدد المتخط:

Using the Genetic Algorithm to Estimate the Parameters of the Mixed Multiple Nonlinear Regression Model:

من المعروف أن الخوارزمية الجينية إحدى الخوارزميات التطورية الذكية التي تقوم بانتخاب الحلول المفضلة من بين عدد كبير من الحلول وإجراء بعض التداخلات والتبديلات بين هذه الحلول بهدف تكوين حلول أفضل. وبشكل عام فإن الخوارزمية الجينية تتضمن عدداً من الخطوات الأساسية لحل مختلف المسائل. ويكون الاختلاف في أسلوب صياغة وتطبيق كل خطوة من هذه الخطوات حسب المسألة ومجال تطبيقها وفيما يأتي توضيح لخطوات الخوارزمية الجينية المقترحة:

1. البيانات الأولية (Initial Data): وهي قراءة لقيم متغيرات مسألة الانحدار اللاخطي المتعدد عندما يتبع حد الخطأ التوزيع الأسّي المزدوج وتتضمن:

- متجه مشاهدات متغير الاستجابة (Response variable)، ذو سعة $(n \times 1)$.
- المصفوفة غير العشوائية والتي تمثل مشاهدات المتغيرات التوضيحية (Explanatory variables) ذات سعة $(n \times u)$.

إذ يكون الجزءان معاً المصفوفة التي تمثل الإدخال للمسألة، وتصبح المصفوفة بالشكل الآتي:

$$T = \begin{bmatrix} (t - T_1) & (t - T_1)^2 & \dots & (t - T_1)^u & Y_{u+1} \\ (t - T_1)^2 & (t - T_1)^3 & \dots & (t - T_1)^{u+1} & Y_{u+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (t - T_1)^{n-u} & (t - T_1)^{n-u+1} & \dots & (t - T_1)^{n-1} & Y_n \end{bmatrix} \dots (12)$$

تجدر الإشارة إلى أنه يجب أن تتم معايرة البيانات (Normalization Data) قبل إدخالها للشبكة؛ لأنه في حالات كثيرة تكون البيانات المعدة للتدريب ذات قيم عالية، عندئذ ينصح دائماً بمعايرة مدخلات ومخرجات الشبكة لتصبح ضمن المدى $[-1, 1]$ ، إذ يستخدم القانون الآتي والخاص بمعايرة البيانات:

$$\hat{T} = 2 * \left(\frac{T - T_{min}}{T_{max} - T_{min}} \right) - 1 \dots (13)$$

إذ إن $(\hat{T})$ تمثل القيم بعد المعايرة و (T) تمثل القيم الأصلية، كما أن (T_{max}) و (T_{min}) يمثلان أكبر وأصغر قيمة من قيم (T) على التوالي.

2. إنشاء الجيل الابتدائي (Initial Generation): ويُعد النقطة الأولى لحل المسألة، وتتم عملية بناء الجيل الابتدائي بطريقة عشوائية طبقاً للتوزيع المنتظم من خلال توليد قيم عشوائية لمتجه معاملات نموذج الانحدار اللاخطي (β) طبقاً للتوزيع المنتظم في الفترة $U[-1, 1]$ ، ومن ثم توليد قيم عشوائية لمعاملات التباين (σ^2) طبقاً كذلك للتوزيع المنتظم.

ومن ثم تنشأ عدة كروموسومات ابتدائية بحجم (N_{pop}) ، ويمثل كل كروموسوم حل متكامل للمسألة، كما أن طول الكروموسوم يكون بعدد معاملات المسألة. والمقطع أدناه يوضح هيئة الكروموسوم لمسألة انحدار اللاخطي المتعدد عندما يتبع حد الخطأ التوزيع الأسّي المزدوج.

β_1	β_2	...	β_u	σ^2
-----------	-----------	-----	-----------	------------

الشكل (1): كروموسوم خاص بمسألة الانحدار اللاخطي المتبع توزيع الأسّي المزدوج.

3. دالة الهدف (Objective Function): إن دالة الهدف الخاصة بمسألة الانحدار اللاخطي عندما يتبع توزيع الأسّي المزدوج هي عبارة عن دالة الخطأ التربيعي (MSE) .

