Journal of Business Economics for Applied Research, Vol. (7), No. (Conference): 548-564
Doi: https://dx.doi.org/10.37940/BEJAR.2025.7.1.34

Journal of Business Economics for Applied Research
مجلة اقتصاديات الاعمال للبحوث التطبيقية
ISSN 2709-6475

Using Clustering Techniques in Quantile Regression for Fixed Effect Model
Simulation Study

Haider Zuham Jabor[(1)], Prof. Dr. Lekaa Ali Al-Alawy[(2)]
University of Baghdad[(1)], University of al-shaab[(2)]
(1) haidar.jabr2101p@coadec.uobaghdad.edu.iq (2) lekaa.alalawy@alshaab.edu.iq

Key words:
Panel Data, Quantile Regression, Fixed Effects, Clustering Technique, K-Medoids, K-means.

ARTICLE INFO
Article history:
Avaliabble online | 25 ***May***. 2025

*Corresponding author:
Lekaa Ali Al-Alawy
University of al-shaab

Abstract:
This study explores the use of quantile regression as an analytical approach for panel data with fixed effects, integrating it with clustering techniques to identify latent subgroups (clusters) of units that share similar characteristics or behaviors. To achieve this, two popular clustering algorithms—K-Means and K-Medoids—were employed and compared in terms of their effectiveness in uncovering hidden clusters among units. Quantile regression is a flexible and powerful statistical tool that allows the analysis of relationships between variables at different points of the distribution of the dependent variable, rather than focusing solely on the conditional mean as in traditional regression models. This flexibility enables a deeper understanding of heterogeneity in the data, especially in the presence of non-uniform distributions or outliers.
In this study, a Monte Carlo simulation approach was adopted to evaluate the performance of integrating quantile regression with the aforementioned clustering methods under various scenarios. The Mean Squared Error (MSE) criterion was used to compare model accuracy across different settings, including variations in time periods and proportions of outliers. The results demonstrated that selecting an appropriate clustering technique and properly tuning the number of clusters can significantly improve the accuracy of estimating quantile regression parameters. In particular, the K-Medoids method outperformed K-Means in scenarios with a high proportion of outliers, while the median quantile level (Q50) proved to be the most stable among the quantiles studied.

Journal of Business Economics for Applied Research, Vol. (7), No. (Conference):548-564
Doi: https://dx.doi.org/10.37940/BEJAR.2025.7.1.34

استعمال تقنيات العنقدة في الانحدار التقسيمي لنموذج الأثر الثابت
دراسة محاكاة

الباحث: حيدر زحام جبر — haidar.jabr2101p@coadec.uobaghdad.edu.iq
جامعة بغداد – كلية الإدارة والاقتصاد

أ.د. لقاء علي العلوي — ekaa.alalawy@alshaab.edu.iq
جامعة الشعب

المستخلص

يتناول هذا البحث استخدام الانحدار التقسيمي كأحد أساليب تحليل البيانات الطولية لنماذج الأثر الثابت، ودمجه مع تقنيات العنقدة بهدف تحديد مجموعات فرعية (عناقيد) من الوحدات ذات الخصائص أو السلوكيات المتشابهة، لتحقيق ذلك، تم توظيف خوارزميتي التجميع K-Means و K-Medoids ومقارنة فاعليتهما في الكشف عن هذه العناقيد الكامنة بين الوحدات. يُعد الانحدار التقسيمي أداةً إحصائيةً مرنةً وقويةً تتيح لنا تحليل العلاقة بين المتغيرات عند مواقع مختلفة من توزيع المتغير التابع، بدلًا من الاقتصار على تقدير المتوسط الشرطي كما في الانحدار التقليدي. تُمكّن هذه المرونة من الحصول على فهم أعمق للتباينات في البيانات، خاصةً في حالة التوزيعات غير المتجانسة أو وجود قيم شاذة (Outliers). في هذا البحث، تم اعتماد منهجية المحاكاة Monte Carlo لتقييم أداء الدمج بين الانحدار التقسيمي وطرق التجميع المذكورة تحت ظروف مختلفة. استُخدم معيار متوسط مربعات الخطأ (MSE) لمقارنة دقة النماذج عبر سيناريوهات متنوعة من حيث طول الفترة الزمنية ونسبة القيم الشاذة. أظهرت النتائج أن اختيار تقنية التجميع المناسبة وضبط عدد العناقيد يمكن أن يحسّن بشكل ملحوظ من دقة تقدير معالم النموذج التقسيمي، حيث تفوقت طريقة K-Medoids في حالة وجود نسبة مرتفعة من القيم الشاذة، في حين أثبت الوسيط (الانحدار عند Q50 أنه الأكثر استقرارًا بين الشرائح الكمية المدروسة.

الكلمات المفتاحية: بيانات بانل، الانحدار الكمي، التأثيرات الثابتة، تقنية التجميع، ك-ميدويدز، ك-مينز.

1- مقدمة

شهدت السنوات الأخيرة تزايدًا ملحوظًا في الاهتمام بتحليل البيانات الطولية في العديد من المجالات الأكاديمية كعلم الاقتصاد والعلوم الاجتماعية والبحوث الصحية. تمتاز البيانات الطولية بأنها تجمع بين خصائص البيانات المقطعية والسلاسل الزمنية، إذ تتضمن مشاهدات متعددة لمجموعة من الوحدات (مثل الدول أو الشركات) عبر فترات زمنية مختلفة. يتيح هذا النوع من البيانات تتبع التغيرات الديناميكية في العلاقات بين المتغيرات عبر الزمن وبين الوحدات المختلفة، مما يوفر فهمًا أعمق للتباينات الفردية التي قد لا تكون مرئية في التحليلات التقليدية. على سبيل المثال، قد تختلف استجابات الشركات للتغيرات الاقتصادية بمرور الزمن وبين الصناعات المختلفة، الأمر الذي يجعل التحليل الطولي أكثر ملاءمة لالتقاط هذه الفروقات مقارنةً بالتحليل المقطعي أو الزمني منفردًا.

تعد النماذج المعتمدة على التأثيرات الثابتة من أهم الأساليب المستخدمة لتحليل البيانات الطولية، حيث تساعد هذه النماذج في التقليل من التحيز الناجم عن تأثيرات غير مرصودة تؤثر على المتغير التابع. ومن خلال ذلك، تتيح نماذج التأثيرات الثابتة فهماً دقيقاً للعلاقات بين المتغيرات عبر الزمن وداخل الكيانات المتعددة. ومع تزايد تعقيد البيانات، أصبح من الضروري تطبيق تقنيات تحليلية تأخذ في الاعتبار هذا التفاوت في البيانات بين الكيانات وبين الزمن. من هنا يأتي دور الانحدار التقسيمي كأداة قوية لتحليل العلاقات بين المتغيرات في البيانات الطولية. يُعَد الانحدار التقسيمي أداة إحصائية

مرنة تتيح تحليل العلاقات بين المتغيرات عند مستويات مختلفة من توزيع المتغير التابع. بدلاً من التركيز على المتوسط الشرطي كما يحدث في الانحدار التقليدي، يتيح الانحدار التقسيمي تحليل تأثير المتغيرات المستقلة على مختلف النقاط في توزيع المتغير التابع، مثل الربع الأدنى، الربع الأعلى، أو الوسيط. هذه الميزة تجعل الانحدار التقسيمي مناسبًا لتحليل البيانات ذات التوزيعات غير المتجانسة أو التي تحتوي على نقاط شاذة. على سبيل المثال، في دراسة أثر الدخل على الإنفاق الاستهلاكي، قد نجد أن العلاقة بين الدخل والإنفاق تختلف بين الأشخاص ذوي الدخل المنخفض والأشخاص ذوي الدخل المرتفع، وهو ما يمكن تحليله باستخدام الانحدار التقسيمي. أعمال مثل Koenker وBassett (1978) قدمت أسس الانحدار التقسيمي، حيث ركزت على استخدامه لفهم العلاقات المختلفة عبر التوزيع الشرطي. كما أن Koenker (2004) طبق الانحدار التقسيمي على البيانات الطولية، مما مهد الطريق لاستخدامه في سياقات جديدة تتعلق بالتغيرات الهيكلية عبر الزمن داخل الوحدات.

