مجلة
كلية التــراث الجامعة

مجلة علمية محكمة
متعددة التخصصات نصف سنوية
العدد الحادي والأربعون
30 نيسان 2025
ISSN 2074-5621

رئيس هيئة التحرير
أ.د. جعفر جابر جواد

مدير التحرير
أ.م. د. حيدر محمود سلمان

رقم الايداع في دار الكتب والوثائق 719 لسنة 2011

مجلة كلية التراث الجامعة معترف بها من قبل وزارة التعليم العالي والبحث العلمي بكتابها المرقم (ب 3059/4) والمؤرخ في (2014/ 4/7)

اختيار أفضل متغيرات انموذج الانحدار الذاتي AR(p) بأستعمال طرق الجزاء

رنـين رافد محـي الديـن أ.د. فـراس احمـد محمـد

جامعة بغداد /كلية الادارة والاقتصاد / قسم الإحصاء

المستخلص

يعتبر سوق العراق للاوراق المالية مصدر تمويل المشاريع الاستثمارية ومخزن ادخار رؤوس الاموال ومؤشرات سوق الاوراق المالية بشكل خاص تهم المستثمرين اضافة الى العديد من الاطراف التي تتعامل في اسواق رأس المال ،وأن مشكلة الأبعاد الكبيرة تُعدّ من المشاكل المهمّة التي تعاني منها البيانات وبيانات موشر السوق لقطاع لمصارف بشكلٍ خاص ، وعملية تحديد المتغيرات التي تؤثر فعلياً بصورة عامة ومتغيرات الأنحدارالذاتي في السلاسل الزمنية بصورة خاصة في البيانات الكبيرة (ذات الابعاد العالية) ، يساعدنا كثيراً في دقة النتائج التحليلية للأساليب والنماذج المستخدمة حيث يقلل من الأخطاء كونه يحقق الهدف المنشود، سيتم التطرق في هذا البحث الى اختيار افضل المتغيرات الموجودة في السلسلة الزمنية بالأضافة الى التقليل من ابعادها عن طريق اختيار المتغيرات التي تمثل النموذج، حيث يعتبر اصعب جانب من ناحية التعامل مع بيانات السلسلة الزمنية هو اختيار النموذج وتحديد متغيراته ومعلماته . وبصورة عامة، تكون الغاية الاساسية هو تحديد المتغيرات المهمة في مجموعة البيانات وتجاهل المتغيرات الاخرى من خلال اعتبارها غير ذات صلة او زائدة عن الحاجة ، حيث يتم ذلك عن طريق استخدام طرق الجزاء والمتمثلة بطريقة الجزاءLasso وطريقة Adaptive Lasso وطريقة Double Adaptive Lasso وتم العمل على بيانات اسبوعية وللفترة من 2017-2021 حسب مامتوفر وبأستخدام برنامج لغة (Python) ومن خلال النتائج التي ظهرت تفوقت طريقة (Double Adaptive Lasso) على نظيرتيها بالأعتماد على اقل قيمة لمعايير المقارنة والمتمثلة بـ (RMSE)

الكلمات المفتاحية : سوق العراق للأوراق المالية ، اختيار افضل المتغيرات ، Lasso ، Double Adaptive Lasso ، Lasso

1- المبحث الاول : منهجية البحث

1-1 المقدمة Introduction

مع التقدم التكنولوجي، أصبح الوصول إلى المعلومات أكثر سهولة في الوقت الحاضر،من خلال جمع وتحليل كميات كبيرة من البيانات خلال فترة زمنية معينة ، وتُعدّ مشكلة الأبعاد الكبيرة من المشاكل المهمّة التي تعاني منها البيانات، حيث أن عملية تحديد المتغيرات التي تؤثر فعلياً بصورة عامة ومتغيرات الأنحدارالذاتي في السلاسل الزمنية بصورة خاصة في البيانات الكبيرة (ذات الابعاد العالية) ، يساعد هذا في دقة النتائج التحليلية للأساليب والنماذج المستخدمة من خلال التقليل من الاخطاء كونه يحقق الهدف المنشود (الوصول الى نموذج اكثر دقة من النموذج الاعتيادي) ، سيتم التطرق في هذه الرسالة الى اختيار افضل المتغيرات المكونة للسلسلة الزمنية بالأضافة الى التقليل من ابعادها عن طريق اختيار المتغيرات التي تمثل الانموذج، حيث يعتبر اصعب جانب من ناحية التعامل مع بيانات السلسلة الزمنية هو اختيار النموذج وتحديد متغيراته ومعلماته ، حيث اجتذب موضوع اختيار المتغيرات في المساحات عالية الابعاد (غالبا ما تكون بالمئات أو الالاف من الابعاد) اهتماماً في السنوات السابقة، وهو شائع في العديد من المشكلات الحقيقية ، إذ أن عملية اختيار مجموعة فرعية مثالية من البيانات او المتغيرات وفقً لمعيار محدد وميزات وخصائص معينة من مجموعة من البيانات يحب ان تكون وفقاً للغرض من اختيار الخاصية او الميزة ، حيث تهدف هذه العملية الى تحسين دقة التنبؤ لأستخراج البيانات المتسخدمة . وبصورة عامة ، تكون الغاية الاساسية هو تحديد الميزات المهمة في مجموعة البيانات وتجاهل الميزات الاخرى من خلال اعتبارها غير ذات صلة او زائدة عن الحاجة.

أن الحل المناسب لتسهيل عملية تحليل ودراسة بيانات كهذه هو استعمال مجموعة من طرائق الجراء التي تساعدنا على الاختيار منها طريقة Least Absolute Shrinkage and Selection Operator أقل انكماش مطلق وعامل التحديد والتي يشار اليها بصورة مختصرة بالرمز بـ (LASSO) ، اقترحت هذه الطريقة عام 1996 من قبل العالم (Tibshirani) حيث تعتبر من اكثرالطرق التي يتم استخدامها في اختيار المتغيرات وتقدير النموذج للبيانات ذات الابعاد العالية وكطريقة جديدة للتقدير في النماذج الخطية، حيث يعتبر هذا الاسلوب واسع الانتشار نسبياً، اذ تم تطبيقه في العديد من الابحاث البيولوجية والطبية لأكتشاف الارتباطات المحتملة بين عوامل الخطر والامراض ذات الصلة ، فضلاً عن تحسين التنبؤ والتحقق من صحة النتائج .

في هذا البحث سيتم التطرق إلى طرق تحليل البيانات عالية الأبعاد والنماذج المعقدة، باستعمال طريقة **(lasso)** وطريقة **(adaptive lasso)** وطريقة **Double adaptive lasso**) وسيتم المقارنة بين الطريق الثلاثة واستعراض اهم ميزات استخدام كلً منهم في تحليل البيانات عالية الابعاد.

2-1 مشكلة البحث Problem of the Research

أن مشكلة البيانات عالية الابعاد موجودة في مختلف المجالات ومنها في مجال الإحصاء والتعلم الآلي وتحليل البيانات، وأن في كثير من الاحيان تعاني نماذج السلاسل الزمنية من الافراط في عدد المراتب التي من الممكن ان تكون بعضها غير معنوية مما يتطلب التركيز على المعالم المعنوية واستبعاد وازاحة المعالم الغير معنوية .