4. اختيار الآباء (Selection): تم اختيار الآباء في هذه الخوارزمية على أساس انتقاء الفرد ذي الرتبة الأعلى (Highest Rank Selection) إذ تعتمد هذه الطريقة على اختيار احدى الافراد التي تمتلك أعلى هدف، بينما يتم انتقاء الفرد الآخر عشوائياً، وتسمح هذه الطريقة بتحسين الكروموسومات بحيث تمتلك بعض الجينات الجيدة من الفرد الذي يمتلك الكلفة الأكبر. (Goldberg, 1998)

5. **التداخل الابدالي (Crossover):** يتمثل عمل التداخل باختيار فردين وربطهما خطياً لإنتاج نوعين جديدين من الأفراد (الأبناء) وهذا النوع من التداخل يسمى (Arithmetic Crossover)، ويتم بتطبيق المعادلتين الآتيتين:

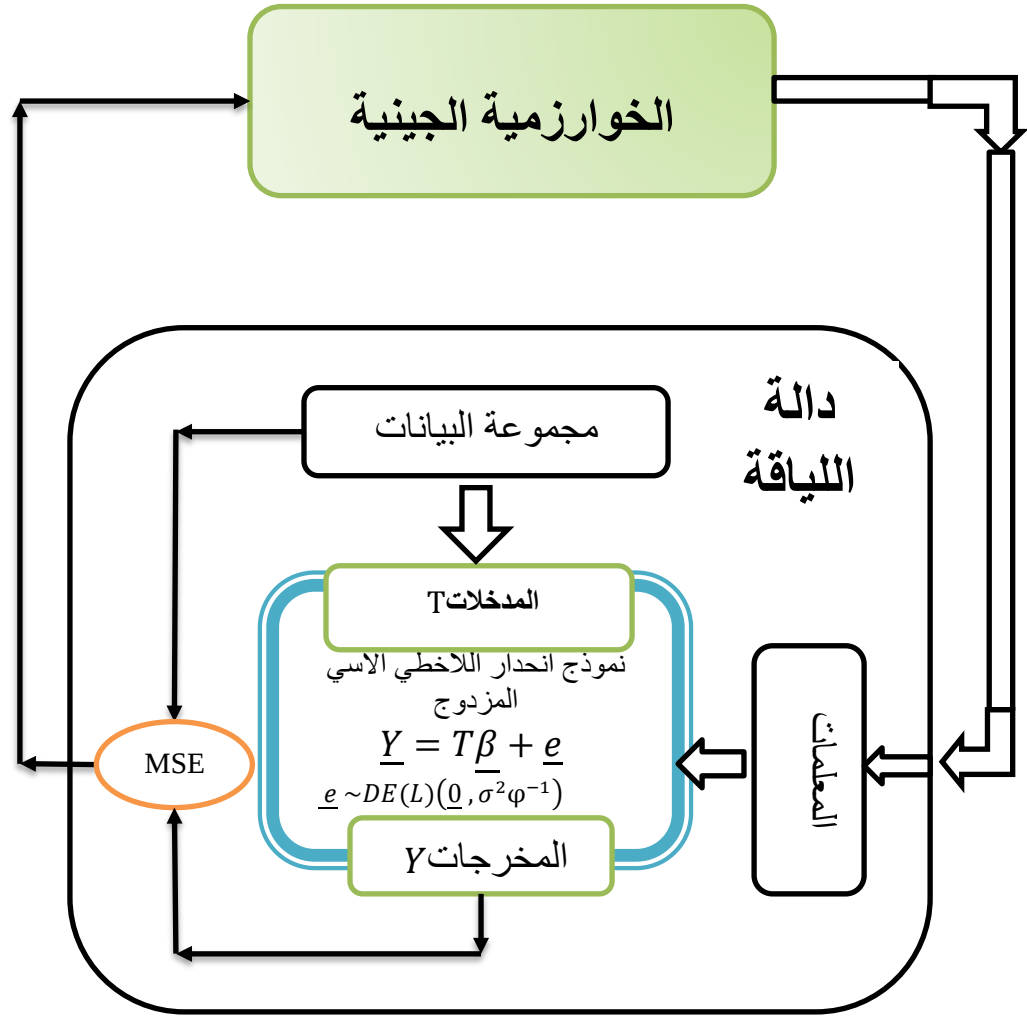
$$\text{offspring 1} = a * \text{Parent 1} + (1 - a) * \text{Parent 2} \quad \dots (14)$$

$$\text{offspring 2} = (1 - a) * \text{Parent 1} + a * \text{Parent 2} \quad \dots (15)$$

إذ أن (a): يمثل عدداً عشوائياً محصوراً بين الصفر والواحد (ويتم اختياره قبل كل عملية تداخل).