ومع تطبيق الانحدار التقسيمي على البيانات الطولية، أصبح بالإمكان تحليل العلاقة بين المتغيرات عبر الزمن وبين الكيانات المختلفة. يمكن للنماذج المستخدمة في الانحدار التقسيمي أن تأخذ في الاعتبار التأثيرات الثابتة التي تخص كل كيان (مثل الدولة أو الشركة) وتقدير تأثير هذه المتغيرات عبر نقاط مختلفة من توزيع المتغير التابع. أحد التحديات التي يواجهها التحليل باستخدام الانحدار التقسيمي في البيانات الطولية هو كيفية التعامل مع التأثيرات الثابتة لكل كيان. الحلول الشائعة تشمل استخدام نماذج الانحدار التقسيمي للأثر الثابت، التي تتيح تقدير التأثيرات الفردية لكل كيان على حدة مع الحفاظ على تأثير المتغيرات المستقلة مشتركًا بين جميع الكيانات. هناك العديد من الباحثين والدراسات التي تناولوا هذا الجانب واستخدموه في عدة مجالات اذ يعد أداة متعددة الاستخدامات يمكنها معالجة العديد من التحديات في تحليل البيانات الطولية. على سبيل المثال، قدم (Galvao,A., 2011) نموذجاً للانحدار التقسيمي الديناميكي مع تأثيراته الثابتة، واقترح (Lamarche 2010) طرقاً لمقاومة التأثيرات الشاذة من خلال طرق جزائية في تقدير الانحدار للبيانات الطولية. كما تناول (Kato et al. , 2012) الخصائص الإحصائية الكبيرة لتقديرات الانحدار التقسيمي في ظل التأثيرات الفردية، وطور (Galvao et al., 2013) طرقاً لتقدير نماذج الانحدار التقسيمي المراقب مع تأثيرات ثابتة. لاحقاً اقترح (Galvao & Wang , 2015) مقدرات فعالة بالاعتماد على منهجية المسافة الدنيا. من جهة أخرى، اهتمت أبحاث حديثة بمعالجة حالات خاصة مثل النماذج المكانية (Spatial) والتفاعلات غير الخطية ; فقدم (Dai & Jin , 2021) انحداراً تقسيمياً بعدياً للنماذج المكانية مع تأثيرات ثابتة، بينما ركز (Powell , 2022) على تطوير تقديرات للانحدار التقسيمي مع تأثيرات ثابتة غير إضافية (non-additive). وفي سياق أحدث، طور (Chen,2024) أسلوب التقدير بخطوتين لنماذج الانحدار التقسيمي مع تأثيراته الثابتة وتداخلات زمنية (interactive effects)، كما قدم (Yang et al. , 2024) اطاراً يجمع بين النماذج ذات المعاملات الدالية (functional-coefficient models) والتجميع الكامن في الانحدار التقسيمي للبيانات الطولية ، حيث استخدمن خوارزميات عنقودية لكشف البنية التجميعية الكامنة بين الكيانات.هذه التطورات الحديثة تؤكد الحيوية المستمرة لمجال الانحدار التقسيمي للبيانات الطولية وأهمية دمج التقنيات المختلفة لتعزيز دقة النماذج.

على الرغم من تعدد الدراسات السابقة في مجال الانحدار التقسيمي للبيانات الطولية، هناك فجوة بحثية تتعلق بتحديد المجموعات الفرعية ضمن البيانات الطولية بناءً على أنماط التأثيرات المقدرة. التعرف على هذه المجموعات (العناقيد) يساهم في زيادة مرونة النمذجة وتحسين كفاءة التقدير عن طريق تجميع الوحدات ذات السلوك المتشابه. اقترح (Zhang et al. , 2019) منهجية للتجميع قائمة على الانحدار التقسيمي بهدف تحديد المجموعات الكامنة من الوحدات ذات المعاملات المتجانسة، وبيّنت نتائجهم إمكانية تحسين دقة التقدير من خلال عملية جمع المعلومات داخل كل مجموعة. المنهجية المتبعة في هذه الدراسة تتقاطع مع عمل Zhang وزملائه من حيث المبدأ العام (دمج التجميع مع الانحدار التقسيمي)، لكنها تختلف في اعتماد خوارزميات تجميع بسيطة وشائعة

K-means)و (K-medoidsواختبارها تحت ظروف تتضمن نسب مختلفة من القيم الشاذة لم تُفحص بعمق في الدراسات السابقة. حيث تُعَد خوارزمية k-means واحدة من أكثر طرق العنقدة شيوعًا وسهولة في الاستخدام. تعتمد هذه الطريقة على تقسيم البيانات إلى عدد من المجموعات بحيث يتم تقليل المسافة بين النقاط داخل كل مجموعة. تتميز هذه الطريقة ببساطتها وكفاءتها، لكنها تعاني من الحساسية للقيم الشاذة، مما يجعل استخدامها في بعض الحالات غير مثالي. في حين تعد k-medoids تحسينًا على طريقة k-means، حيث تعتمد على اختيار النقاط المركزية (medoids) من بين نقاط البيانات الفعلية، بدلاً من استخدام متوسط المسافات كما هو الحال في k-means. هذه الطريقة أكثر متانة في التعامل مع البيانات التي تحتوي على نقاط شاذة أو توزيعات غير متماثلة.

بناءً على ما تقدم تهدف دراستنا إلى دمج تقنية الانحدار التقسيمي مع أساليب التجميع لتحديد العناقيد الكامنة بين الوحدات وتحليل سلوكها المختلف عبر الشرائح الكمية، مع التركيز بشكل خاص على تأثير القيم الشاذة على أداء كل من خوارزميتي K-means و K-medoids في هذا السياق.

2- نماذج البيانات الطولية

البيانات الطولية تمثل نوعًا من البيانات التي تجمع بين خصائص البيانات المقطعية والبيانات الزمنية .هذا يعني أن البيانات الطولية تحتوي على بعدين: الأول يمثل الزمن، والثاني يمثل الوحدات المقطعية (مثل الدول أو الشركات)(Muslim, 2009) . يتم الحصول على هذه البيانات من خلال مراقبة ظاهرة معينة عبر عدة وحدات على مدى فترات زمنية متعددة. ويعد هذا النوع من البيانات مهمًا جدًا لأنه يسمح بتحليل العلاقة بين المتغيرات عبر الزمن والوحدات المختلفة بشكل أعمق من البيانات المقطعية أو الزمنية بمفردها (Wooldridge, J. M. 2010).

تكمن أهمية البيانات الطولية في أنها توفر معلومات غنية عن سلوك الأفراد أو الوحدات على مر الزمن، وتأخذ في الاعتبار التباينات بين الوحدات المختلفة. تتمثل بعض المزايا الرئيسية لهذه البيانات في أنها تسمح بالتحكم في التباين الفردي الذي قد يؤدي إلى نتائج منحازة إذا ما تم تجاهله. كما أنها توفر عددًا أكبر من درجات الحرية وتحسن من كفاءة التقديرات الإحصائية، مما يجعلها أداة قوية في دراسة تأثير المتغيرات غير الملحوظة(Baltagi, B. H. ,2008).

نماذج البيانات الطولية تتضمن دمج النماذج المقطعية والزمنية في معادلة واحدة، حيث يمكن تمثيل نموذج البيانات المقطعية بالمعادلة التالية:

البيانات المقطعية بالمعادلة التالية:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \epsilon_i \qquad (1)$$

وتمثيل نموذج السلاسل الزمنية بالمعادلة:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \epsilon_t \qquad (2)$$

أما نموذج البيانات الطولية فيدمج المعادلتين (1) و (2) على النحو التالي:(Abdulrazak, 2012)

$$Y_{it} = \beta_{0(i)} + \sum_{j=1}^{k} \beta_j X_{j(it)} + \varepsilon_{(it)} \qquad (3)$$

حيث ان:

Y_{it} : تمثل متغير الاستجابة للمقطع العرضي i عند الفترة الزمنية t.

$\beta_{0(i)}$: يمثل الحد الثابت للمقطع العرضي i.

β_j : يمثل معاملات الميل للأنموذج.

$X_{j(it)}$: تمثل المتغيرات التوضيحية j للمقطع العرضي i عند الفترة الزمنية t.

$\varepsilon_{(it)}$: يمثل الخطأ العشوائي للمقطع العرضي i عند الفترة الزمنية t.

يؤدي دمج الأبعاد المقطعية والزمنية في نماذج البيانات الطولية إلى ظهور تحديات مثل مشكلة عدم التباين (heteroscedasticity)أو مشكلة الارتباط بين الأخطاء وعدم استقرار البيانات. تنشأ هذه المشاكل لأن الجمع بين هذين البعدين يؤدي إلى تداخل في التباين على مر الزمن وفي الوحدات المقطعية المختلفة. على سبيل المثال، تشير مشكلة عدم التباين إلى الحالة التي يختلف فيها تباين الأخطاء بين الوحدات المقطعية أو على مدار الزمن، مما يعقد حسابات الخطأ المعياري ويجعل النتائج أقل موثوقية. بالإضافة إلى ذلك، قد تظهر مشكلة ارتباط الأخطاء عندما تكون الأخطاء في نفس الوحدة الزمنية أو بين الوحدات غير مستقلة عن بعضها البعض، مما قد يؤدي إلى نتائج منحازة إذا لم يتم التعامل معها بشكل صحيح.