3-1 هدف البحث Objective of the Research

الوصول الى نموذج سلسلة زمنية اكثر كفاءة من النموذج الاعتيادي من خلال استبعاد المتغيرات التي ليس لها تأثير بالنموذج ، وهذا يُحسن النماذج الإحصائية، التي يمكن رؤيتها ليس فقط في التنبؤ ولكن أيضًا في تفسير النموذج وتقليل العمليات الحسابية .كما يلعب إختيار المتغيرات دوراً مهمًا في النمذجة الإحصائية عندما يكون هناك عدد معين من المتغيرات ذات الأهمية في النموذج من بين مجموعة كبيرة من متغيرات المرتده زمنياً .

المبحث الثاني : الجانب النظري :

1-2 نموذج الانحدار الذاتي The Autoregressive Model (AR)

نموذج الانحدار الذاتي هو احدى النماذج الرياضية المهمة التي تسعى الى تحقيق افضل تمثل للبيانات، ويُعبر هذا النموذج عن القيمة الحالية للسلسلة y_t بدلالة القيم السابقة للسلسلة $(y_{t-1}, y_{t-2}, \dots, y_{t-p},)$ والقيمة الحالية للخطأ ε_t [4][7]، حيث يستعمل هذا النموذج في العديد من المجالات ، منها تفسير ظاهره معينة سواءً كانت اقتصادية او طبيعية. ان الهدف الرئيسي من تحليلنا لنماذج السلاسل الزمنية هو امكانية الوصول الى نموذج رياضي يعبرعن البيانات ، وان نموذج الانحدار الذاتي (AR) من ضمن النماذج حين تصبح قيمة السلسلة الحالية (دالة) في قيمتها السابقة بإلاضافة الى بعض الاخطاء، فأن النماذج التي تكونت خلال هذه المرحلة تسمى نماذج انحدار ذاتي، فأذ كانت y_t تعبرعن قيمة السلسلة الحالية و$(y_{t-1}, y_{t-2}, \dots, y_{t-p})$ تمثل نفس القيم للسلسلة في فتراتها السابقة ، حيث ان y_t تتأثر بالقيم السابقة لها ويمكننا توضيح هذه العلاقة للسلسة الزمنية المنفصلة $\mathrm{y} = [y_t]_{t=1}^{n}$ لنموذج الانحدار الذاتي للرتبة (p) [3] [16] والذي يعرف بالصيغة الاتية:-

$$\mathrm{Y_t} = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \qquad (1-1)$$

ويمكننا اعادة كتابة المعادلة اعلاه بدلالة الارتداد الخلفي[9]:-

$$(1 - \phi_1\beta_1 - \cdots - \phi_p\beta_p)\mathrm{Y_t} = \varepsilon_t$$

$$\phi_p(\beta)\mathrm{Y_t} = \varepsilon_t \qquad (2-1)$$

حيث ان:-

Y_t : قيمة مشاهدات الظاهرة المدروسة بالزمن t

ϕ_0: الحد الثابت

$(\phi_1, \phi_2, .., \phi_p)$: هي معلمات النموذج

Y_{t-k} : قيم الظاهرة المرتدة زمنيا خلال الزمن t و $k=1,2,3,...,p$

ε_t :الخطأ العشوائي او التشويش الذي يتوزع توزيع طبيعي بمعدل (صفر) وتباين ثابت (σ_ε^2) اذ أن ((ε_t iid.N(0, σ_ε^2) ~

$$E(\varepsilon_t) = 0$$

$$E(\varepsilon_t^2) = \sigma_\varepsilon^2$$

$$E(\varepsilon_t, \varepsilon_j) = 0 \text{ for } t \neq j$$

2-2 أختيار المتغيرات Variables selection

أن موضــوع اختبــار المتغيــرات مــن المواضــيع التــي تمحــور اهتمــام الكثيــر حولهــا فــي وقتنــا الحــالي ، حيــث يــتم اســتخدامها فــي تقليــل عــدد المتغيــرات المســتخدمة فــي التحليــل الاحصــائي وازالــة او أســتبعاد المتغيــرات الغيــر ذات اهميــة والتــي لا تــؤثر بدرجــة كبيــرة علــى النتــائج ،فــي العديــد مــن المجــالات التــي تمتلــك بيانــات كبيــرة نكــون بحاجــة الــى تحديــد عــدد المتغيــرات المعنويــة (المهمــة) حيــث ويســاعدنا ذلــك فــي تقليــل الكلفــة والجهــد وبالأضــافة لتحســين جــودة التحليــل . ومــن بــين هــذه المجــالات (الهندســة الوراثيــة ، الماليــة ، المجــالات الطبيــة ، الزراعــة ، البيئة ، علم الاجتماع) وغيرها من المجالات الاخرى .

تعتبــر عمليــة اختيــار المتغيــر جــرءً مهــم فــي العديــد مــن المجــالات ، حيــث تظهــر اهميتــه متــى مــا تكــون الدالــة الحقيقــة دالــة غيــر خطيــة ومــن المعــروف أن ادخــال عــددً كبيــراً مــن المتغيــرات بمعادلــة الانحــدار لأي دراســة او ظــاهرة يكلــف جهــداً ووقتــاً وثمنــاً ووجــود بعــض المتغيــرات غيــر الاساســية فــي تأثيرهــا المتغيــر المعتمــد أو يكــون تأثيرهــا مشــابهة لتــأثير متغيــرات اخــرى ، أو كــون لهــذه المتغيــرات ارتبــاط عــالي فيمــا بينهــا حيــث يجعــل تأثيرهــا غيــر معنــوي ممــا يــدعونا الى ازالة (استبعاد) مثل هذه المتغيرات

هنــاك نوعــان اساســيان مــن طرائــق اختبــار المتغيــرات منهــا مــا تعــرف بأســم الطرائــق التقليديــة والتــي تتضــمن طريقــة الحــذف العكســي حيــث تبــدأ هــذه الطريقــة بأعتمــاد جميــع المتغيــرات فــي معادلــة الأنحــدار ومــن ثــم يــتم تباعــاً حــذف تلــك المتغيــرات بأســتخدام قيمــة F الجدوليــة ، أمــا الطريقــة الثانيــة فتعــرف بطريقــة الاختبــار الامــامي المباشــر حيــث يــتم مــن خــلال هــذه الطريقــة ادخــال المتغيــرات واحــداً تلــو الاخــر الــى معادلــة الانحــدار ، أمــا الطريقــة الثالثــة فتعــرف بطريقــة الاختبــار التــدريجي ، بالاضــافة الــى معيــارين مهمــين همــا (معيــار اكــايكي المعلومــات AIC ، معيــار بيــز للمعلومــات BIC) وهــي عــادة مــا تســتخدم فــي السلاســل الزمنيــة فــي تحديــد افضل رتبة او افضل متغيرات توضيحية

و بالنســبة للنــوع الثــاني مــن طرائــق اختبــار المتغيــر فتعــرف بطرائــق التنظــيم أذ تقــوم هــذه الطرائــق بفــرض حــداً للجــراء علــى المعلمــات ليــتم تســهيل عمليــة تقــديرات المعلمــات وتقلــيص المتغيــرات ومــن هــذه الطــرق هــي طريقــة لاســو وطريقــة لاســو المكيــف ولاســو المكيــف المزدوج وغيرها من الطرق .