6. **الطفرة (Mutation):** تم اجراء الطفرة في هذه الخوارزمية بالاعتماد على طفرة التوزيع المنتظم، إذ يتم اختيار موقع ضمن المقطع (الكروموسوم) ويتم استبدال قيمته بشكل عشوائي وضمن الحدود الخاصة به وباحتمالية مقدارها $(p(m) = 0.1)$.

7. **مقياس التوقف (Stop Criterion):** يستمر تكوين الأجيال المتعاقبة (الخطوات 4-7) بهدف تحسين الحل (أي بجعله أكثر اقتراباً إلى الحل الأمثل) إلى أن يتحقق شرط التوقف الذي يعتمد على انتهاء العدد المحدد من التكرارات، كما يمكن إيقاف الخوارزمية في حالة عدم تحسن الحل، والشكل (2) يوضح طريقة ربط الخوارزمية الجينية مع نموذج انحدار اللاخطي عندما يتبع توزيع الأسّي المزدوج .



الشكل (2) : كيفية ربط الخوارزمية الجينية بنموذج انحدار اللاخطي الاسي المزدوج المتعدد.

The Practical Side :

(6) الجانب التطبيقي:

يتناول هذا الجانب تطبيق ما تم التوصل اليه في الجانب النظري على العديد من البيانات الحقيقية المتعلقة بأسواق العراق للأوراق المالية والخاصة في العديد من القطاعات المختلفة بالإضافة الى مناقشة تلك النتائج.

أ- البيانات (Data) :

تم في هذا البحث استخدام ست مجاميع من البيانات لغرض اثبات فعالية نموذج الانحدار اللاخطي الأسّي المزدوج المتعدد في التنبؤ بأسعار الأسهم والجدول (1) يوضح تفاصيل البيانات المستخدمة.

الجدول (1): تفاصيل البيانات الخاصة بالأسهم.

ت	اسم الشركة	الرمز	عدد المشاهدات	نوع المشاهدات
1	المصرف التجاري العراقي	BCOI	40	شهري
2	الخليج للتأمين	TZNI	38	شهري
3	بغداد العراق للنقل العام	SBPT	44	شهري
4	انتاج الالبسة الجاهزة	IRMC	35	شهري
5	صناعة المواد الانشائية الحديثة	IMCM	39	شهري
6	نقل المنتجات النفطية	SIGT	42	شهري

قبل إجراء عملية التنبؤ تم مطابقة بيانات شركات القطاعات المختلفة لتوزيع الأسى المزدوج من خلال اختبار حسن المطابقة (**Goodness of Fit Test**) والمتمثل باختبار كولموجروف سميرنوف، وكانت قيم الاختبار لجميع الشركات أصغر من القيمة الجدولية تحت مستوى معنوية ($\alpha = 0.05$)، مما يدل على أن البيانات مطابقة للنموذج المستخدم في البحث والجدول (2) يبين قيم الاختبار الخاص بحسن المطابقة.

الجدول (2): قيم اختبار كولموجروف سميرنوف.

ت	اسم الشركة	الرمز	قيم اختبار كولموجروف سميرنوف	القيمة الجدولية
1	المصرف التجاري العراقي	BCOI	1.9871	$D(0.05,40) = 0.2102$
2	الخليج للتأمين	TZNI	2.0014	$D(0.05,38) = 0.2172$
3	بغداد العراق للنقل العام	SBPT	1.5214	$D(0.05,44) = 0.2043$
4	انتاج الالبسة الجاهزة	IRMC	2.0102	$D(0.05,35) = 0.2242$
5	صناعة المواد الانشائية الحديثة	IMCM	1.8214	$D(0.05,39) = 0.2172$

6	نقل المنتجات النفطية	SIGT	1.4422	$D(0.05,42) = 0.2043$
---	----------------------	------	--------	-----------------------

ولتقييم أداء التنبؤ لبيانات الأسهم المختلفة، تم تقسيم البيانات على مجموعتين هما: مجموعة التدريب ومجموعة الاختبار، وهاتان المجموعتان مستقلتان لا وجود لأي ترابط بينهما؛ إذ تستخدم بيانات مجموعة التدريب لغرض بناء نموذج الانحدار اللاخطي وتقدير معلماته، في حين يتم استخدام بيانات الاختبار لتقييم كفاءة النماذج المستخدمة، وقد تم تقسيم البيانات إلى (80%) و (20%) لبيان هيكلية مجموعتي التدريب والاختبار للبيانات على التوالي.