أما فيما يتعلق بالمعاملات في نماذج البيانات الطولية، فهي تعكس العلاقة بين المتغيرات المستقلة والمعتمدة. وتكمن المشكلة في كيفية التعامل مع هذه المعاملات، فهل يتم التعامل معها على أنها ثابتة (أي أنها لا تتغير بين الوحدات الزمنية والمقطعية) أم أنها عشوائية (أي أنها قد تختلف بين الوحدات أو على مدار الزمن). ففي نموذج التأثيرات الثابتة، يُفترض أن تكون المعاملات ثابتة لكل وحدة، مما يساعد في التحكم في التغاير غير المرصود بين الوحدات. بينما في نموذج التأثيرات العشوائية، يُفترض أن الفروقات بين الوحدات عشوائية وغير مرتبطة بالمتغيرات التوضيحية، مما قد يؤدي إلى تقديرات أكثر كفاءة إذا كان الافتراض صحيحاً (Wooldridge, J. M. 2010).

تسلط هذه المشاكل الضوء على أهمية الاختيار الدقيق بين نماذج التأثيرات الثابتة أو العشوائية، استخدام اختبارات مثل اختبار المضاعف لاجرانج (Lagrange Multiplier Test) لمقارنة نموذج الانحدار التجميعي مع نماذج التأثيرات الثابتة والعشوائية. كما يتم استخدام اختبار هاسمان (Hausman Test)لاختيار النموذج الأنسب بين التأثيرات الثابتة والعشوائية. يمثل معالجة هذه التحديات عاملاً أساسياً لتحسين قوة ودقة التحليل باستخدام البيانات الطولية (Baltagi, B. H. ,2008).

3- الانحدار التقسيمي:

الانحدار التقسيمي هو طريقة إحصائية حديثة توفر فهماً شاملاً للعلاقة بين المتغيرات المستقلة والمتغير التابع عبر عدة نقاط مختلفة من توزيع الاستجابة. على عكس الانحدار الخطي الذي يتنبأ بالمتوسط الشرطي، يمكن للانحدار التقسيمي تقدير التأثيرات في نقاط مختلفة مثل الوسيط أو القيم العليا والدنيا للتوزيع، مما يجعله أداة تحليلية مرنة وقوية(Majid, 2018) . ويعتبر الانحدار التقسيمي مفيدًا بشكل خاص عندما تختلف تأثيرات المتغيرات المستقلة في قيم عالية أو منخفضة من المتغير التابع. ويُستخدم على نطاق واسع في مجالات مثل العلوم الاقتصادية، المالية، البيولوجيا، والعلوم الاجتماعية، حيث من الضروري فهم التأثيرات عبر توزيع البيانات بالكامل وليس فقط المتوسط.

تعود أصول تحليل الكميات إلى القرن التاسع عشر عندما قدم فرانسيس غالتون الكميات كطريقة لدراسة توزيع البيانات بشكل أكثر تفصيلاً من المتوسطات البسيطة. ومع ذلك، لم يتم تطوير مفهوم الانحدار بناءً على الكميات خلال ذلك الوقت، حيث هيمن الانحدار الخطي على التحليل الإحصائي في القرنين التاسع عشر والعشرين. في أواخر السبعينيات، وضع Koenker وBassett مفهوم الانحدار التقسيمي في ورقة بحثية نُشرت عام 1978، والتي وضعت إطارًا رياضيًا قويًا للانحدار التقسيمي، ومنذ ذلك الحين، اكتسب الانحدار التقسيمي انتشارًا واسعًا في مجالات مثل الاقتصاد والصحة العامة.

تعرف الكميات على انها نقاط إحصائية تقسم البيانات إلى أجزاء متساوية، وتوفر فهماً أعمق لتوزيع البيانات. من أشهر هذه الكميات الوسيط والذي يقسم البيانات إلى نصفين متساويين. تشمل الأنواع الشائعة للكميات الأرباع التي تقسم البيانات إلى أربعة أجزاء متساوية، والمئينات التي

تقسمها إلى 100 جزء متساوٍ، والعشيرات التي تقسمها إلى عشرة أجزاء. لكل نوع من هذه الكميات استخدامات محددة، حيث توفر كل منها نظرة مختلفة على توزيع البيانات (Ibrahim, 2021) .

تتمثل الفائدة الرئيسية لاستخدام الكميات في قدرتها على تقديم فهم أعمق وأكثر دقة للبيانات، خاصة عندما تكون البيانات منحرفة أو تحتوي على قيم متطرفة، مما يجعلها تمثل التوزيع بشكل أفضل.

لنفترض أن دالة التوزيع التراكمي (CDF) للمتغير العشوائي Y هي $F(y) = p(Y \leq y)$، وتحتوي على دالة الكثافة الاحتمالية $f(y)$. يتم تعريف الكمّي $\tau - th$ على النحو التالي (Huang and Nguyen, 2018):

$$Q_\tau(y \mid x) = F^{-1}(\tau) = inf\{y: F(y) > \tau \qquad (4)$$

حيث $0 < \tau < 1$ ، في حالة النموذج الانحدار التقسيمي في الربع الثاني والذي يمثل الوسيط $Q_{\tau=0.5}(y \mid x)$. من ناحية أخرى، تهدف نماذج الانحدار التقليدية (مثل الانحدار المتوسط) إلى تقدير القيمة المتوقعة لـ Y عن طريق تقليل دالة الفقدان الخاصة بالخطأ التربيعي $E(y - E(y \mid x))^2$ ، في حين يهدف نموذج الانحدار التقسيمي إلى تقليل دالة فقدان الخطأ المطلق $E \mid y - E(y \mid x) \mid$.

لكن الانحدار التقسيمي يركز على نمذجة الكم الشرطي $Q_\tau(y \mid x)$ بحيث يحقق الشرط التالي:

$$P(Y \leq Q\_\tau) = \tau \qquad (5)$$

لتلبية هذا الشرط، تتضمن الطريقة تقليل الفرق بين Y وQ_τ باستخدام دالة فقدان كما يلي:

$$\rho_\tau(e) = e \cdot (\tau - I\{e \leq 0\}) = \begin{cases} \tau|e| & ; e > 0 \\ (1-\tau)|e| & ; e < 0 \end{cases} \qquad (6)$$

$$= \frac{|e| + (2\tau - 1)e}{2}$$

حيث ان:

$I\{\cdot\}$: يمثل دالة المؤشر القياسية.

لحساب الكمات المرصودة Q_τ ، نقوم بتقليل القيمة المتوقعة لدالة الخسارة:

$$E[\rho_\tau(e)] = E[\rho_\tau(Y - Q_\tau(\mathrm{y} \mid x) = \mathrm{F}^{-1}(\tau)\mu)] =$$

$$\int_{-\infty}^{\infty} \rho_\tau(y - \mu) f(y) dy$$

$$= \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(y_i - \mu) \qquad (7)$$

في هذه الصيغة، نبحث عن القيمة المثلى لـ μ التي تقلل القيمة المتوقعة لدالة الخسارة. يمكن صياغة نموذج الانحدار التقسيمي على النحو التالي:

$$y_i = x_i^T \beta(\tau) + e_i(\tau), i = 1,2, \dots, n \qquad (8)$$

حيث يمثل e_i الخطأ العشوائي، وتكون كثافته $f_\tau(e_i)$ تحقيق الشرط الاحتمالي التالي:

$$P(e_i(\tau) < 0) = \int_{-\infty}^{0} f_\tau(e_i) de_i = \tau \qquad (9)$$

للحصول على تقديرات معلمات الانحدار التقسيمي $\hat{\beta}(\tau)$ ، يتم حل مشكلة التقليل التالية:

$$\hat{\beta}(\tau) = arg \min_{\beta} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i^T \beta(\tau)\right) \qquad (10)$$

تسمح هذه العملية بتحديد معلمات الانحدار التي تقلل الخطأ الخاص بالكم، مما يوفر إطارًا متينًا لفهم النقاط المختلفة في توزيع المتغير التابع.

4- الانحدار التقسيمي للبيانات الطولية

ان اختيار نموذج التأثيرات الثابتة في هذه الدراسة مبني على الرغبة في السيطرة على الخصائص النوعية الثابتة وغير المشاهدة لكل وحدة، كما أن استخدام التأثيرات الثابتة مناسب لأن افتراضات التأثيرات العشوائية قد لا تكون صحيحة في سياق الانحدار التقسيمي، إذ تتطلب استقلالية التأثيرات عن المتغيرات المستقلة وهو شرط يصعب تحقيقه عملياً. يوفّر نموذج التأثيرات الثابتة قدراً أكبر من الموثوقية في تقدير تأثير المتغيرات عبر مختلف الكميات مع التخلص من تأثير الفروق الثابتة بين الوحدات (مثل الحجم أو الثقافة في حالة الشركات أو الدول) مما يقلل التحيز في التقدير من التحديات المعروفة في تقدير نماذج الانحدار التقسيمي ذات التأثيرات الثابتة ، فعلى الرغم من أن طريقة "تحويل داخل الوحدة" (Within-Transformation) يمكن تطبيقها بسهولة في النماذج الخطية من خلال إزالة المتوسط لكل وحدة، إلا أن هذا النهج يصبح أكثر تعقيدًا في الانحدار التقسيمي نظرًا للطبيعة غير المتماثلة لدالة خسارة الكميات (Check Function). تعني عدم الخطية في هذه الدالة، لذا فإن إزالة المتوسط أو توسيط البيانات لكل وحدة قد يؤدي إلى تشويه توزيع الأخطاء أو شكل دالة الانحدار التقسيمي (Galvao et al., 2018).