3-2 دالة الجزاء (Penalty Function)

قـام العـالم (Tibshirani) عـام 1996 بـأقتراح طريقـة تقلـيص عرفـت بأسـم (اقـل انكمـاش مطلـق لأختيـار العامـــــل) [12] ومصـــــــطلح (LASSO) يرمـــــز الـــــى الحـــــروف الاولـــــى لمفهـــــوم " Least Absolute Shrinkage and Selection Operator " تقـوم هـذه الطريقـة بمنـع حـدوث مشـكلة التحيـز الـى الاعلـى (Over fitting) أي أن النمـوذج مـع العديـد مـن المتغيـرات الداخلـة تكـون دون المسـتوى الامثـل أي (غيـر مهمـة) عنـدما يكـون النمـوذج الصـحيح مبعثـر (sparse) بمعنـى أن المتغيـر المعتمـد فقـط يعتمـد علـى عـددً قليـلاً مـن المتغيـرات الداخلـة أي (المعنويـة) ، أن دالـة الجـزاء تعتمـد علـى معلمـة الجـزاء (∝) وعـن طريـق هـذه المعلمـة يـتم وضـع حـد جـزاء والـذي مـن خلالـه يـتم الحصـول علـى عـدد كبيـر مـن المعلمـات و بالنسـبة للمعلمـات الاقـل مـن قيمـة معلمـة الجـزاء تكـون قيمتهـا مسـاوية الـى الصفر عندها يكون اختيار النموذج المناسب تلقائياً [6][10]

4-2 طريقة لاسو [4] [11] Lasso Method

لاسو هي تقنية جذابة للتنظيم واختيار المتغيرات للبيانات عالية الابعاد [8] ، وهي احدى طرق التنظيم التي تعرف بأنها (اضافة دالة جراء للتحكم بتعقيد النموذج ، حيث يعبر عنه بـ (penalty) وأن عمل هذا الجزء يكون بتقليل تباين النموذج مع الزيادة الكبيرة في التحيز) ، تلقت هذه الطريقة الكثير من الأهتمام للدراسات النظرية والتطبيقات الواقعية نظراً لقدرتها على إجراء تقدير المعلمة واختيار النموذج في وقت واحد ، كما انها تعتبر من الطرق الفعالة حيث تعمل هذه الطريقة على اختيار المتغيرات اضافة لتقدير المعالم ، بالأضافة الى انها تعمل بمبدأ تصغير مجموع مربعات الخطأ تبعاً للقيد الذي يمثل (المجموع المطلق للمعلمات) [5] ، كما أن دالة الجزاء(Lasso) تقوم بجعل بعض المعالم مساوية الى الصفر في حال كانت معلمة الجزاء كبيرة ليتم تقليصها تباعاً لمقدار معين بالأعتماد على معلمة الجزاء (∝) لغرض التحكم بمقدار التقليص(الانكماش) (shrinkage) [17]

ويمكن كتابته بالشكل الاتي:-

$$\widehat{\emptyset}_{lasso} = argmin_{\emptyset}\left\{\sum_{t=p+1}^{n}(y_t - \sum_{j=1}^{p}\emptyset_j\, y_{t-j})^2 + \propto \sum_{j=1}^{p}|\emptyset_j|\right\} \qquad (3-1)$$

∝ : معلمة الجزاء

$\propto \sum_{j=1}^{p}|\emptyset_j|$: دالة الجزاء

أن مقدر لاسو يعمل على جعل قيمة المعلمات الغير معنوية مساوية الى الصفر ويعمل على انكماش او تقليص باقي المعلمات بمقدار معين ، أضافة الى تقدير معالم النموذج ويقوم بأختيار المتغيرات الداخلة في النموذج لزيادة القدرة التفسيرية لنموذج الانحدار الذاتي الذي تم استخدامه في تفسير الظاهرة المدروسة .

5-2 طريقة لاسو التكيفي Adaptive Lasso

اللاسو التكيفي هو تطوير لطريقة لاسو LASSO التي تُستخدم في اختيار المتغيرات وتنظيم النماذج الاحصائية ، لقد تطرقت العديد من الدراسات الى تحسين اداء LASSO ، وفي عام 2006 اقترح العالم (**Zou**) [18] نسخة جديدة من LASSO والمعروفة بأسم لاسو التكيفي LASSO Adaptive وكالأتي:-

$$\widehat{\emptyset}_{lasso(1)} = \arg\min_{\emptyset\in R^P}\left\{\sum_{t=p+1}^{n}(y_t - \sum_{j=1}^{p}\emptyset_j\, y_{t-j})^2 + \propto \sum_{j=1}^{p} w_j|\emptyset_j|\right\} \qquad (4-1)$$

أذ أن

$\propto \sum_{j=1}^{p} w_j|\emptyset_j|$: تمثل دالة الجزاء مع معلمة الجزاء $\propto$

حيث $w_j \geq 0$ و $w_j \in w$ و $j=1,\ldots,p$

w: يعرف بالاوزان المكيفة (Adaptive weight) والتي تعتمد على تقديرات اولية لمعاملات النموذج وهذه الاوزان تسمح للعقوبة بأن تكون اكثر مرونة حيث يؤدي ذلك الى تحسين أداء النموذج في اختيار المتغيرات وتقليل التحيز (bias) في التقديرات ، كما بين العالم (**Zou**) أنه اذا تم اختيار الاوزان (w_j) بكفاءة وبطريقة يعتمد فيها على البيانات فأن طريقة لاسو التكيفية يمكن أن تحقق ما يسمى بـ (*oracle properties*) وخاصية الكفاءة (*Efficient*) بحيث ينفذ بطريقة كما لو كان النموذج الصحيح معروف ، وتم اقتراح استعمال الاوزان المقدرة بالصيغة الاتية:-

$$w_j = \frac{1}{|\widehat{\emptyset}_j|^{\gamma}} \qquad (5-1)$$

حيث أن :-

$\widehat{\emptyset}_j$ تمثل تقديرات (OLS) وهي مقدر اولي متسق لـ $\emptyset$ [13]

γ : قيمة معلمة التحكم في قوة العقوبة وقيمتها اكبر من الصفر بمعنى $\gamma > 0$ وعادة ما تكون $\gamma = 1$ او $\gamma = 2$

أن أختيار الاوزان w_j مهم جداً لذا قام العالم (**Zou**) بأستعمال تقديرات المربعات الصغرى الاعتيادية لحسابها ($\widehat{\emptyset}_{j\,ols}$)

6-2 طريقة لاسو التكيفي المزدوج Double Adaptive Lasso

لاسو التكيفي المزدوج هو امتداد لطريقة اللاسو التكيفي Adaptive Lasso حيث يتم تحسين عملية اختيار المتغيرات وتقدير المعاملات في النماذج الاحصائية عن طريق استخدام (اوزان تكيفية مزدوجة) حيث تعزز هذه الطريقة من قدرة النموذج على التعامل مع البيانات المعقدة وتحسين دقة التقديرات خاصة في نماذج الانحدار الذاتي مثل AR(p) [14] ، أن الفكرة الاساسية لأستخدام هذه الطريقة هي تقدير المعاملات $\phi_1, \phi_2, \ldots, \phi_p$ المذكورة في المعادلة (1-1) بدقة ، اضافة الى تقليل عدد المعاملات غير المعنوية (اجبار بعضها على الصفر)

نبدأ بالحصول على تقديرات اولية لمعاملات الانموذج $\widehat{\emptyset}_1, \widehat{\emptyset}_2, \ldots, \widehat{\emptyset}_p$ بأستخدام طريقة المربعات الصغرى الاعتيادية (OLS) ليتم استخدامها في حساب الاوزان التكيفية الاولى $w_j^{(1)}$ لكل معامل $\emptyset_j$ وكالاتي:-

$$w_j^{(1)} = \frac{1}{|\widehat{\emptyset}_j|^{\gamma}} \qquad (6-1)$$

نقوم بتقدير المعاملات بأستخدام لاسو المكيف Adaptive Lasso حيث نضيف عقوبة تكيفية الى دالة الجراء:

$$\widehat{\emptyset}_{lasso(1)} = \arg\min_{\emptyset \in R^p} \left\{ \sum_{t=p+1}^{n} (y_t - \sum_{j=1}^{p} \emptyset_j\, y_{t-j})^2 + \propto \sum_{j=1}^{p} w_j^{(1)} |\emptyset_j| \right\} \qquad (7-1)$$