ب- معايير المقارنة:

لغرض المقارنة ما بين النماذج المستخدمة لغرض التنبؤ بأسعار الأسهم المالية، تم استخدام عدد من معايير اختبار النماذج ومعرفة مدى صلاحيتها ومدى قرب القيم الحقيقية من القيم المقدرة فيها. وقد استخدم كل من متوسط مربعات الخطأ (MSE) ومتوسط القيم المطلقة للأخطاء (MAE) في المعادلات (16) و (17) على التوالي.

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad \dots (16)$$

$$MAE = \frac{\sum_{i=1}^n (|y_i - \hat{y}_i|)}{n} \quad \dots (17)$$

إذ إن (y) تمثل القيمة الحقيقية و ($\hat{y}$) تمثل القيمة المقدرة باستخدام الأساليب المختلفة، بينما تمثل (n) عدد المشاهدات المستخدمة.

ج- تقدير معلمات نموذج الانحدار اللاخطي الأسّي المزدوج المتعدد والتنبؤ به:

تم في هذا المبحث استعمال نموذج الانحدار الخطي الطبيعي المتعدد ونموذج انحدار (t) المتعدد والشبكات العصبية الاصطناعية لغرض المقارنة مع نموذج انحدار اللاخطي الأسّي المزدوج المتعدد. (Zhang,1998) و يبين الجدول (3) نتائج المقارنة باستخدام المعيارين في المعادلة (16) و (17)، بينما تبين الأشكال (3) و (4) و (5) و (6) و (7) و (8) رسم القيم الحقيقية والقيم المتنبئ بها. (عيسى، 2000).

الجدول (3): نتائج مقارنة نموذج الانحدار اللاخطي الآسي المزدوج المتعدد مع نموذج الانحدار الطبيعي المتعدد ونموذج انحدار (t) المتعدد والشبكات العصبية الاصطناعية.

البيانات	المعيار	نموذج الانحدار اللاخطي الآسي المزدوج		نموذج الانحدار الخطي الطبيعي		نموذج الانحدار الخطي (t)		الشبكات العصبية الاصطناعية	
		اختبار	تدريب	اختبار	تدريب	اختبار	تدريب	اختبار	تدريب
BCOI	MSE	0.4298	0.4421	0.9941	0.4620	0.9721	0.4501	0.9941	0.4523
	MAE	0.4298	0.3140	0.9941	0.4431	0.9643	0.3911	0.9941	0.4210
TZNI	MSE	1.5527	1.2245	15.2412	1.3098	2.2245	1.2539	15.2412	1.2838
	MAE	1.5527	1.0021	15.2412	1.1147	2.0024	1.0999	15.2412	1.1021
SBPT	MSE	1.3080	0.7642	11.3942	0.8002	3.3698	0.7758	11.3942	0.7821
	MAE	3.3080	0.6631	11.3942	0.7820	5.3241	0.6829	11.3942	0.7088
IRMC	MSE	0.7724	0.8422	0.7700	3.2892	0.7926	1.8952	0.8021	1.9952
	MAE	0.7620	0.8071	0.7501	3.0014	0.7809	0.9075	0.7991	1.0098
IMCM	MSE	0.8290	0.9871	1.9987	1.2347	0.8923	1.0021	1.1147	2.0142
	MAE	0.8541	0.9001	0.8998	1.0021	0.9982	0.1080	1.9327	1.8215
SIGT	MSE	2.1874	2.3210	4.4422	4.5741	3.8921	3.9987	4.2479	4.0147
	MAE	2.0047	2.1012	4.4148	4.4231	2.7147	3.7720	4.0078	3.9921

1- نموذج الانحدار الخطي الطبيعي المتعدد: تم استعمال برنامج (SPSS) لتقدير معلمات نموذج الانحدار الخطي المتعدد الاعتيادي لمجاميع البيانات (الجدول (3)) وتم تقدير المعلمات لمجموعة التدريب، بينما تم اختبار المعلمات المقدرة باستخدام بيانات الاختبار. (Sarle and Warren, 1994)

2- نموذج الانحدار الخطي (t): تم استعمال برنامج (SPSS) لتقدير معلمات نموذج الانحدار الخطي (t) لمجاميع البيانات (الجدول (3)) وتم تقدير المعلمات لمجموعة التدريب ، بينما تم اختبار المعلمات المقدرة باستخدام بيانات الاختبار.