بالإضافة إلى ذلك، يظهر ما يُعرف بمشكلة "المعلمات العرضية" (Incidental Parameters) عند تقدير التأثيرات الثابتة الفردية لكل وحدة، خاصة في النماذج غير الخطية مثل الانحدار التقسيمي. تنشأ هذه المشكلة بسبب العدد الكبير من المعلمات مقارنة بحجم العينة لكل وحدة، مما يؤدي إلى انحياز في التقدير، لا سيما عندما يكون عدد الوحدات كبيرًا (Arellano & Bonhomme, 2016). علاوة على ذلك، فإن التحويلات الخطية الشائعة في الانحدار الخطي، مثل طرح المتوسط الخاص بكل وحدة، لا يمكن تطبيقها بسهولة في الانحدار التقسيمي بسبب القيود المنهجية المتعلقة بالمتغيرات التفسيرية وتوزيع الأخطاء (Zhang et al., 2019).

تتطلب هذه التحديات استخدام تقنيات تقدير مبتكرة للتغلب على القيود الكامنة في الانحدار التقسيمي مع التأثيرات الثابتة لبيانات السلاسل الزمنية المقطعية (البيانات الطولية). ومن بين هذه التقنيات، استخدام منهجية التقدير بخطوتين لتقدير نماذج الانحدار التقسيمي للبيانات الطولية مع تأثيرات ثابتة وهي مشابه لما قدمه (Canny ,2011)و (Zhang et al. , 2019) و (Chen , 2024).

تقدم طريقة التقدير بخطوتين حلاً عمليًا للتغلب على مشكلة المعلمات العرضية في إطار الانحدار التقسيمي. تقوم هذه الطريقة بتبسيط عملية التقدير من خلال فصل إزالة التأثيرات الثابتة عن تقدير نموذج الانحدار التقسيمي نفسه، مما يساعد على تجنب التعقيدات النظرية والحسابية. لذا ولتقدير معالم نموذج الانحدار التقسيمي ذو التأثيرات الثابتة، سيتم اتباع اسلوب التقدير المكون من خطوتين:

الخطوة الأولى :يتم تقدير التأثيرات الثابتة لكل وحدة مقطع عرضي (كيان) بشكل مستقل. ويتم تقدير التأثيرات الثابتة باستخدام طريقة المربعات الصغرى، حيث يتم فصل التباين الناتج عن التأثيرات الثابتة عن المتغيرات الأخرى.

$$\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - x'_{it}\hat{\beta}_{OLS}\right) \qquad (11)$$

الخطوة الثانية :بعد تقدير التأثيرات الثابتة في الخطوة الأولى، يتم دمج هذه التأثيرات مع القيم الأصلية للمتغير التابع لإزالة التباين الناتج عن التأثيرات الثابتة أي ان $(\hat{y}_{it} = y_{it} - \hat{\alpha}_i)$ بعد ذلك، يتم تطبيق نموذج الانحدار التقسيمي لتقدير معاملات الميل أو الانحدار وفق المعادلة الاتية:

$$\hat{\beta}(\tau) = \arg\min_{\beta} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \rho_\tau(\hat{y}_{it} - x'_{it}\tilde{\beta}) \qquad (12)$$

حيث ان:

$$\rho_\tau(u) = u(\tau - I(u < 0))$$

$\rho_\tau(u)$: دالة الخسارة الخاصة بالانحدار التقسيمي.

بانتهاج هذه العملية المكونة من خطوتين، نتجنب التحيز الناتج عن التقدير المشترك لعدد كبير من المعالم العرضية مع المعالم الأخرى. لقد أظهرت بحوث سابقة أن مثل هذه الإستراتيجية توفر تقديرات أكثر اتساقاً لمعالم الانحدار التقسيمي في البيانات الطولية. بعد تقدير معاملات الانحدار التقسيمي عبر مختلف الكميات لكل وحدة، سنحصل على مجموعة من المعاملات تصف كيفية تأثير المتغيرات المستقلة على المتغير التابع عند كل كمية. عند هذه المرحلة، يمكن أن تبدأ عملية تجميع الوحدات المتشابهة بناءً على هذه المعاملات المقدّرة، كما سنوضّح في القسم التالي.

5- تقنيات التجميع (العنقدة)

بعد الحصول على معاملات الانحدار التقسيمي المقدّرة للوحدات، يأتي دور تقنيات التجميع (Clustering)لتحديد المجموعات الفرعية من الكيانات التي تظهر أنماطاً متشابهة. في هذه الدراسة، تم اختيار خوارزميتين شهيرتين للتجميع هما K-Means وK-Medoids. يعود اختيارنا لهاتين الطريقتين إلى كونهما من أكثر الخوارزميات استخداماً وانتشاراً، فضلاً عن اختلاف خصائصهما فيما يتعلق بالتأثر بالقيم الشاذة:

عندما يُستخدم تحليل التجميع لتصنيف الحالات حيث يشترك أعضاء المجموعات في خصائص مشتركة، يصبح من السهل على الباحث التنبؤ بسلوك هذه الحالات بناءً على عضويتها في المجموعة، نظرًا لتشابه خصائص أعضاء كل مجموعة. في العادة، لا يكون عدد المجموعات أو عضوية الحالات في هذه المجموعات معروفًا مسبقًا في تحليل التجميع. ولهذا السبب، يُعتبر التجميع عملية تصنيف غير موجهة أو تعلّم غير موجه (Unsupervised Learning). هذا النهج يساعد الباحث على فهم البنية الطبيعية للمجموعات واكتشاف الأنماط الكامنة بينها (Tan, Steinbach, & Kumar, 2006). استخدم في هذه الدراسة نوعين من طرق التجميع وهي:

أ- **طريقة K-means:** هي خوارزمية شائعة وبسيطة للتعلم غير الموجه تُستخدم لتقسيم مجموعة بيانات إلى kمجموعات عنقودية، حيث تنتمي كل نقطة بيانات إلى المجموعة التي يكون متوسطها (centroid) الأقرب للنقطة. يعمل هذا المتوسط كنموذج أولي (Prototype) للمجموعة العنقودية (Jambudi, T., & Gandhi, S., 2021). تهدف الخوارزمية إلى تقليل التباين داخل المجموعة(Within-Cluster Variance) ، والذي يُقاس من خلال مجموع المسافات المربعة بين نقاط البيانات ومراكز المجموعات العنقودية الخاصة بها. تتميز خوارزمية K-means بالبساطة والفعالية، ويمكن تلخيص خطواتها الرئيسية كالتالي:

1. **تحديد المراكز الأولية(Initial Centroids) :** في البداية، يتم اختيار kمراكز أولية، حيث kهو عدد المجموعات المطلوب تحديده، وهو معطى يحدده المستخدم.
2. **تعيين النقاط إلى المجموعات:** يتم تعيين كل نقطة بيانات إلى المركز الأقرب لها بناءً على مسافة محددة (عادةً المسافة الإقليدية). النقاط المخصصة لكل مركز تُشكل مجموعة عنقودية(Cluster) .
3. **تحديث المراكز(Centroid) :** بعد تعيين النقاط، يتم تحديث مركز كل مجموعة بناءً على متوسط النقاط في المجموعة.

4. **التكرار حتى الاستقرار:** تُكرر الخطوتان السابقتان (التعيين والتحديث) حتى لا تتغير عضوية النقاط بين المجموعات، بمعنى أن مراكز المجموعات تصبح مستقرة ولا تتغير.

ب- طريقة K-medoids: خوارزميةk-Medoids ، والمعروفة أيضًا باسم Partitioning Around Medoids (PAM)، هي خوارزمية للتجميع مشابهة لخوارزمية .k-means بدلاً من استخدام متوسط النقاط في المجموعة كمركز(centroid) ، تختار k-medoids نقاط بيانات فعلية كمراكز .(medoids) هذا يجعل الخوارزمية أكثر قوة في التعامل مع الضوضاء والقيم الشاذة مقارنةً بـ.k-means تهدف هذه الخوارزمية إلى تقليل مجموع الفروقات بين نقاط البيانات والمراكز الخاصة بها. تم اقتراح طريقة k-medoids للتجميع من قبل Kaufman و .(1987) Rousseeuwتسعى الطريقة إلى العثور على kأعضاء ممثلين من مجموعة البيانات لعكس هيكل البيانات. لتطبيق هذه الطريقة، صمم المؤلفون برنامج PAM (Partitioning Around Medoids)الذي يتألف من مرحلتين، البناء والتبديل (Maechler et al., 2017).