بعد الحصول على تقديرات جديدة $\widehat{\emptyset}_j^{(1)}$ من المعادلة (1ـــ7) نقوم بحساب الاوزان التكيفية الثانية $w_j^{(2)}$وكالاتي:-

$$w_j^{(2)} = \frac{1}{\left|\widehat{\emptyset}_j^{(1)}\right|^{\gamma}} \qquad (8-1)$$

نقوم بعدها بتطبيق عقوبة مزدوجة عن طريق استخدام الاوزان التكيفية الثانية في دالة الجزاء المعدلة وكالاتي:-

$$\widehat{\emptyset}_{lasso(2)} = \arg\min_{\emptyset \in R^P} \left\{ \sum_{t=p+1}^{n} (y_t - \sum_{j=1}^{p} \emptyset_j \, y_{t-j})^2 + \propto \sum_{j=1}^{p} w_j^{(2)} |\emptyset_j| \right\} \qquad (9-1)$$

ليتم بعدها تقدير المعاملات النهائية $\phi_1, \phi_2, \ldots, \phi_p$ عن طريق تقليل دالة الجزاء المعدلة $\widehat{\emptyset}_{lasso(2)}$.

المبحث الثالث : الجانب التطبيقي

بالأعتماد على برنامج لغة python تم تطبيق طرق اختيار المتغيرات (طريقة لاســو و طريقة لاســو التكيفي وطريقة لاســو التكيفي المزدوج) على بيانات حقيقة متمثلة بمؤشــر الســوق الاســبوعي لقطاع المصــارف، تم الحصــول عليها من [1] والمأخوذه من الموقع الرسـمي لسـوق العراق للأوراق المالية [2] للمدة من 2017-2021 حسب توافر البيانات وبواقع (260) مشاهدة.

1-3 تحليل البيانات باستخدام انموذج الانحدار الذاتي AR(P):

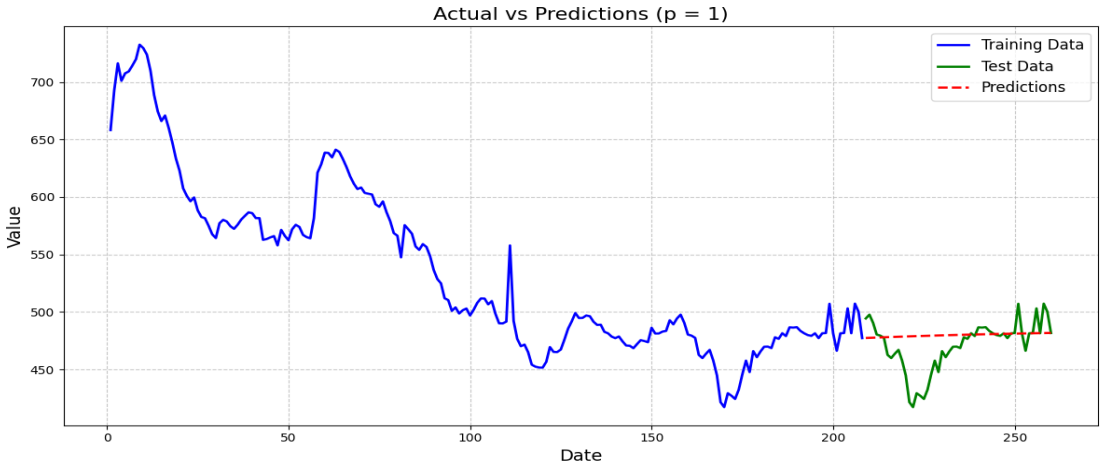
بعد أن قمنا برسم السلسلة الزمنية وظهرت عدم استقراريتها في الوسط الحسابي تم اجراء اختبار الاستقرارية والمتمثل باختبار ديكي -فولر للتأكد من عدم الاستقرارية وقمنا بأعادة رسمها بعد أن اخذ الفروق لها ليتبين استقرارها عند الفرق الاول ، سيتم في هذه المرحلة تحليل بيانات السلسلة الزمنية المستقرة بأستخدام انموذج الانحدار الذاتي **AR(p)** لغرض تحديد افضل رتبة (**p**) من خلال أستخدام المعايير الثلاثة HQIC ,AIC,BIC حيث ظهرت النتائج الموضحة في الجدول ادناه وكما يلي:-

p	AIC	BIC	HQIC
1	1597.869	1607.867	1601.912
2	1581.418	1594.729	1586.801
3	1568.643	1585.258	1575.363
4	1563.338	1583.247	1571.392
5	1557.093	1580.285	1566.476
6	1548.771	1575.237	1559.48
7	1543.822	1573.551	1555.852
8	1534.257	1567.241	1547.605
9	1526.365	1562.592	1541.027
10	1521.36	1560.819	1537.331
15	1492.054	1547.52	1514.516
20	1466.273	1537.475	1495.122
25	1437.181	1523.837	1472.307
30	1410.091	1511.908	1451.381
35	1375.87	1492.541	1423.203
40	1344.415	1475.621	1397.665
45	1317.592	1462.999	1376.626
50	1289.379	1448.634	1354.054
55	1262.225	1434.96	1332.393
60	1210.677	1396.504	1286.178

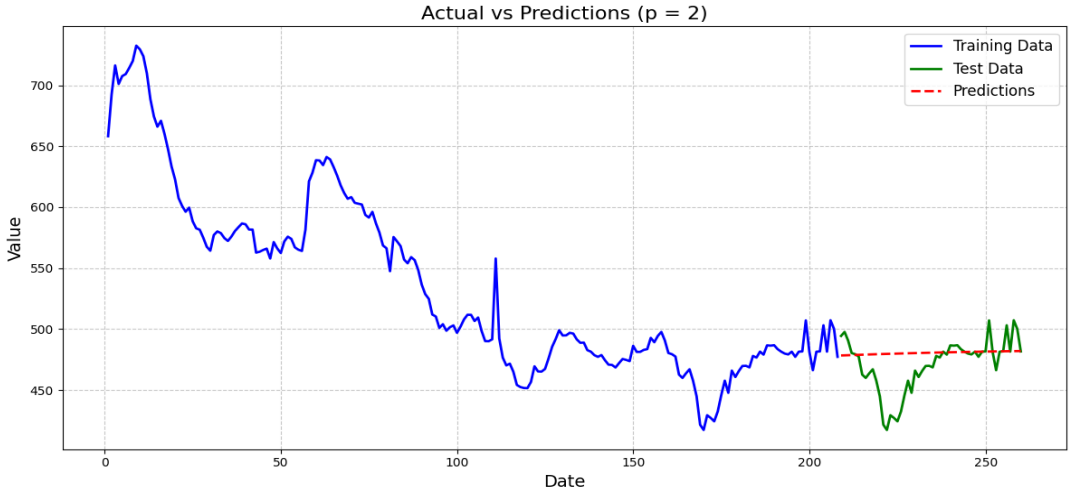
65	1181.604	1380.115	1262.269

الجدول(3-1) يوضح عدد الارتدادات الزمنية (**p**) وقيم معياري AIC,BIC, HQIC

مــن خــلال الجــدول (3-1) يتضــح أن الرتبــة الصــحيحة تقابــل أقــل قيمــة لمعايير التشــخيص ومــن ملاحظتنــا لنتــائج الجــدول اعــلاه فــأن أقــل قيمــة للمعــايير الثلاثــة ظهــرت عنــدما كانــت قيمــة **p =65** حيــث كلمــا زادت قيمــة الرتبــة قلــت قيمــة المعــايير الثلاثــة، وبســبب كثــرة الارتدادات الزمنية التي تم العمل عليها سيتم استعراض جزء من النتائج والرسوم