3- **الشبكات العصبية الاصطناعية:** ولغرض التنبؤ استخدمت شبكة عصبية متعددة الطبقات ذات انتشار الامامي (multi-layer feed forward neural network) و بطبقة مخفية واحدة واستخدام عقدة اخراج واحدة واستعمال تنبؤ أمامي متعدد الخطوات باستخدام إجراءات تكرارية (الكسو، 2005).

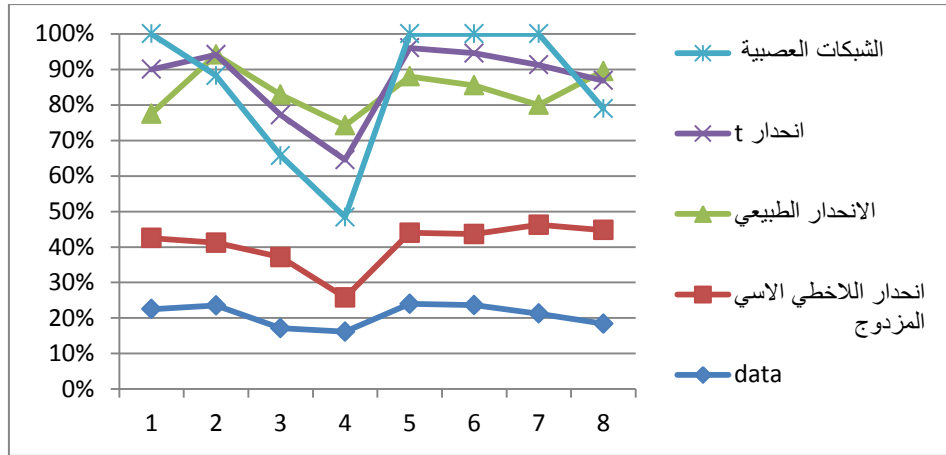
كما تم اختيار عقد ادخال بقدر عدد المتغيرات المستقلة، مع عقد مخفية تتفاوت ما بين [3 – 11] بزيادة بمقدار اثنين عن سابقتها، لذا فإن ما مجموعه (8) نموذج شبكة عصبية تم اختبارها قبل الوصول الى التركيبة النهائية لنموذج الشبكة العصبية. وقد تم استخدام برنامج (Matlab R2020a) لبناء الشبكة وتقدير الاوزان واختبارها.

4- **نموذج انحدار اللاخطي الأسّي المزدوج المتعدد:** تم تحديد معلمات الخوارزمية الجينية قبل البدء بعملية تقدير المعلمات وذلك لغرض الحصول على أفضل نتيجة وأسرع تقارب والجدول (4) يوضح معلمات الخوارزمية الجينية.

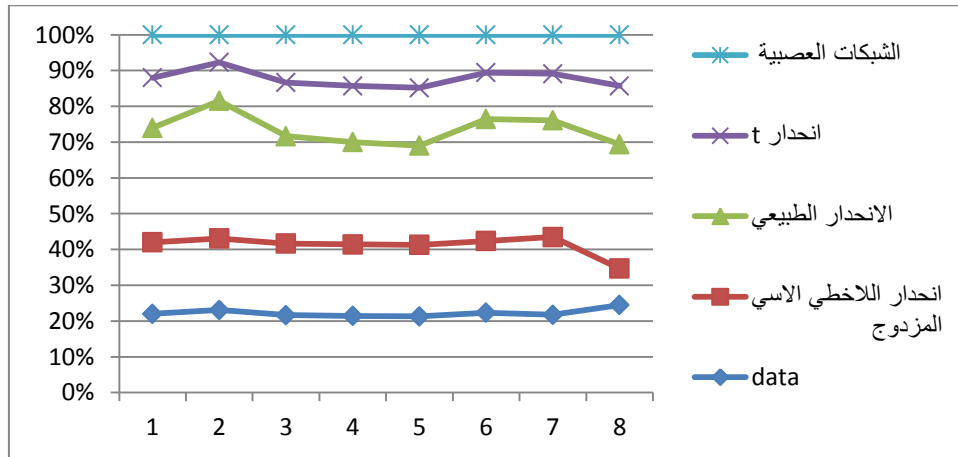
الجدول (4): معلمات الخوارزمية الجينية.