تهدف مرحلة البناء إلى الحصول على المجموعة الأولية من (medoids) عن طريق تقليل متوسط المسافات بين الكائنات والنقطة الممثلة لها. يتم اختيار أول نقطة ممثلة على أنها الأكثر تمركزًا في البيانات، ثم تُحدد النقاط الأخرى بشكل تكراري. في مرحلة التبديل، يتم تحسين المجموعة الأولية من (medoids) عن طريق استبدال كل (medoid) مختار بكائن غير مختار وتقليل متوسط المسافات لجميع (medoids) المحتملة.

6- التجميع اعتماداً على معاملات الانحدار التقسيمي

بينما يساعد الانحدار التقسيمي في فهم تأثير المتغيرات عبر كميات مختلفة من التوزيع، إلا أنه لا يحدد تلقائيًا المجموعات الفرعية داخل البيانات التي قد تظهر سلوكيات أو خصائص مختلفة. وهنا يأتي دور تقنيات التجميع. تتيح هذه التقنيات تجميع الكيانات (مثل الأفراد أو الشركات أو الدول) في مجموعات فرعية بناءً على التشابه في خصائصها أو سلوكياتها. في سياق البيانات الطولية مع الانحدار التقسيمي، يمكن تطبيق التجميع على معاملات الانحدار التقسيمي المقدرة. من خلال تجميع هذه المعاملات، يمكننا تحديد مجموعات فرعية مميزة داخل البيانات المقطعية تستجيب بشكل مختلف لنفس المتغيرات التفسيرية.

الخطوة الأولى لتطبيق الانحدار التقسيمي على البيانات الطولية هي تعريف النموذج. لنفترض أن y_{it}يمثل المتغير التابع للوحدة iفي الزمن t ، وأن x_{it}هو متجه ذو أبعاد pمن المتغيرات المستقلة للوحدة iفي الزمن t يتم تعريف نموذج الانحدار التقسيمي الخطي للبيانات الطولية والناتج من المعادلة (3) و (8) كما يلي:

$$Q_\tau(y_{it} \mid x_{it}) = \alpha_i + x'_{it}\beta_{g_i}(\tau) \qquad (13)$$

حيث ان:

$$i = 1,2,\dots,N \quad , \quad t = 1,2,\dots,T \quad , \quad g = \{1,2,\dots,G\} \quad , \quad 0 < \tau < 1$$

$Q_\tau(y_{it} \mid x_{it})$:تمثل الكمية الشرطية τ^{th} للمتغير التابع y_{it} بشرط المتغيرات المستقلة x_{it} .

α_i : تمثل التأثيرات الثابتة الفردية للمقطع العرضي i .

$\beta_{g_i}(\tau)$:تمثل معاملات المجموعة عند العنقود g_i للكمية τ .

g_i :تمثل المجموعة التي تنتمي إليها المقطع العرضي i .

G: تمثل العدد الإجمالي للعناقيد.

بعد إجراء الانحدار التقسيمي عبر كميات مختلفة للبيانات المقطعية، يتم الحصول على مجموعة من المعاملات لكل كمية. تصف هذه المعاملات كيف تؤثر المتغيرات المستقلة على المتغير التابع عند نقاط مختلفة من توزيعه.

Journal of Business Economics for Applied Research, Vol. (7), No. (Conference):548-564
Doi: https://dx.doi.org/10.37940/BEJAR.2025.7.1.34

الخطوة التالية هي تطبيق تقنيات التجميع على هذه المعاملات المقدرة لتحديد المجموعات الفرعية داخل البيانات المقطعية. الفكرة هي تجميع الكيانات (مثل الأفراد أو الشركات أو الدول) بناءً على التشابه في معاملات الانحدار التقسيمي عبر الكميات المختلفة. ويتم تقديم الخوارزميات المستخدمة كالتالي:

الخطوة الأولى - تعريف النموذج: استخدم نموذج الانحدار التقسيمي الخطي للبيانات الطولية مع التأثيرات الثابتة وفقًا للمعادلة (13).

الخطوة الثانية - تقدير نماذج الانحدار التقسيمي: لكل مقطع في البيانات، يتم تقدير نموذج الانحدار التقسيمي عند كميات مختلفة (مثل النسبة 25%، 50%، و75%). هذا يؤدي إلى مجموعة من المعاملات لكل كمية ولكل مقطع. حيث يتم إيجاد المعالم باستخدام المعادلات (11) و (12).

الخطوة الثالثة - إعداد بيانات المعاملات للتجميع: جمع المعاملات لكل مقطع عبر الكميات المختلفة في مصفوفة β_j، بحيث تمثل الصفوف المقاطع العرضية وتمثل الأعمدة المعاملات المتعلقة بكل كمية.

الخطوة الرابعة – تطبيق خوارزميات التجميع:

اولاً: تطبيق خوارزمية K-means للتجميع:

- **التهيئة :**اختيار kمراكز أولية عشوائيًا من مجموعة البيانات.
- **خطوة التعيين :**يتم تعيين كل نقطة بيانات β_i إلى أقرب مجموعة c_i بناءً على المسافة الإقليدية
- $c_i = \arg\min_j \left\| \beta_i - mean(\beta)_j \right\|^2$ (14)

حيث ان $mean(\beta)_j$ مركز العنقود j .

- **خطوة التحديث :**يتم تحديث مركز كل مجموعة jكمتوسط للنقاط المخصصة للمجموعة

$$mean(B)_j = \frac{1}{|C_j|} \sum_{\beta j \in C_j} \beta_i \qquad (15)$$

- **التقارب :**يتم التحقق مما إذا كانت المراكز قد استقرت. إذا استقرت، تكون الخوارزمية قد تقاربت؛ وإلا، عُد إلى خطوة التعيين.

ثانياً: تطبيق خوارزمية K-medoids للتجميع:

- **التهيئة :**اختيار k عشوائيًا كمراكز أولية من مجموعة البيانات.
- **خطوة التعيين :**تعيين كل نقطة بيانات β_i إلى أقرب مركز c_iلتقليل إجمالي المسافة

$$c_i = arg\min_j \| \beta_i - medoid(\beta_j) \|^2 \quad (16)$$

- **خطوة التحديث :**لكل مجموعة j ، يتم تحديد medoid جديدًا عن طريق تقليل المسافة الإجمالية بين جميع النقاط داخل المجموعة:

$$medoid(\beta_j) = \arg\min_{\beta_k \in C_j} \sum_{\beta_i \in C_j} \| \beta_i - \beta_k \|^2 \qquad (17)$$

- **التقارب :**التحقق مما إذا كانت medoidsقد استقرت (أي لم تحدث تغييرات كبيرة في مواقعها). إذا لم تستقر، يتم تكرار خطوات التعيين والتحديث حتى يتم الوصول إلى التقارب.

الخطوة الخامسة - تحديد المجموعات الفرعية وحساب المتغير المتوقع: بعد التجميع، يتم تعيين كل مقطع إلى مجموعة محددة j ، مما يشير إلى أن المقاطع داخل نفس المجموعة تشترك في أنماط مشابهة للانحدار التقسيمي. يتم حسب القيم المتوقعة $\hat{y}$ لكل مجموعة jباستخدام المعاملات الخاصة بالمجموعة:

حيث β_tهو مركز المجموعة j .

Journal of Business Economics for Applied Research, Vol. (7), No. (Conference): 548-564
Doi: https://dx.doi.org/10.37940/BEJAR.2025.7.1.34

الخطوة السادسة ـ حساب متوسط مربعات الخطأ:(MSE) احسب متوسط مربعات الخطأ لكل كمية τ_i باستخدام المعادلة:

$$MSE(\tau_i) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \hat{y}_{it}(\tau_i)\right)^2 \qquad (19)$$

الخطوة السابعة ـ اختيار أفضل نموذج: يتم تحديد المعاملات الخاصة بالمجموعة β_t التي تقلل من متوسط مربعات الخطأ (MSE) لكمية معينة τ ويتم اعتبار النموذج المحدد هو الأفضل لتمثيل أنماط البيانات المرتبطة بالكميات.

7- المحاكاة

في هذا البحث، تم اعتماد المحاكاة كمنهج أساسي لتقييم أداء نماذج الانحدار التقسيمي للبيانات الطولية مع الأخذ بعين الاعتبار تأثير تطبيق طرق تجميع هي K-Means، K-Medoids استُخدم أسلوب مونت كارلو (Monte Carlo) كإطار محاكاة محكم نظرًا لقدرته على تمثيل سلوك الظواهر بشكل دقيق ومسيطر عليه. تم تصميم برنامج المحاكاة باستخدام لغةMATLAB ، حيث شملت التجربة خطوات تفصيلية لضمان دقة التقدير والتقييم.

بدأت المحاكاة بتحديد حجم العينة بناءً على سيناريوهات مختلفة لأحجام المقاطع العرضية (20،30) وفترات زمنية (6، 18)، بهدف تمثيل سيناريوهات متنوعة وتحليل تأثير تغيير الأبعاد على أداء النماذج. ولتطبيق طرق العنقدة، تم تحديد عدد العناقيد مسبقًا بخمسة مستويات (2، 3، 4، 5، 6). كما تم اعتماد قيم أولية للمعالم $(\alpha_0 = -0.70)$، $(\beta_1 = 0.166)$، $(\beta_2 = 0.0022)$، $(\beta_3 = -0.182)$.