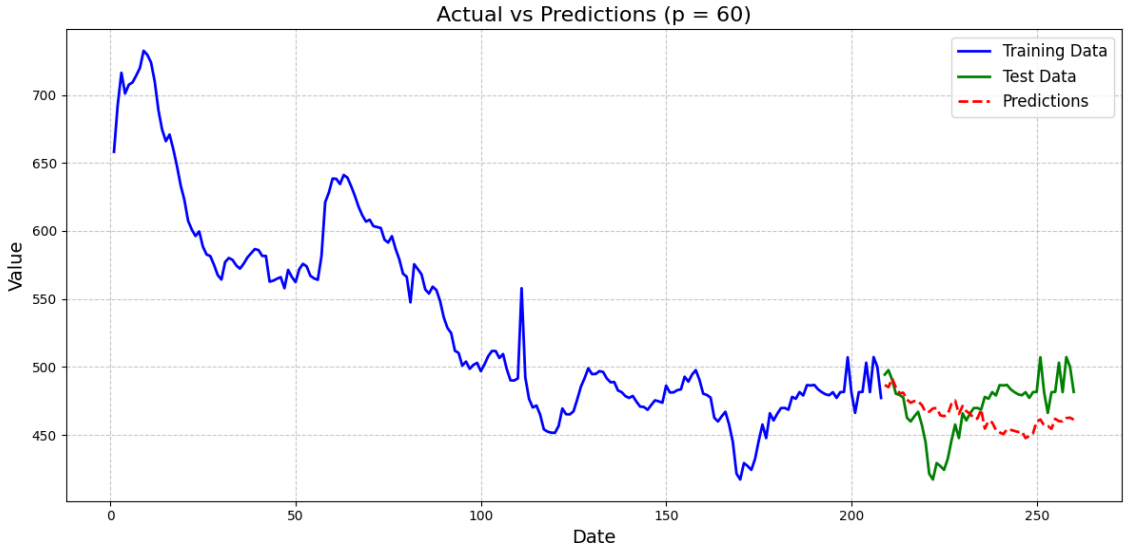


الشكل (3-1) يوضح القيم الفعلية والقيم التنبؤية عند (p=1)

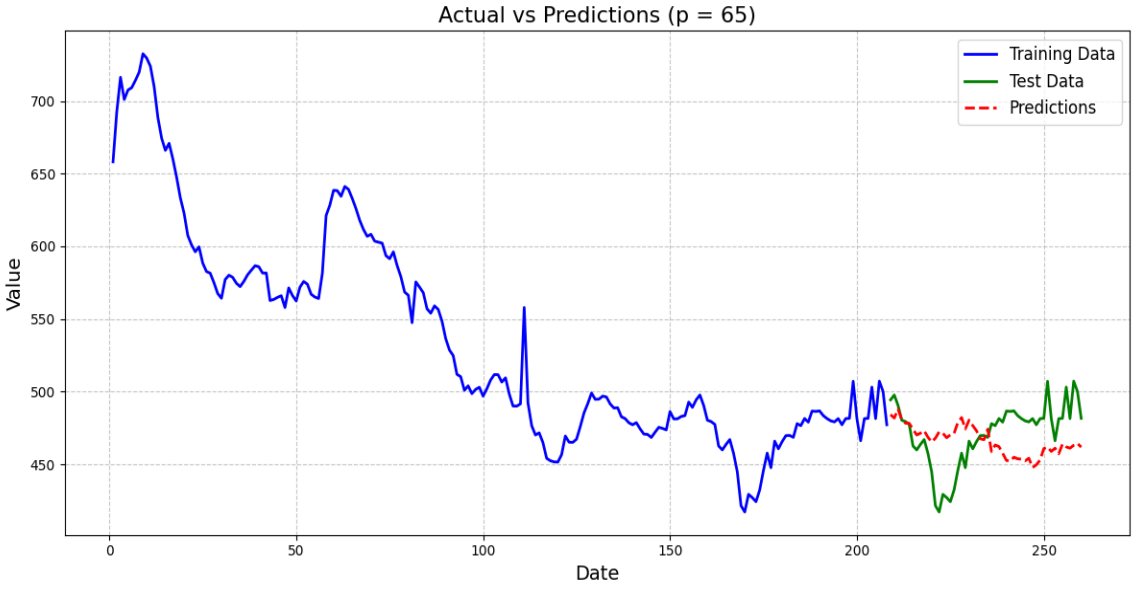


الشكل (3-2) يوضح القيم الفعلية والقيم التنبؤية عند (p=2)

لغرض توضح الفرق بين قيم الرتب المأخوذه تم اضافة عندما كانت قيمة (p=1) و(p=2) حيث يتبين من خلال الشكل (3-1) و(3-2) أن القيم التنبؤية في الشكلين اعلاه لم تتأثر بالتغييرات الحاصلة حيث ظهرت وكانها تسلك سلوك المتوسط ، اما عندما قمنا بتحسين قيمة الرتب وصولا الى الرتبة (p=60) و(p=65) فأن القيم التنبؤية بدأت تتحسس للتغيرات وكما موضحة بالشكل (3-3) و(3-4) ، حيث يدل ذلك على أنه كلما ازدادت قيمة الرتبة ازداد التذبذب بالقيم التبؤية.



الشكل (3-3) يوضح القيم الفعلية والقيم التنبؤية عند (p=60)



الشكل (3-4) يوضح القيم الفعلية والقيم التنبؤية عند (p=65)

2-3 تحديد المعالم المعنوية الافضل بأستخدام طرق Lasso

1-2-3 المعالجة المسبقة للبيانات (Data Pre-Processing):

بعد أن تم تهيئة البيانات وتحديد المتغير التابع (مؤشر السوق) والمتغيرات المستقلة (عدد الاتدادات) و أستخراج الارتدادات بأستخدام انموذج الانحدار الذاتي AR(p) تأتي مرحلة المعالجة المسبقة للبيانات التي بدورها تنقسم الى مرحلتين،الاولى (تقسيم البيانات (Data Splitting)) والثانية (تقييس البيانات (Features Scaling)) اذ تتم في المرحلة الاولى تقسيم البيانات الى مجموعتين تدريب (Training) واختبار(Testing) حيث (80%) تدريب و(20%) اختبار [15] كما موضح بالشكل (3-4) عندما كانت

قيمة (p=65) ، ومن ثم يتم أجراء تقييس للبيانات (Features Scaling) باستخدام طريقة (standard scalar) حيث تقوم بتوحيد البيانات بحيث يكون المتوسط $\mu = 0$ وانحراف معياري $\sigma = 1$ والصيغة الرياضية له كالتالي:-

$$X_{scaled} = \frac{X - \mu}{\sigma}$$

ليتم تهيئتها لغرض ادخالها الى النماذج المستخدمة.

