

مجلة كلية التراث الجامعة

مجلة علمية محكمة
متعددة التخصصات نصف سنوية
العدد الأربعون

30 آب 2024
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)

مقارنة مقدرات Lts و s الحصينة لنماذج البيانات الطولية

أ.د. لقاء علي محمد

حنين نعمه جاسم

كلية الإدارة والاقتصاد/جامعة بغداد

قسم الإحصاء

المُلخص

يعد نموذج التأثيرات الثابتة واحد من أكثر النماذج للبيانات الطولية شيوعاً في العديد من المجالات الإحصائية والاقتصادية، وفي حالة وجود القيم الشاذة في البيانات يكون من الصعب تقدير معاملات الانموذج بطريقة المربعات الصغرى لأنها تعطي تقديرات غير كفوءة، لذلك يتم استخدام الطرائق الحصينة لكونها مقاومة للقيم الشاذة. وفي بحثنا هذا تم الاعتماد على طريقتين حصينتين وهما طريقة المربعات الصغرى المشدبة (Lts) وطريقة (s)، وباستخدام أسلوب المحاكاة للمقارنة بين الطرائق مع أحجام عينات مختلفة متمثلة ($n=25$, $n=100$) متمثلة بالمقاطع العرضية و ($t=8$) متمثلة بالسلاسل الزمنية بحيث يكون لدينا حجمين عينة (200، 800)، مع فرض ثلاث نسب تلوث البيانات (10 %، 20 %، 30 %). ومن خلال استخدام معيار المقارنة متوسط مربعات الخطأ (mse) تبين أن طريقة (Lts) كانت الأفضل في جميع أحجام العينات وعند نسب التلوث المختلفة.

الكلمات المفتاحية: أنموذج التأثيرات الثابتة، التقديرات الحصينة، طريقة المربعات الصغرى، طريقة المربعات الصغرى المشدبة، طريقة (s).

Abstract

The fixed effects model is a widely utilized approach for analyzing panel data within various statistical and economic domains. However, when outliers are present in the dataset, estimating model parameters through the least squares method (ols) becomes challenging, as it yields inefficient estimates. Consequently, robust methods are employed due to their resistance to the influence of outliers. This study focuses on two robust methods, namely the trimmed least squares method (LTS) and the (S) method. A simulation approach is employed to compare these methods across varying sample sizes ($n=25$, $n=100$), considering cross-sectional data ($t=8$) and time series data. This results in two sample size scenarios (200, 800), with three different data contamination rates (10%, 20%, 30%). Using the mean square error (MSE) as the benchmark for comparison, the research reveals that the (LTS) method outperforms the (S) method across all sample sizes and contamination levels.

Keyword: Fixed effects model; robust estimation; least squares method; trimmed least squares method (Lts); (s) method.

المقدمة

تتمتع البيانات الطولية بأهمية كبيرة خصوصاً في البحوث الإحصائية والاقتصادية، حيث تسمح لنا بتحليل البيانات بصورة دقيقة للتغيرات التي تحدث في مجموعة من المقاطع العرضية خلال فترة من الزمن. ومع ذلك، فإن تحليل البيانات الطولية يمكن أن يواجه صعوبات، لا نها تعتمد على بعدين هو البعد الزمني والبعد المكاني، ومن بين هذه الصعوبات هي وجود القيم الشاذة (Outliers) في البيانات. حيث وجود القيم الشاذة في البيانات يؤثر على طرق التقدير التقليدية حيث يؤدي إلى تقديرات غير دقيقة. لذلك كان من الضروري البحث عن طرق بديلة لا تتأثر بوجود القيم الشاذة بدل من حذفها كبديل عن الطرق التقليدية وهذه الطرائق تسمى بالطرق الحصينة والتقديرات الناتجة تدعى بالمقدرات الحصينة حيث تتشابه جميع التقديرات الحصينة بنقطين رئيسيتين: تعطي أوزان أقل للملاحظات الشاذة لتقليل من تأثيرها على عملية التقدير، وتعتمد على استخدام أسلوب التكرار، يمكن أن تظهر القيم الشاذة في المتغير التوضيحية أو في متغيرات الاستجابة وكلاهما يؤثر على عملية التقدير.

منهجية البحث

2-1 مشكلة البحث:

تتمثل مشكلة هذا البحث في وجود قيم شاذة في أنموذج التأثيرات الثابتة للبيانات الطولية، مما يؤدي وجودها الى عدم دقة في عملية تحليل البيانات مما يؤدي إلى تقديم تقديرات غير دقيقة. يهدف البحث إلى حل هذه المشكلة من خلال استخدام مقدرات حصينة تساعد في تحسين استقرار ودقة أنموذج التأثيرات الثابتة.