افترض أن المتغيرات التوضيحية تتبع توزيعات مختلفة لضمان تمثيل سيناريوهات متنوعة، حيث وُلد المتغير الأول من توزيع طبيعي$x_1 \sim N(7.19,4.1)$، والمتغير الثاني من توزيع بواسون $x_2 \sim P(6.483)$ ، والمتغير الثالث من توزيع منتظم $x_3 \sim U(0,22)$.

أما الأخطاء العشوائية، فقد وُلدت من توزيع طبيعي أساسي $\varepsilon_{it} \sim N(0,0.2)$ مع إضافة نسبة من القيم الشاذة المولدة من توزيع طبيعي مختلف $\theta \sim N(2,3)$. لتحقيق التوزيع العشوائي للأخطاء، تم دمج القيم الأساسية والشاذة بطريقة تضمن تمثيل الظاهرة بدقة. تم توليد المتغير التابع باستخدام القيم الأولية المفترضة للمعالم والمتغيرات التوضيحية والخطأ العشوائي وفق المعادلة:

أجريت عملية المحاكاة لتقدير معالم الانحدار التقسيمي واستخدام طرق العنقدة لتحديد المجموعات الفرعية ذات الخصائص المتشابهة. كما تم اعتماد معيار متوسط مربعات الخطأ (MSE) كأداة للمقارنة بين الطرق المختلفة لتحديد الطريقة المثلى. لضمان استقرار النتائج، تكررت المحاكاة 1000 مرة لكل سيناريو.

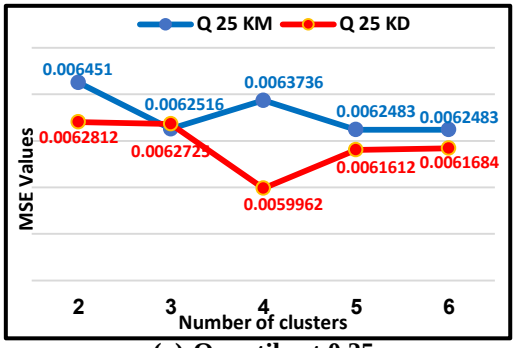
جدول (1): متوسط مربعات الخطأ (MSE) لنموذج الانحدار التقسيمي بالاعتماد على العنقدة (K-means، K-medoids) لـ K من العناقيد وعدد المقاطع العرضية n=20 ولفترات زمنية (t=6,18) ونسبة شواذ (40%)

time	K	Q 25		Q 50		Q 75	
		KM	KD	KM	KD	KM	KD
6	2	0.006451	0.0062812	0.0052475	0.0051282	0.0052752	0.0051725
	3	0.0062516	0.0062725	0.0052067	0.00511	0.0051018	0.0050379
	4	0.0063736	0.0059962	0.0057465	0.0051985	0.0051168	0.0050532
	5	0.0062483	0.0061612	0.0055933	0.0051283	0.0054093	0.0053792
	6	0.0062483	0.0061684	0.0054948	0.0051297	0.0053295	0.0051168
18	2	0.0047799	0.0046146	0.0040611	0.004387	0.0041827	0.0039821

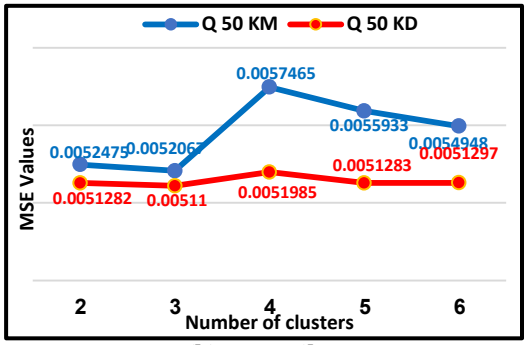
	3	0.0047734	0.0046101	0.0041337	0.0040404	0.0041634	0.0039166
	4	0.004803	0.0046797	0.0041231	0.0038678	0.0041934	0.004118
	5	0.0046797	0.0049569	0.004139	0.0038978	0.0042416	0.0039866
	6	0.0046797	0.0049048	0.0038712	0.0041437	0.0041555	0.004098

يتضح من خلال الجدول رقم (1) ولعدد مقاطع عرضية (n=20) ان في الفترة الزمنية t=6، تتفاوت دقة الطريقتين (KMوKD) عبر الكميات المختلفة، مع ملاحظة تفوق KD بشكل عام، خصوصًا عند نسبة شواذ مرتفعة (40%) .عند الكمية (Q25) ، تكون قيم متوسط مربعات الخطأ (MSE)مرتفعة نسبيًا لكلا الطريقتين بسبب تأثير القيم الدنيا وتوزيع البيانات غير المنتظم. ومع ذلك، تظهر KD أداءً أفضل قليلاً منKM ، حيث تُظهر مقاومة أكبر للقيم الشاذة. عند الكمية (Q50)، تكون القيم أقل بشكل عام، مما يعكس استقرار الوسيط ككمية مركزية أقل تأثرًا بالقيم الشاذة، وتظهر KD تفوقًا طفيفًا على KM عند نسب الشواذ العالية. أما عند الكمية (Q75)، فتكون القيم أقل من (Q25) وقريبة من (Q50)، مع تفوق ملحوظ لـ KD مقارنة بـKM ، مما يعكس قدرتها على التعامل بشكل أفضل مع القيم المرتفعة في التوزيع. من ناحية عدد العناقيد K، تكون القيم الأقل لـ MSE عند عدد عناقيد منخفض مثل K=2أو K=3، مما يعكس قدرة هذه الإعدادات على تجميع البيانات بشكل أكثر دقة في ظل فترة زمنية قصيرة، بينما يؤدي زيادة عدد العناقيد إلى ارتفاع القيم نسبيًا، مما يعكس تشتتًا أكبر داخل المجموعات.

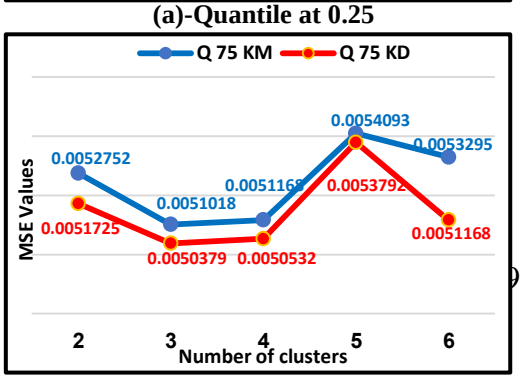
في الفترة الزمنية t=18، تصل النماذج إلى أقصى درجات الدقة والاستقرار، مع تحقيق أقل قيم لمتوسط مربعات الخطأ (MSE) مقارنة بالفترات الزمنية الأقصر. عند الكمية (Q25)، تُظهر KD أداءً مستقرًا وأفضل منKM ، مما يعكس قدرتها على التعامل مع التوزيع غير المتوازن للقيم الدنيا، خاصة عند نسب الشواذ المرتفعة. عند الكمية (Q50)، تكون القيم الأدنى لكلا الطريقتين، مع تفوق طفيف ومستمر لـKD ، مما يؤكد قدرتها على تقديم نتائج دقيقة عند الكمية المركزية. عند الكمية (Q75)، تتحسن القيم بشكل كبير وتُظهر KD أداءً أفضل مقارنة بـKM ، حيث تكون أكثر استقرارًا في مواجهة القيم المرتفعة والشواذ. بالنسبة لعدد العناقيد K، يظهر أن K=3وK=4 يحققان أداءً متوازنًا، مع انخفاض قيمMSE ، بينما يؤدي زيادة Kإلى تشتت أكبر في العناقيد، خاصة عند وجود نسب شواذ مرتفعة.