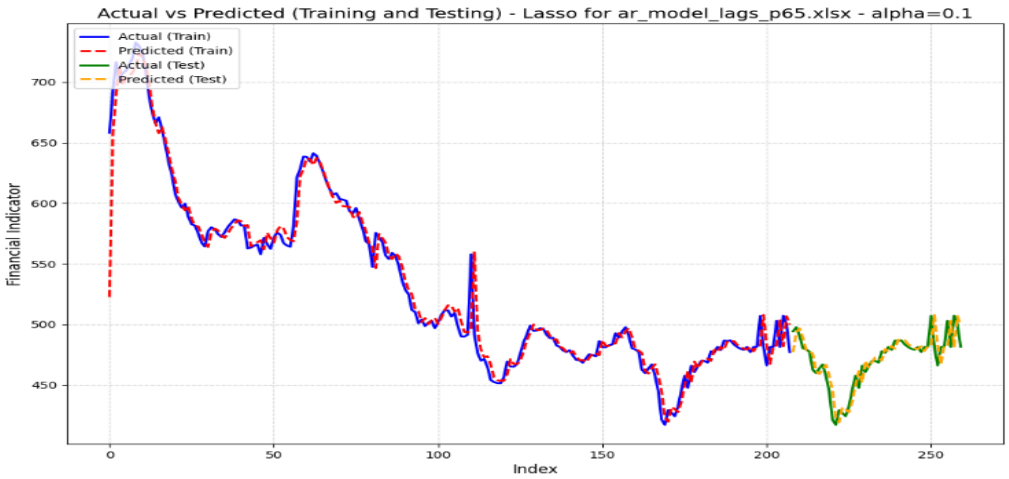
2-2-3 انموذج لاسوLasso :

بعد تهيئة البيانات واستخراج الارتدادات واجراء المعالجة المسبقة ، تم بناء انموذج لاسو لغرض تحديد اي المتغيرات التي لها تأثير على دقة الانموذج وايها التي ليس لها تأثير، وبأستخدام عدة قيم من الفا (∝) (0.001،0.01،0.05،0.1،0.5) ظهرت النتائج الموضحة بالجدول الاتي:-

	∝	RMSE
lags_p=65	0.001	10.96413316
	0.01	10.96395921
	0.05	10.95398892
	0.1	10.93315823
	0.5	11.01855339

الجدول (3-2) يوضح بعض مقاييس الدقة لكل قيمة من قيم الفا لأنموذج لاسو

من نتائج الجدول (3-2) يتضح أن اقل قيمة لمقايس الدقة ظهرت عندما كانت قيمة ∝= 0.1 عندها سنقوم برسم القيم الفعلية والقيم التنبؤية عند (p=65) و (∝= 0.1) وكالأتي:



الشكل (3-5) يوضح القيم الفعلية والقيم التنبؤية عند (p=65) و (∝= 0.1)

من ملاحظتنا للرسم المبين في الشكل (3-5) يتبين بأن تأثر بالتقليص الحاصل واضح جدا بأستخدام طريقة لاسو حيث تم استبعاد عدداً من المتغيرات وكما مبينة في الجدول (3-3) وكالاتي:

الارتدادات الغير معنوية	∝	

Lags p=65	0.001	[]
	0.01	['lag_46']
	0.05	['lag_6', 'lag_33', 'lag_42']
	0.1	['lag_3', 'lag_5', 'lag_8', 'lag_11', 'lag_17', 'lag_18', 'lag_19', 'lag_24', 'lag_32', 'lag_35', 'lag_39', 'lag_41', 'lag_49', 'lag_52', 'lag_53', 'lag_60', 'lag_61']
	0.5	['lag_14', 'lag_24', 'lag_25', 'lag_26', 'lag_34', 'lag_35', 'lag_36', 'lag_37', 'lag_38', 'lag_42', 'lag_46', 'lag_47', 'lag_48', 'lag_49', 'lag_50', 'lag_53', 'lag_56', 'lag_57', 'lag_58', 'lag_61', 'lag_65']

الجدول (3-3) يوضح الارتدادات الزمنية الغير معنوية لأنموذج لاسو لكل قيمة من قيم (∝) عندما p=65

من الجدول (3-3) اعلاه نلاحظ عندما كانت قيمة (∝= 0.1) فأن طريقة لاسو استبعدت المتغيرات

['lag_3', 'lag_5', 'lag_8', 'lag_11', 'lag_17', 'lag_18', 'lag_19', 'lag_24', 'lag_32', 'lag_35', 'lag_39', 'lag_41', 'lag_49', 'lag_52', 'lag_53', 'lag_60', 'lag_61']

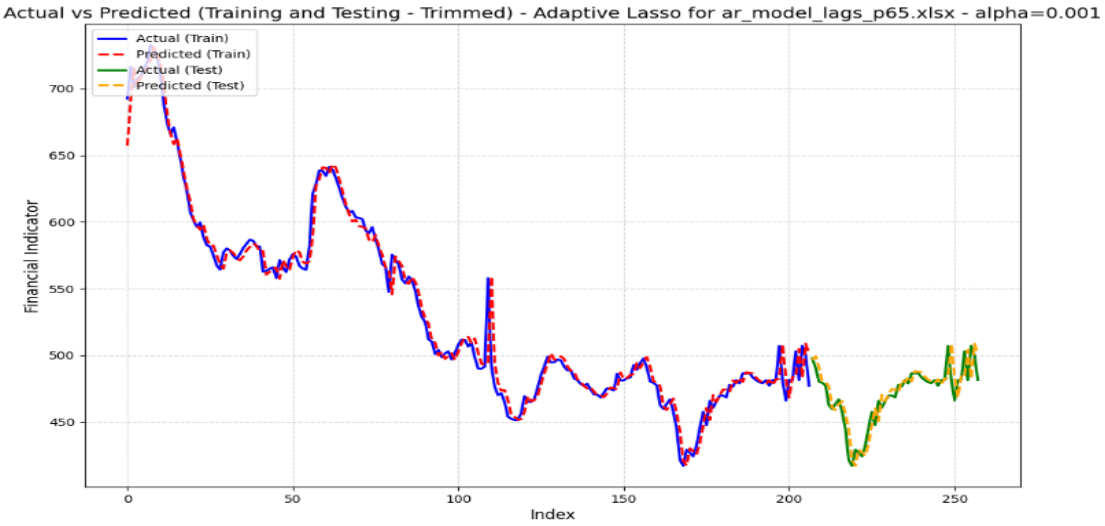
3-2-3 انموذج لاسو التكيفي Adaptive lasso :

سنقوم في هذه المرحلة ببناء انموذج لاسو التكيفي عندما كانت قيمة p=65 ولنفس القيم من ∝ظهرت النتائج كما موضحة بالجدول الاتي:-

	∝	RMSE
lags_p=65	0.001	12.47898
	0.01	15.1645
	0.05	14.13161
	0.1	15.17139
	0.5	14.7402

الجدول (3-4) يوضح مقاييس الدقة لكل قيمة من قيم الفا لأنموذج لاسو المكيف

من نتائج الجدول (3-4) يتضح أن اقل قيمة لمقاييس الدقة ظهرت عندما كانت قيمة ∝= 0.001 عندها نقوم برسم القيم الفعلية والقيم التنبؤية عند (p=65) و (∝= 0.001) وكالاتي:-



الشكل (3-6) يوضح القيم الفعلية والقيم التنبؤية عند (p=65) و (∝= 0.001)

مــن ملاحظتنــا للرســم فــي الشــكل (3-6) يتبــين بــأن تــأثر بــالتقليص الحاصــل واضــح جــدا بأســتخدام طريقــة لاســو التكيفــي حيــث تــم اســتبعاد عــدداً مــن المتغيــرات الغيــر معنويــة وكمــا مبينة في الجدول (3-5) وكالاتي:

	∝	الارتدادات الغير معنوية
lags_p=65	0.001	['lag_8', 'lag_14', 'lag_23', 'lag_29', 'lag_33', 'lag_34', 'lag_37', 'lag_46', 'lag_47', 'lag_49', 'lag_51', 'lag_57', 'lag_61', 'lag_64']
	0.01	['lag_4', 'lag_7', 'lag_14', 'lag_25', 'lag_29', 'lag_31', 'lag_35', 'lag_37', 'lag_46', 'lag_49', 'lag_52', 'lag_56', 'lag_61', 'lag_63']
	0.05	['lag_12', 'lag_13', 'lag_21', 'lag_31', 'lag_33', 'lag_38', 'lag_41', 'lag_46', 'lag_48', 'lag_49', 'lag_52', 'lag_53', 'lag_54', 'lag_56', 'lag_59', 'lag_62']
	0.1	['lag_6', 'lag_7', 'lag_21', 'lag_28', 'lag_29', 'lag_30', 'lag_31', 'lag_32', 'lag_36', 'lag_41', 'lag_42', 'lag_44', 'lag_46', 'lag_49', 'lag_52', 'lag_53', 'lag_56', 'lag_58', 'lag_61', 'lag_64']
	0.5	['lag_1', 'lag_3', 'lag_4', 'lag_6', 'lag_7', 'lag_8', 'lag_9', 'lag_10', 'lag_11', 'lag_12', 'lag_13', 'lag_14', 'lag_15', 'lag_17', 'lag_20', 'lag_22', 'lag_24', 'lag_25', 'lag_26', 'lag_27', 'lag_28', 'lag_29', 'lag_30', 'lag_31', 'lag_32', 'lag_34', 'lag_35', 'lag_36', 'lag_39', 'lag_42', 'lag_43', 'lag_44', 'lag_45', 'lag_46', 'lag_47', 'lag_48', 'lag_49', 'lag_50', 'lag_51', 'lag_52', 'lag_53', 'lag_54', 'lag_55', 'lag_56', 'lag_58', 'lag_59', 'lag_60', 'lag_61', 'lag_62', 'lag_63', 'lag_64', 'lag_65']

الجدول (3-5) يوضح الارتدادات الزمنية الغير معنوية لأنموذج لاسوالمكيف لكل قيمة من قيم (∝) عندما p=65

من الجدول (3-5) اعلاه نلاحظ عندما كانت قيمة (∝= 0.001) فأن طريقة لاسو التكيفي استبعدت المتغيرات ['lag_8', 'lag_14', 'lag_23', 'lag_29', 'lag_33', 'lag_34', 'lag_37', 'lag_46', 'lag_47', 'lag_49', "lag_51', 'lag_57', 'lag_61', 'lag_64']

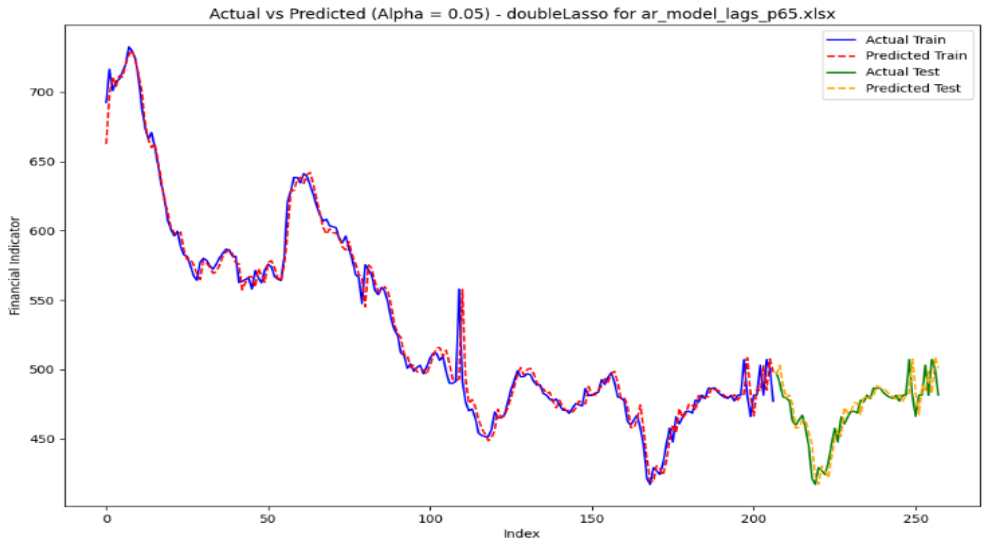
3-2-4 انموذج لاسو التكيفي المزدوج Double Adaptive Lasso

سنقوم في هذه المرحلة ببناء انموذج لاسو التكيفي المزدوج عندما كانت قيمة p=65 ولنفس القيم من ∝ ظهرت النتائج كما موضحة بالجدول الاتي:-

	∝	RMSE
lags_p=65	0.001	11.27795
	0.01	11.11946
	0.05	10.81015
	0.1	10.90179
	0.5	10.9017

الجدول (3-6) يوضح مقاييس الدقة لكل قيمة من قيم الفا لأنموذج لاسو التكيفي المزدوج

من نتائج الجدول (3-6) يتضح أن اقل قيمة لمقايس الدقة ظهرت عندما كانت قيمة ∝= 0.05 عندها نقوم برسم القيم الفعلية والقيم التنبؤية عند (p=65) و (∝= 0.05) وكالاتي:-



الشكل (3-7) يوضح القيم الفعلية والقيم التنبؤية عند (p=65) و (∝= 0.05)

من ملاحظتنا للرسم في الشكل (3-7) يتبين بأن تأثر بالتقليص الحاصل واضح جدا بأستخدام طريقة لاسو المكيف المزدوج حيث تم استبعاد عدداً من المتغيرات الغير معنوية وكما مبينة في الجدول (3-7) وكالاتي:

	∝	الارتدادات الغير معنوية
lags_p=65	0.001	[]

	0.01	['lag_19', 'lag_22', 'lag_26', 'lag_29', 'lag_33', 'lag_46', 'lag_49', 'lag_51']
	0.05	['lag_5', 'lag_10', 'lag_12', 'lag_13', 'lag_17', 'lag_18', 'lag_19', 'lag_20', 'lag_22', 'lag_24', 'lag_26', 'lag_27', 'lag_29', 'lag_33', 'lag_34', 'lag_36', 'lag_37', 'lag_39', 'lag_43', 'lag_45', 'lag_46', 'lag_49', 'lag_51', 'lag_53', 'lag_55', 'lag_57']
	0.1	['lag_2', 'lag_5', 'lag_6', 'lag_8', 'lag_12', 'lag_13', 'lag_15', 'lag_16', 'lag_17', 'lag_18', 'lag_19', 'lag_20', 'lag_22', 'lag_23', 'lag_24', 'lag_26', 'lag_27', 'lag_28', 'lag_29', 'lag_31', 'lag_33', 'lag_34', 'lag_36', 'lag_37', 'lag_38', 'lag_39', 'lag_40', 'lag_41', 'lag_42', 'lag_43', 'lag_45', 'lag_46', 'lag_48', 'lag_49', 'lag_50', 'lag_51', 'lag_53', 'lag_54', 'lag_55', 'lag_57', 'lag_59', 'lag_60', 'lag_61', 'lag_62', 'lag_63', 'lag_64', 'lag_65']
	0.5	['lag_2', 'lag_3', 'lag_4', 'lag_5', 'lag_6', 'lag_7', 'lag_8', 'lag_10', 'lag_11', 'lag_12', 'lag_13', 'lag_14', 'lag_15', 'lag_16', 'lag_17', 'lag_18', 'lag_19', 'lag_20', 'lag_21', 'lag_22', 'lag_23', 'lag_24', 'lag_25', 'lag_26', 'lag_27', 'lag_28', 'lag_29', 'lag_30', 'lag_31', 'lag_32', 'lag_33', 'lag_34', 'lag_35', 'lag_36', 'lag_37', 'lag_38', 'lag_39', 'lag_40', 'lag_41', 'lag_42', 'lag_43', 'lag_45', 'lag_46', 'lag_48', 'lag_49', 'lag_50', 'lag_51', 'lag_52', 'lag_53', 'lag_54', 'lag_55', 'lag_56', 'lag_57', 'lag_58', 'lag_59', 'lag_60', 'lag_62', 'lag_63', 'lag_64', 'lag_65']