2-2 هدف البحث:

الهدف الرئيسي لهذا البحث هو تحسين دقة تقديرات أنموذج التأثيرات الثابتة داخل المجموعة للبيانات الطولية. يتم ذلك من خلال استخدام مقدرات حصينة تمكن النموذج من التعامل بشكل فعال مع القيم الشاذة الموجودة في هذه البيانات.

3- الجانب النظري:

1-3 البيانات الطولية:

البيانات الطولية (panel data) هي مجموعة من البيانات المأخوذة عبر فترة زمنية، حيث انها تجمع بين مميزات الخاصة بالبيانات المقطعية وبيانات السلاسل الزمنية. ويمكن تصنيفها الى بيانات متزنة (Balanced Panel Data) عندما تكون جميع المشاهدات قابلة للقياس خلال فترات زمنية متساوية. وبيانات غير متزنة (Unbalanced Panel Data) عندما تكون جميع المشاهدات قابلة للقياس خلال فترات زمنية مختلفة. وتتميز البيانات الطولية مقارنة بنماذج المقاطع العرضية ونماذج السلاسل الزمنية باحتوائها على محتوى معلومات أكثر، مما تساهم في الحصول على تقديرات ذات موثوقية عالية. تتميز بمزايا مختلفة حيث تتمكن بالتحكم بتجانس التباين الفردي الذي يظهر في المقاطع العرضية والسلاسل الزمنية. بالإضافة الى ذلك فهي تظهر كفاءة أعلى وتمتد درجات حرية أكثر.

3-2 نموذج التأثيرات الثابتة للبيانات الطولية:

يعتبر انموذج التأثيرات الثابتة واحد من اهم النماذج المستخدمة في تحليل البيانات الطولية ويمكن معرفته من خلال الصيغة التالية:

$$y_{it} = \beta_{0(i)} + X'_{it}\beta + \varepsilon_{it} \quad ; \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T \quad (1)$$

$$\varepsilon_{it} \sim N(0, \sigma^2_{\varepsilon})$$

حيث ان:

(y_{it}) يمثل متغير الاستجابة وان (X'_{it}) تمثل المتغيرات المستقلة، بينما ($\beta_{0(i)}$) تمثل معلمة الحد الثابت التي تعتبر متغيرات عشوائية التي تتغير في كل مقطع عرضي خلال فترات زمنية، فيما تعتبر (β) قيمة الميل الحدي التي تظل ثابتة عبر المقاطع العرضية خلال الزمن. وان (ε_{it}) يمثل الخطأ العشوائي ولنموذج التأثيرات الثابتة اكثر من نوع وفي بحثنا هذا ركزنا على نموذج التأثيرات الثابتة داخل المجموعة ويستخدم عملية التمرکز للتخلص من التأثيرات الفردية والتي تقوم بحساب انحرافات القيم عن متوسطها.

نقوم بإيجاد متوسط المشاهدات عبر الزمن

$$\bar{y}_i = \beta_{0(i)} + \bar{x}_i\beta + \bar{\varepsilon}_i \quad (2)$$

حيث ان

$$\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$$

$$\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}$$

$$\bar{\epsilon}_i = T^{-1} \sum_{t=1}^T \epsilon_{it}$$

وعليه يكون النموذج بالصيغة التالية:

$$\dot{y}_{it} = \dot{x}_{it}\beta + \dot{\epsilon}_{it} \quad (3)$$

اذ ان

$$\dot{y}_{it} = (y_{it} - \bar{y}_i)$$

$$\dot{x}_{it} = (x_{it} - \bar{x}_i)$$

$$\dot{\epsilon}_{it} = (\epsilon_{it} - \bar{\epsilon}_i)$$

3-3 القيم الشاذة:

تعد القيم الشاذة أحد المشاكل التي قد يواجهها الباحث تحليل البيانات، حيث يمكن أن تؤثر على نتائج التحليل والتفسير الإحصائي للبيانات إذا لم يتم التعامل معها بشكل صحيح. تُعرف القيم الشاذة بأنها تلك المشاهدات في العينة التي تكون قيمتها صغيرة أو كبيرة جداً مقارنة ببقية مشاهدات العينة مما يؤدي وجودها الى مشكلة في التقدير، حيث ان طرق التقدير التقليدية ومنها طريقة المربعات الصغرى تكون غير كفؤة ولهذا نلجأ الى استخدام طرق أكثر كفؤة ومنها الطرائق الحصينة.