(a)-Quantile at 0.25

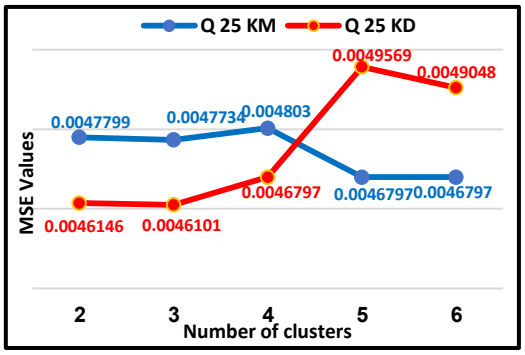


(b)-Quantile at 0.50

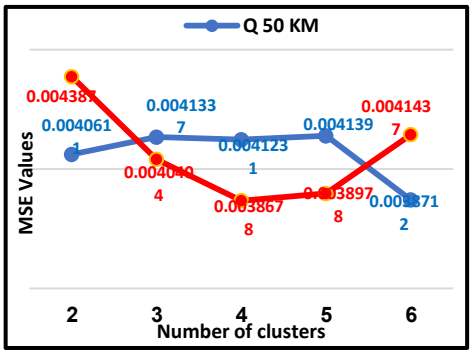


(c)-Quantile at 0.75

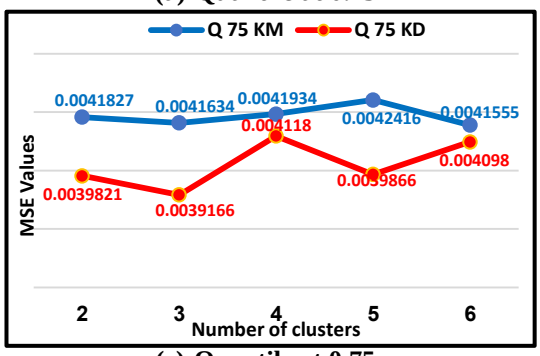
شكل (1): متوسط قيم مربعات الخطأ MSEوفقًا لعدد العناقيد ولكل طريقة تجميع للانحدار التقسيمي عند المقطع العرضي n = 20 والوقت t = 6.



(a)-Quantile at 0.25



(b)-Quantile at 0.50



(c)-Quantile at 0.75

شكل (2): متوسط قيم مربعات الخطأ MSEوفقًا لعدد العناقيد ولكل طريقة تجميع للانحدار التقسيمي عند المقطع العرضي n = 20 والوقت t = 18.

من خلال مقارنة الطريقتين (KM وKD) عبر الكميات الثلاث (Q25، Q50، Q75)، يتضح تفوق KD في جميع الكميات، يظهر هذا التفوق بشكل أكثر وضوحًا عند الكميات (Q25وQ75)، حيث تكون البيانات أكثر عرضة للتأثر بالقيم الشاذة. أما عند (Q50)، فتكون الفروقات بين الطريقتين أقل، مع استمرار التفوق الطفيف لـ KD. يعكس ذلك أن KD هي الخيار الأمثل عند التعامل مع البيانات التي تحتوي على قيم شاذة أو توزيعات غير متجانسة، مما يجعلها أكثر ملاءمة لتحليل البيانات الطولية المعقدة عند (n=20). الشكل (1) و(2) يوضحان القيم المثلى لـMSE عند كل مجموعة ويلاحظ ان أفضل القيم تقع بين k=3 وk=4.

جدول (2): متوسط مربعات الخطأ (MSE) لنموذج الانحدار التقسيمي بالاعتماد على العنقدة (K-means،K-medoids) لـ K من العناقيد وعدد المقاطع العرضية n=30 ولفترات زمنية (t=6,18) ونسبة شواذ (40%).

جدول (2): متوسط مربعات الخطأ (MSE) لنموذج الانحدار التقسيمي بالاعتماد على العنقدة (K-means،K-medoids) لـ K من العناقيد وعدد المقاطع العرضية n=30 ولفترات زمنية (t=6,18) ونسبة شواذ (40%)

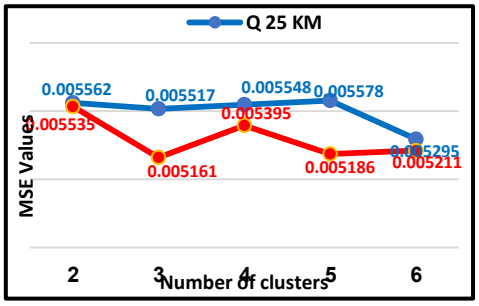
time	K	Q 25		Q 50		Q 75	
		KM	KD	KM	KD	KM	KD
6	2	0.005562	0.005535	0.004413	0.005043	0.005272	0.004335
	3	0.005517	0.005161	0.004173	0.004166	0.004866	0.004226
	4	0.005548	0.005395	0.004467	0.004870	0.004717	0.004335
	5	0.005578	0.005186	0.005043	0.004735	0.004564	0.004461
	6	0.005295	0.005211	0.005378	0.004293	0.004594	0.004399
18	2	0.003955	0.003865	0.003441	0.003693	0.003486	0.003363

	3	0.003949	0.003853	0.003527	0.003385	0.003454	0.003218
	4	0.003994	0.003855	0.003547	0.003394	0.003497	0.003235
	5	0.003932	0.003921	0.003388	0.003422	0.003488	0.003278
	6	0.003929	0.003929	0.003533	0.003432	0.003318	0.003264

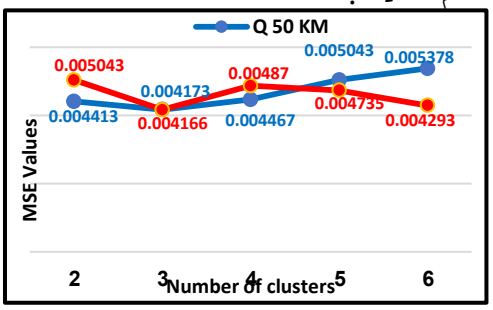
من الجدول رقم (2) ولعدد مقاطع عرضية (n=30) يتضح انه في الفترة الزمنية t=6، تظهر القيم الأعلى لمتوسط مربعات الخطأ (MSE) مقارنة بالفترات الزمنية الأطول، مما يعكس تأثير قصر الفترة الزمنية على دقة النماذج. بالنسبة للكمية (Q25)، تكون القيم مرتفعة نسبيًا، لكن KD تُظهر تفوقًا بسيطًا علىKM ، مما يعكس مقاومتها الأفضل لتأثير الشواذ. عند الكمية (Q50)، تكون القيم أقل وتُظهر أداءً أكثر استقرارًا، مع تفوق ملحوظ لـKD، أما عند (Q75)، تكون القيم أقل مقارنة بـ (Q25)، مع استمرار تفوق KD الذي يظهر قدرتها على التعامل مع القيم الأعلى في التوزيع. بالنسبة لعدد العناقيد K، تحقق القيم المثلى عند K=3أو K=4، حيث يتم ملاحظة انخفاض في القيم، بينما تؤدي زيادة Kإلى ارتفاع طفيف بسبب زيادة تشتت العناقيد.

في الفترة الزمنية t=18، تصل القيم إلى أدنى مستوياتها عبر الفترات الزمنية، مما يشير إلى تحسين كبير في دقة النماذج واستقرارها مع زيادة طول الفترة الزمنية. عند نسبة شواذ 40%، تظل KDمتفوقة على KM في جميع الكميات تقريبًا، مما يعكس مقاومتها الأفضل لتأثير القيم الشاذة. بالنسبة لعدد العناقيد K، يُلاحظ أن K=3أو K=4يحققان الأداء الأفضل، مع انخفاض كبير في قيم MSE.

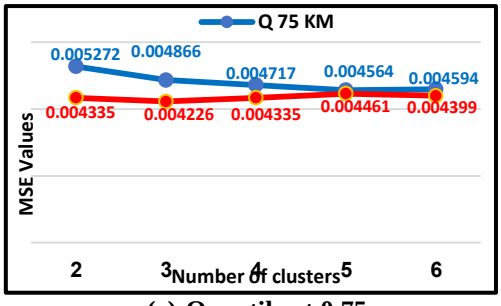
عبر الفترات الزمنية المختلفة، يظهر أن زيادة طول الفترة الزمنية تؤدي إلى تحسين كبير في دقة النماذج وتقليل متوسط مربعات الخطأ (MSE). يظهر تفوق KD بشكل ملحوظ علىKM ، عند نسب الشواذ المرتفعة(40%) ، مما يعكس قدرتها على التعامل مع القيم الشاذة وتقليل تأثيرها. بالنسبة للكميات المختلفة، يُعد (Q50) الأكثر استقرارًا ودقة، بينما تتأثر (Q25) بالقيم الدنيا للتوزيع، وتُظهر (Q75) قدرة جيدة للنماذج على التعامل مع القيم العليا. من حيث عدد العناقيد، تحقق K=3أو K=4توازنًا مثاليًا بين دقة التجميع واستقرار النتائج، مما يجعلها الخيار الأمثل في معظم الحالات.