الجدول (3-7) يوضح الارتدادات الزمنية الغير معنوية لأنموذج لاسو المكيف المزدوج لكل قيمة من قيم (∝) عندما p=65

من الجدول (3-7) نلاحظ عندما كانت قيمة (∝= 0.05) فأن طريقة لاسو التكيفي المزدوج استبعدت المتغيرات['lag_5', 'lag_10', 'lag_12', 'lag_13', 'lag_17', 'lag_18', 'lag_19', 'lag_20', 'lag_22', 'lag_24', 'lag_26', 'lag_27', 'lag_29', 'lag_33', 'lag_34', 'lag_36', 'lag_37', "lag_39', 'lag_43', 'lag_45', 'lag_46', 'lag_49', 'lag_51', 'lag_53', 'lag_55', 'lag_57]

3-3 تقييم النماذج (Result Evaluation) :

من خلال استخراجنا الرتبة (**P=65**) وبعد استخدامنا لطرق الجزاء الثلاثة وبالاعتماد على مقياس الدقة (**RMSE**) المستخدم لقياس دقة كل نموذج ظهرت النتائج الاتية :-

Model	∝	RMSE	Nonzero parameter
Lasso	0.1	10.93315823	48
Adaptive Lasso	0.001	12.47898	51
Double Adaptive Lasso	0.05	10.81015	39

جدول (3-6) معايير الدقة لافضل النماذج التنبؤية

لمعرفة أي النماذج المستخدمة افضل في اختيار المتغيرات ، تم الاعتماد على مقياس الدقة الجذر التربيعي لمتوسط مربعات الخطأ RMSE حيث نلاحظ مايأتي:-

من الجدول (3-6) نلاحظ أنه بأستخدام انموذج (Double Adaptive Lasso) فأنه يعطي نتائج متقاربة للنتائج المستحصل عليها بأنموذج (Lasso) مع وجود افضلية نسبية لأنموذج (Double Adaptive Lasso) حيث بلغت قيمة الجذر التربيعي لمتوسط مربعات الخطأ بأستخدام انموذج (Double Adaptive Lasso) (RMSE=10.81015) بينما ظهرت قيمة الجذر التربيعي لمتوسط مربعات الخطأ بأستخدام انموذج (Lasso)

(RMSE = 10.93315823) أما قيمة الجذر التربيعي لمتوسط مربعات الخطأ بأستخدام انموذج (Adaptive Lasso) فقد بلغت (RMSE = 12.47898) و بلغت قيمة معلمة الجزاء بأستخدام انموذج(Double Adaptive Lasso) (∝=0.05) وقيمة معلمة الجزاء بأستخدام انموذج (Lasso) (∝ =0.1) بينما كانت قيمة معلمة الجزاء بأستخدام انموذج(Adaptive Lasso) (∝=0.001) أما عدد المعاملات غير الصفرية فقد بلغت (39) معامل بأستخدام انموذج (Double Adaptive Lasso) و (48) معامل بأستخدام انموذج (Lasso) وبلغ (51) معامل بأستخدام انموذج (Adaptive Lasso).

3-4الاستنتاجات (Conclusion)

في هذا البحث، تم استخدام ثلاثة من طرق الجزاء والمتمثلة بطريقة الجزاءLasso وطريقة Adaptive Lasso وطريقة Double Adaptive Lasso والتي لها اهمية بالغه في اختيار افضل متغيرات نماذج السلاسل الزمنية حيث تم استخدام طريقة Double Adaptive Lasso المقترحة من قبل (Liu) في عام 2014 حيث تبين أن منهجية لاسو التكيفي المزدوج له دقة حسابية اعلى بكثير نظيرتيه . نستنتج ومن خلال نتائج الجانب التطبيقي أن النتائج المستحصل عليها بأستخدام انموذج (Double Adaptive Lasso) متقاربة للنتائج المستحصل عليها بأنموذج (Lasso) حيث تفوق أنموذج (Double Adaptive Lasso) على نظيرتيه كونه يعطي افضلية نسبية عن طريق استعمال مقياس الدقة (RMSE) اقل عند المقارنة .

المصادر (References):-

1- عبدالحسين , سما حيدر(2022)م " مقارنة بعض تقنيات التعلم العميق للتنبؤ بالمؤشرات المالية " رسالة ماجستير في تقنيات المعلوماتية الادارية

2- الصفحة الرسمية لسوق العراق للأوراق المالية

http://www.isx-iq.net/isxportal/portal/aboutUs.ht

3- Box, G. E. P., Jenkins, G. M., Ljung, G. M. & Reinsel, G. C. (2016). “Time Series Analysis Forecasting and Control”, John Wiley & Sons.

4- Brockwell, P. J., & Davis, R. A. (2009). *Time series: theory and methods*: Springer science & business media

5- Jatherine, Stuart (2011) " Robust Regression", Springer

6- Liu, X. (2011), " penalized variable selection for semiparametric regression models ". submitted in partial fulfillment of the requirements for the degree doctor of philosophy – nuniversity of Rochester , new york .

7- Liu, Zi Zhen, (2014) " The Doubly Adaptive LASSO Methods for Time Series Analysis" Electronic Thesis and Dissertation Repository.

8- Meinshausen, N. and Yu, B. (2009). Lasso-type recovery of sparse representation for high-dimensional data. The Annals of Statistics, 37, 246.27

9- Montgomery, D. C., Jennings, C. L.& Kulahci, M. (2015). “Introduction to time series analysis and forecasting”, John Wiley & Sons, Inc; Second Edition

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

11- Roth, Volker. (2004) "The generalized LASSO." IEEE transactions on neural networks, VOL (15), NO. (1): 16-28.

12- Tibshirani , Robert, (1996) ," Regression Shrinkage and Selection via the Lasso" , J. R. Statist. Soc. B 58, No. 1, pp. 267-288.

13- Wang, T., Xu.P.,and Zhu , L.(2013) ," Penalized Minimum Average Variance Estimation " . statistics sinica 23, pp.543-569.

14- Wang, H., & Leng, C. (2007). Unified LASSO estimation by least squares approximation. *Journal of the American Statistical Association*, *102*(479), 1039-1048.

15- Yue, J., Wei, B.&Rao,Y.(2017)"A deep learning framework for financial time series using stacked autoencoders and long-short term memory" PLoS ONE 12(7):e0180944.

16- Yule, G.U (1927). On a Method of Investigating Periodicities in Disturbed series, With special Reference to wolfer's Sunspot Numbers. *Philosophical*

17- Zhang, C. H. and Zhang, S. (2014) Confidence intervals for low dimensional parameters in high-dimensional linear models. Journal of the Royal Statistical Society, Series B, 76.

18- Zou, H. (2006). The adaptive LASSO and its oracle properties. J. American Statistical Association, 101: 1418-1429