4- طرائق التقدير

1-4 طريقة المربعات الصغرى لتقدير معلمات نموذج التأثيرات الثابتة:

واحدة من اهم الطرق وأكثرها استخداما لتقدير المعلمات هي طريقة المربعات الصغرى (OLS) حيث تعطي أفضل تقدير خطي غير متحيز (BLUE). وفي سياق تقدير انموذج التأثيرات الثابتة وصيغته في المعادلة رقم (3) وبعد تطبيق طريقة المربعات الصغرى على الانموذج نحصل على تقديرات للمعلمات كما بالصيغة التالية:

$$\hat{\beta}_{(fe)} = \sum_{i=1}^N \sum_{t=1}^T (\dot{x}_{it} \dot{x}_{it}')^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T \dot{x}_{it}' \dot{y}_{it} \right) \quad (4)$$

2-4 طريقة S:

تعتبر طريقة (S) من بين اهم طرائق التقدير الحصينة لكونها تمتلك نقطة انهيار عالية، حيث قام كل من روسيو و يوهاي (Rousseeuw and Yochai) بتقديمها في عام (1984)، ويعتمد مبدا عمله على تقليل مجموع الكلي للأخطاء. حيث يمكن تعريفها من خلال:

$$\hat{\beta}_s = \arg \min_{\beta} \hat{\sigma}_s(e_{it}) \quad (5)$$

حيث يتم تحديد اقل مقدر حصين للتشتت ($\hat{\sigma}_s$) من خلال

$$\min_{\beta} \sum_{i=1}^n p\left(\frac{e_{it}}{\hat{\sigma}_s}\right) \quad (6)$$

اذ

$$\hat{\sigma}_s = \sqrt{\frac{1}{NK} \sum_{i=1}^N \sum_{t=1}^T w_{it} e_i^2} \quad (7)$$

حيث ان :

(K : قيمة ثابتة وتساوي (0.199)) ، $w_i = \omega_\sigma = \frac{p(U_i)}{U_i^2}$ ، والتقدير الاولي يكون

$$\hat{\sigma}_s = \frac{\text{median} |e_{it} - \text{median}(e_{it})|}{0.6754} \quad (8)$$

اذ ان (p) دالة وزن وزن توكي (Tukey's biweight function)

$$p(x) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^2} & \text{if } |x| \leq c \\ \frac{c^2}{6} & \text{if } |x| > c \end{cases} \quad (9)$$

حيث تكون معادلة رقم (6) بعد اشتقاق دالة (p) بالصيغة التالية:

$$\sum_{i=1}^N x_{itj} \Psi \left(\frac{y_{it} - \sum_{j=0}^k X'_{it} \beta}{\hat{\sigma}_s} \right) \quad (10)$$

حيث ان (Ψ) دالة مشتقة من دالة (p) من خلال الصيغة التالية:

$$\Psi(u_i) = p'(u_i)$$

$$= \begin{cases} u_i [1 - \left(\frac{u_i}{c}\right)^2]^2 & \text{if } |u_i| \leq c \\ 0 & \text{if } |u_i| > c \end{cases} \quad (11)$$

حيث ان

$$c=1.547$$

3-4 طريقة المربعات الصغرى المشدبة (LTS)

يعتبر مقدر المربعات الصغرى المشدبة (LTS) من بين المقدرات المستخدمة بشكل كبير في تقدير المعلمات، لتمييزه بنقطة انهياره العالية والتي تكون (50%). قدمه روسيو (Rousseew) في عام (1984) ويمكن حسابة من خلال الصيغة التالية:

$$\hat{\beta}_{LTS} = \arg \min \sum_{i=1}^h e_{(i)}^2 \quad (12)$$

حيث تمثل $e_{(i)}^2$ مربعات البواقي وان (h) البيانات الجيدة التي لا يمكن تشذيبها ويمكن حسابها حسب الصيغة التالية:

$$h = \frac{NT + P + 1}{NT} \quad (13)$$

وان (h) تقع ضمن الفترة $\left(\frac{NT}{2} < h < NT\right)$

5- الجانب التطبيقي:

لغرض تطبيق طرائق التقدير المدروسة في الجانب النظري ومعرفة مدى ملائمة الطرائق في حالة البيانات الخالية من الشواذ وفي حالة احتوائها على القيم الشاذة ، ولتحقيق ذلك نستعمل أسلوب المحاكاة لغرض تطبيق أكثر عدد ممكن من الحالات المقترحة وبذلك تكون النتائج أكثر شمولية للواقع العملي ، سيتم تنفيذ البرنامج بلغه (R) وباستعمال احجام عينات مختلفة متمثلة ($n=25$ ، $n=100$) مقاطع عرضية و ($t=8$) تمثل السلاسل الزمنية ، وبافتراض نسب تلوث مختلفة (10% ، 20% ، 30%) وبتكرار التجربة (1000) مرة لكل عينة ، وبافتراض عدد المتغيرات ($p=3$) . وكانت النتائج كما يلي:

جدول (1) يمثل قيم متوسط مربعات الخطأ لجميع حجومات العينات وجميع نسب التلوث

قيم الشاذة	N	Estimates	Mse
10%	200	OLS	1.052355976
		S	0.991007318
		LTS	0.923488885
10%	800	OLS	0.997104367
		S	0.937896683
		LTS	0.817286877
20%	200	OLS	1.052355976
		S	0.991007318
		LTS	0.923488885
20%	800	OLS	0.997104367
		S	0.937896683
		LTS	0.817286877
30%	200	OLS	1.011523315
		S	1.001203026
		LTS	0.860724566
30%	800	Ols	1.003898454
		S	0.997995198
		Lts	0.851649523

6- تحليل النتائج:

نلاحظ من الجدول رقم (1) ان النتائج كانت متقاربة في جمع احجام العينات وعند نسب التلوث المختلفة، حيث تبين من النتائج الى كفاءة طريقة (Its) عند حجوم العينات (200، 800) وعند نسب التلوث (10%، 20%، 30%) لانها تمتلك اقل متوسط مربعات الخطأ.

7- الاستنتاجات:

أظهرت تجربة المحاكاة ان أفضل طريقة في تقدير معلمات البيانات الطولية لأنموذج التأثيرات الثابتة كانت (Its) عند جميع أحجم العينات وجميع نسب التلوث لانها تعتمد على الوسط. كانت لحجم العينة تأثير على دقة النتائج، حيث كلما كبر حجم العينة قل متوسط مربعات الخطأ وبالتالي يؤدي الى دقة طريقة (Its). بينت النتائج ان طريقة (Its) كانت الأفضل عند زيادة حجم العينة وقلت نسبة التلوث، حيث أظهرت النتائج ان اقل متوسط مربعات الخطأ كانت عند حجم عينة (800) ونسبة تلوث (10%).

8- التوصيات

نظرا لجودة طريقة (Its) في التقدير نوصي باستخدامها لتقدير معلمات أنموذج التأثيرات الثابتة التي من الممكن ان يحتوي على مشاكل أخرى. نوصي بإجراء دراسات حول البيانات الطولية الغير متزنة والتي تحتوي على القيم الشاذة. يمكن استخدام أكثر من طريقة تقدير حصينة لتقدير أنموذج التأثيرات الثابتة في حالة القيم الشاذة. ننصح بإجراء دراسات إضافية لمعرفة كيفية تأثير القيم الشاذة على تقديرات معلمات أنموذج التأثيرات الثابتة وتقييم مدى كفاءة الطرق المستخدمة في التعامل معها. نوصي الباحثين بإجراء دراسة تطبيقية باستخدام بيانات حقيقية لتقدير أنموذج التأثيرات الثابتة في حالة وجود القيم الشاذة باستخدام طرائق التقدير المدروسة.

9- المصادر:

- الجمال، زكريا يحيى نوري. (2012). اختيار النموذج في نماذج البيانات الطولية الثابتة والعشوائية. المجلة العراقية للعلوم الإحصائية، مج. 12، ع. 21، ص ص. 266-285.
- عمر عبد المحسن علي، & خالد جمال كاظم. (2018). المربعات الصغرى المشدبة الموزونة لتقدير تأثير مياه الصرف الصحي في تلوث مياه نهر دجلة/محافظة واسط. Journal of Economics and Administrative Sciences, 24(109), 486-486.
- م. مائل كامل ثامر. (2020). مقارنة ما بين بعض الطرائق الحصينة وغير الحصينة لتقدير نماذج الانحدار مع تطبيق عملي. مجلة كلية المعارف الجامعة، 30(1)، 915-896.
- Alma, Ö. G. (2011). Comparison of robust regression methods in linear regression. Int. J. Contemp. Math. Sciences, 6 (9), 409-421.
- Bahez, Z. K., & Rasheed, H. A. (2022). Comparing Some of Robust the Non-Parametric Methods for Semi-Parametric Regression Models Estimation. Journal of Economics and Administrative Sciences, 28 (132), 105-117.
- Baltagi, B. H. (2005). Econometric Analysis of Panel Data, John Wiley&Sons Ltd. West Sussex, England.
- Bramati, M. C., & Croux, C. (2007). Robust estimators for the fixed effects panel data model. The econometrics journal, 10 (3), 521-540.



Hamiye Beyaztas, B., & Bandyopadhyay, S. (2020). Data Driven Robust Estimation Methods for Fixed Effects Panel Data Models. arXiv e-prints, arXiv-2011.

Hamiye Beyaztas, B., & Bandyopadhyay, S. (2020). Robust estimation for linear panel data models. *Statistics in medicine*, 39(29), 4421-4438.

Nwakuya, M. T., & Biu, E. O. (2019). Comparative Study of Within-Group and First Difference Fixed Effects Models. *American Journal of Mathematics and Statistics*, 9(4), 177-181.

11- Susanti, Y., Pratiwi, H., Sulistijowati, S., & Liana, T. (2014). M estimation, S estimation, and MM estimation in robust regression. *International Journal of Pure and Applied Mathematics*, 91(3), 349-360.