(a)-Quantile at 0.25

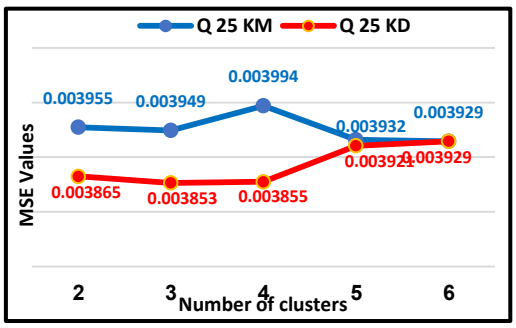


(b)-Quantile at 0.50

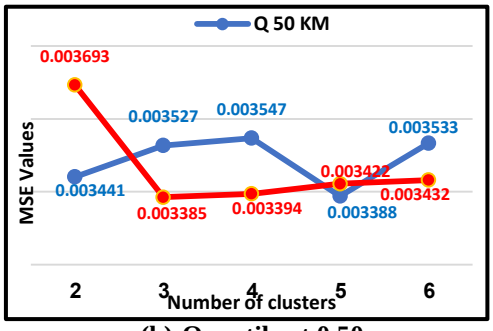


(c)-Quantile at 0.75

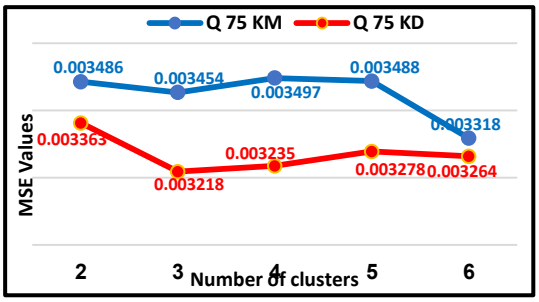
شكل (3): متوسط قيم مربعات الخطأ MSE وفقًا لعدد العناقيد ولكل طريقة تجميع للانحدار التقسيمي عند المقطع العرضي n = 30 والوقت t = 6.



(a)-Quantile at 0.25



(b)-Quantile at 0.50



(c)-Quantile at 0.75

شكل (4): متوسط قيم مربعات الخطأ MSE وفقًا لعدد العناقيد ولكل طريقة تجميع للانحدار التقسيمي عند المقطع العرضي n = 30 والوقت t = 18.

في الشكل (3) و(4)، فنلاحظ نمطاً مشابهاً للحالة السابقة: القيم الأقل لـ MSE تحقق عند K=3، بينما تؤدي زيادة K فوق ذلك إلى ارتفاع طفيف في MSE. ويمكن تفسير ذلك بأن تقسيم الوحدات إلى 3 مجموعات يوازن بين الحصول على تجانس كافٍ داخل كل مجموعة وعدم الإفراط في التقسيم.

8- الاستنتاجات

بناءً على تحليل نتائج تجارب المحاكاة، تكشف نتائج الجداول (1) و(2) أن أداء نماذج الانحدار التقسيمي يتأثر بشكل واضح بعوامل متعددة، أبرزها طول الفترة الزمنية، نسبة القيم الشاذة، عدد العناقيد، ومستوى الكمية المدروس. يظهر تأثير طول الفترة الزمنية في انخفاض متوسط مربعات الخطأ (MSE) بشكل ملحوظ مع زيادة t من 6 إلى 18، مما يشير إلى أن الفترات الزمنية الأطول تساهم في تحسين دقة النماذج واستقرارها.

كما تُظهر نسبة القيم الشاذة تأثيرًا كبيرًا على دقة النماذج، حيث تؤدي النسبة العالية (40%) إلى زيادة القيم غير المتوقعة وتشتيت النتائج. رغم ذلك، أثبتت طريقة KD تفوقها في معالجة البيانات ذات النسب المرتفعة من القيم الشاذة مقارنة بطريقةKM ، حيث سجلت قيمًا أقل وأكثر استقرارًا لمتوسط مربعات الخطأ عبر مختلف السيناريوهات.

أما بالنسبة لتأثير مستويات الكمية، فإن الوسيط (Q50) أثبت أنه المستوى الأكثر استقرارًا ودقة، حيث سجل أدنى قيم لمتوسط مربعات الخطأ، مما يجعله الخيار الأمثل لتحليل تأثير المتغيرات المستقلة على البيانات. من جهة أخرى، أظهرت الكمية (Q25) تأثرًا نسبيًا أكبر بالقيم الدنيا، في حين حققت (Q75) أداءً جيدًا خاصة في تحليل القيم العليا للتوزيع.

وفيما يتعلق بعدد العناقيد، فإن الإعدادات المثلى تتراوح بين K=3وK=4، حيث أظهرت هذه القيم توازنًا مناسبًا بين دقة التجميع وثبات النتائج. بينما يؤدي زيادة عدد العناقيد عن هذا النطاق إلى تشتت أكبر داخل المجموعات وارتفاع في قيم متوسط مربعات الخطأ.

بناءً على هذه النتائج، يمكن استنتاج أن فعالية نماذج الانحدار التقسيمي تتعزز من خلال اختيار تقنيات التجميع المناسبة مع ضبط العوامل المؤثرة. تظهر طريقة KD كخيار أكثر موثوقية عند التعامل مع البيانات التي تحتوي على نسب مرتفعة من القيم الشاذة، بينما يمثل الوسيط (Q50) المستوى الأنسب لتحليل الأنماط العامة.

كما نوصي اعتماد الانحدار التقسيمي في الدراسات الاقتصادية لتحليل تأثير المتغيرات (مثل الدخل، الاستثمار، أو التجارة) عبر شرائح مختلفة من التوزيع، مما يوفر فهماً أدق للتفاوت بين الدول أو الأفراد. واستخدام تقنيات العنقدة لتصنيف الدول أو الشركات ضمن مجموعات متجانسة سلوكيًا، مما يساعد في تخصيص السياسات الاقتصادية وتوجيه الدعم أو القرارات بناءً على خصائص كل مجموعة. كذلك التركيز على تحليل الوسيط (Q50) عند دراسة متغيرات اقتصادية غير متجانسة، مثل الناتج المحلي الإجمالي أو الاستهلاك، كونه يعكس الواقع الاقتصادي بشكل أكثر استقرارًا من المتوسط الذي يتأثر بالقيم الشاذة.

المصادر

1 Abdulrazak, ali S. (2012). Estimation of Missing Data in Panel Data Model with Practical Application. A thesis to the College of Administration and Economics, University of Baghdad.

2 Canay, I. A. (2011). A simple approach to quantile regression for panel data. *Econometrics Journal*, *14*(3), 368-386.

3 Chen, L. (2024). Two-step estimation of quantile panel data models with interactive fixed effects. *Econometric Theory*, *40*(2), 419-446.

4 Dai, X., & Jin, L. (2021). Minimum distance quantile regression for spatial autoregressive panel data models with fixed effects. *Plos one*, *16*(12), e0261144.

5 Galvao, A. F. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, *164*(1), 142-157.

6 Galvao, A. F. and Wang, L. (2015) Efficient minimum distance estimator for quantile regression fixed effects panel data. Journal of Multivariate Analysis, 133, 1-26.

7 Galvao, A. F., Lamarche, C., & Lima, L. R. (2013). Estimation of censored quantile regression for panel data with fixed effects. *Journal of the American Statistical Association*, *108*(503), 1075-1089.

8 Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

9 Huang, M. L., & Nguyen, C. (2018). A nonparametric approach for quantile regression. Journal of Statistical Distributions and Applications, 5, 1-14.

10 Ibrahim, Marwa Khalil. (2021). Bivariate Quantile regression model Estimation for Children Growth in Iraq. The Ph. D to the College of Administration and Economics, University of Baghdad

11 Jambudi, T., & Gandhi, S. (2021). Analysing the effect of different Distance Measures in K-means Clustering Algorithm. *GLS KALP: Journal of Multidisciplinary Studies*, *1*(3), 49-57.

12 Kato, K., Galvao Jr, A. F., & Montes-Rojas, G. V. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, *170*(1), 76-91.
13 Koenker, R. (2004) Quantile regression for longitudinal data. Journal of Multivariate Analysis, 91, 74-89.
14 Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.
15 Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33-50.
16 Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, *157*(2), 396-408.
17 M. Arellano, and S. Bonhomme, (2016). Nonlinear panel data estimation via quantile regressions. The Econometrics Journal, Vol. 19, no. 3, pp. C61–C94.
18 Majid, Haytham H. (2018). Modified Estimators for parameters in TQR Model by using Hierarchical Bayesian method With practical Application. The Ph. D to the College of Administration and Economics, University of Baghdad.
19 Muslim, Basim .S. (2009). Bayesian Analysis For Regression Panel Data Models. The Ph. D to the College of Administration and Economics, University of Baghdad.
20 Powell, D. (2022). Quantile regression with nonadditive fixed effects. *Empirical Economics*, *63*(5), 2675-2691.
21 Schubert, E., & Rousseeuw, P. J. (2019). Faster k-medoids clustering: improving the PAM, CLARA, and CLARANS algorithms. In *Similarity Search and Applications: 12th International Conference, SISAP 2019, Newark, NJ, USA, October 2–4, 2019, Proceedings 12* (pp. 171-187). Springer International Publishing.
22 Tan, P. N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Addison-Wesley.
23 Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). MIT Press.
24 Xiaorong Yang & Jia Chen & Degui Li & Runze Li, (2024). Functional-Coefficient Quantile Regression for Panel Data with Latent Group Structure. Journal of Business & Economic Statistics, Taylor & Francis Journals, vol. 42(3), pages 1026-1040, July.
25 Zhang, Y., Wang, H. J., & Zhu, Z. (2019). Quantile-regression-based clustering for panel data. *Journal of Econometrics*, *213*(1), 54-67.