المعلمة	القيمة
حجم المجتمع (عدد الافراد)	100
عدد التكرارات	2500
احتمالية التعابر	0.75
احتمالية الطفرة	0.46
نوع الانتقاء	Highest Rank Selection
نوع التعابر	Arithmetic
نوع الطفرة	Uniform
دالة اللياقة	MSE

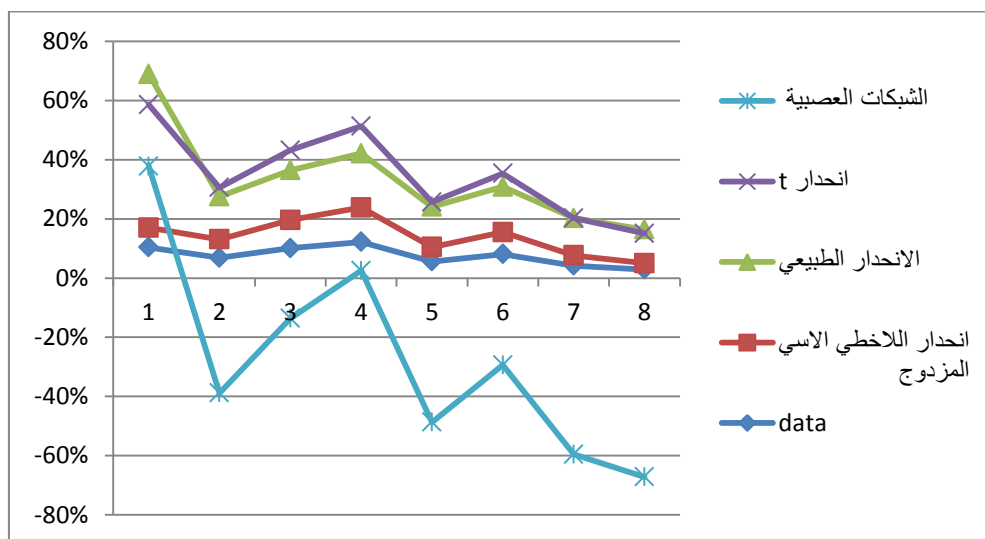
بعد تقدير معلمات نموذج انحدار اللاخطي الأسّي المزدوج المتعدد باستعمال الخوارزمية الجينية من خلال بيانات التدريب، تم اختبار كفاءة النموذج باستخدام المعايير الموضحة سابقاً على بيانات الاختبار لمعرفة أفضل نموذج يمكن استخدامه للتنبؤ بأسعار الأسهم من النماذج الستة المستخدمة.



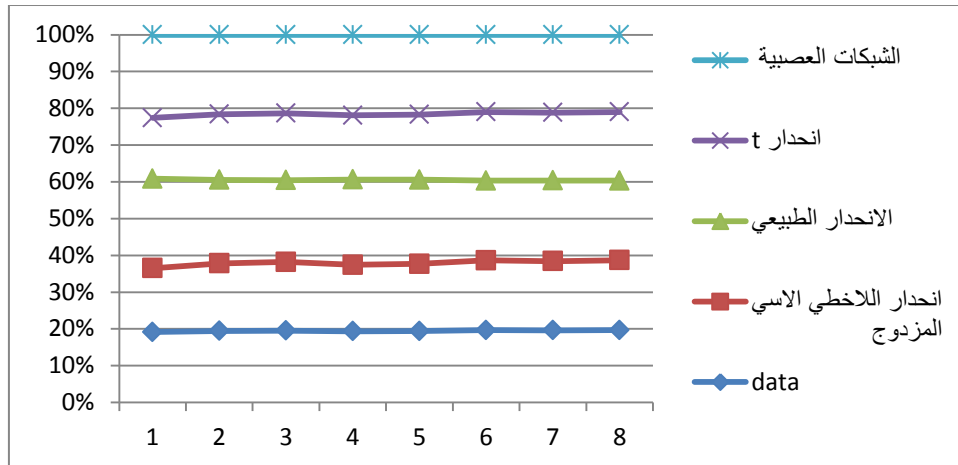
الشكل (3): رسم القيم الحقيقية والقيم المتنبئة لشركة BCOI.



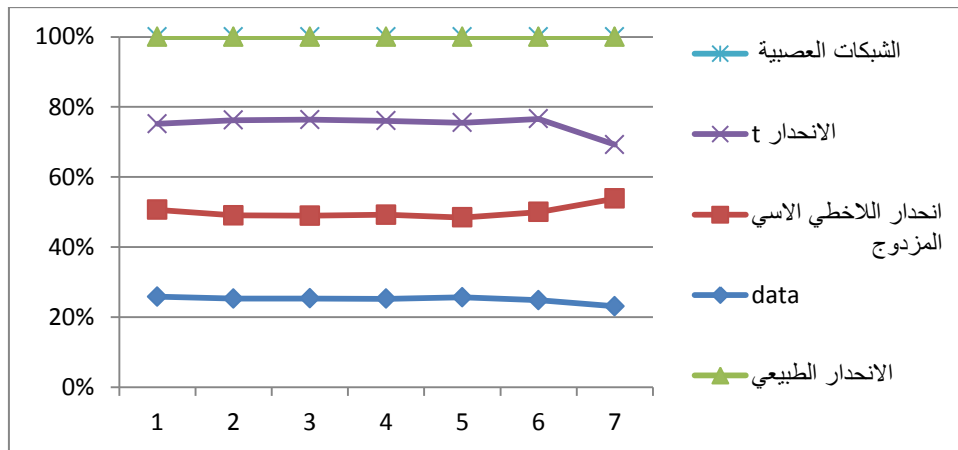
الشكل (4): رسم القيم الحقيقية والقيم المتنبئة لشركة TZNI.



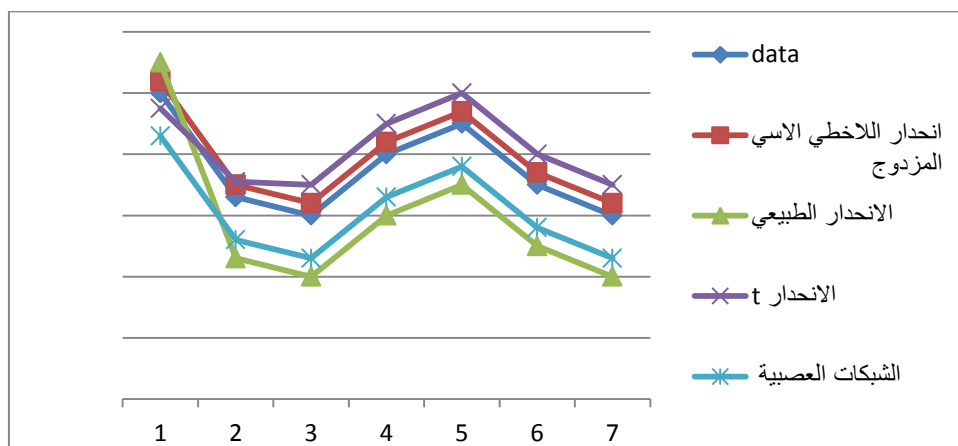
الشكل (5): رسم القيم الحقيقية والقيم المتنبئة لشركة SBPT.



الشكل (6): رسم القيم الحقيقية والقيم المتنبئة لشركة IRMC.



الشكل (7): رسم القيم الحقيقية والقيم المتنبئة لشركة IMCM.



الشكل (8): رسم القيم الحقيقية والقيم المتنبئة لشركة SIGT.

د- مناقشة النتائج

من الجدول (3) ومن ملاحظة الأشكال (3) و (4) و (5) و (6) و (7) و (8) يتبين أن نموذج انحدار اللاخطي الأسّي المزدوج المتعدد أفضل من نموذج الانحدار الخطي الطبيعي ونموذج الانحدار الخطي (t) وأسلوب الشبكات العصبية الاصطناعية في التنبؤ بأسعار الأسهم؛ إذ أثبتت النتائج قدرة نموذج الانحدار اللاخطي الأسّي المزدوج المتعدد على الموازنة بين التدريب والاختبار، وهذا يدل على قوة النموذج وكفاءته مقارنة بنموذج الاعتيادي وأسلوب الشبكات العصبية الاصطناعية اللذين يستخدمان كذلك في مجال التنبؤ.

Conclusions and recommendations:

(7) الاستنتاجات والتوصيات:

تم الاستنتاج من خلال النتائج التي تم التوصل إليها إلى أن استخدام أسلوب الانحدار اللاخطي الأسّي المزدوج المتعدد كبديل ناجح لنموذج الانحدار الخطي الطبيعي المتعدد ونموذج انحدار (t) والشبكات العصبية، وذلك لغرض التنبؤ بأسعار الأسهم في أسواق الأوراق المالية. وتم استخدام الخوارزمية الجينية لتقدير معاملات التوزيع الأسّي المزدوج. وطبقت النتائج التي توصلنا إليها على العديد من البيانات الحقيقية المتعلقة بسوق العراق للأوراق المالية، وقد أظهرت نتائج التنبؤ تفوق نموذج انحدار اللاخطي الأسّي المزدوج المتعدد على نموذج الانحدار الخطي الطبيعي و(t) المتعدد والشبكات العصبية في بيانات التدريب والاختبار، مما يدل على فعالية النموذج لكون توزيع الأسّي المزدوج توزيعاً شاملاً يستطيع احتواء أغلب الظواهر والبيانات. ويوصي الباحث باعتماد نموذج الانحدار اللاخطي الأسّي المزدوج المعمم والخوارزمية الجينية في نمذجة جميع الظواهر الحياتية، واستخدامه كبديل للانحدار الاعتيادي المتعدد.

(8)المصادر:

References:

1. الصفاوي، صفاء يونس و متي، نور صباح، (2011)، "تقدير دوال الانحدار اللامعلمي باستخدام بعض أساليب التمهيد"، المجلة العراقية للعلوم الاحصائية، عدد خاص بوقائع المؤتمر العلمي الرابع لكلية علوم الحاسوب والرياضيات، (20)، 373-392.
2. الغرابي، سليم، (2008)، "ملاحظات حول التوزيع الأسّي"، مجلة العلوم الاقتصادية والادارية، المجلد 14، العدد 50، ص ص (292-303).
3. الكسو، ابتهاج عبد الحميد محمد علي، (2005)، "استخدام الشبكات العصبية في تقدير رتب سلاسل ماركوف مع التطبيق على سلسلة جبل بطمة في محافظة نينوى"، اطروحة دكتوراه غير منشورة كلية علوم الحاسبات والرياضيات، جامعة الموصل.
4. عبدالله، ابتسام كريم وراضي، اسماء خميس، (2022)، "مقارنة طرائق تقدير معاملات توزيع لابلاس الغير متماثل باستعمال دالة خسارة التربيعية وطريقة الإمكان الأعظم" مجلة الادارة والاقتصاد / جامعة المستنصرية.

5. الشريفي، عماد فرهود محي، (2015)، " بعض طرائق التقدير لمعلمة القياس لتوزيع لابلاس : دراسة مقارنة" رسالة ماجستير غير منشورة، قسم الرياضيات، جامعة المستنصرية.
6. عيسى، علام زكي، (2000)، " الشبكات العصبية، البنية الهندسية-الخوارزميات – التطبيقات"، الطبقة الاولى، شعاع للنشر والعلوم ، حلب، سوريا.

Eltoft, T., Kim, T., & Lee, T. W. (2006). On the multivariate Laplace distribution. *IEEE Signal Processing Letters*, 13(5), 300-303.

8. Fogel D., Evolutionary Computation: The Fossil Record. Piscataway NJ: IEEE Press, 1998.
9. Fogel, D. B. (2000). What is evolutionary computation?. *IEEE spectrum*, 37(2), 26-32.
10. Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning, addison-wesley, reading, ma, 1989. *NN Schraudolph and J*, 3(1).
11. Heidenreich, N. B., Schindler, A., & Sperlich, S. (2013). Bandwidth selection for kernel density estimation: a review of fully automatic selectors. *AStA Advances in Statistical Analysis*, 97, 403-433.
12. Lemonte, A. J., & Cordeiro, G. M. (2011). The exponentiated generalized inverse Gaussian distribution. *Statistics & probability letters*, 81(4), 506-517.
13. Mühlenbein, H., Schomisch, M., & Born, J. (1991). The parallel genetic algorithm as function optimizer. *Parallel computing*, 17(6-7), 619-632.
14. Sarle, W. S. (1994). Neural networks and statistical models. In *Proceedings of the 19th annual SAS users group international conference* (Vol. 13).
15. You, J., Zhou, Y., & Chen, G. (2013). Statistical inference for multivariate partially linear regression models. *Canadian Journal of Statistics*, 41(1), 1-22.
16. Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1), 35-62.