المجلة العراقية للعلوم الإحصائية

www.stats.mosuljournals.com

استخدام أوزان بيز لمعالجة مشكلة عدم التجانس في تباين الأخطاء العشوائية في النماذج الخطية

طه حسين علي الزبيدي و كامران حسن أحمد وجيمن ابو بكر عمر

قسم الاحصاء والمعلوماتية ، كلية الإدارة والاقتصاد ، جامعة صلاح الدين ، اربيل ، العراق

معلومات النشر

تاريخ المقالة:
تم استلامه في 11 تموز 2020
تم القبول في 1 ايلول 2020
متاح على الإنترنت في 1 كانون الاول 2020

الكلمات الدالة:
تحليل الإنحدار، مشكلة عدم تجانس التباين، أسلوب بيز
طريقة المربعات الصغرى الموزونة.

المراسلة:
طه حسين علي الزبيدي
drtahaalaa1970@yahoo.com

الخلاصة

تمّ في هذا البحث إقتراح إستخدام أسلوب بيز ذات المعلومات الخبرية في حساب أوزان بيز وتوظيفها في معالجة مشكلة عدم تجانس تباين قيم الخطأ العشوائي عند تقدير معلمات أنموذج الإنحدار الخطي بإستخدام طريقة المربعات الصغرى الموزونة (BWLS). ومن ثم مقارنتها مع الطريقة التقليدية من خلال جانب تجريبي لمحاكاة بيانات مولدة من توزيع طبيعي ولعدة حالات مختلفة فضلاً عن جانب تطبيقي لبيانات حقيقية. وتوصلت نتائج البحث إلى أفضلية الطريقة المقترحة على الطريقة التقليدية بالإعتماد على بعض المعايير الإحصائية من خلال برنامج مصم لهذا الغرض بلغة ماتلاب.

DOI: 10.33899/IQJOSS.2020.167391,

المقدمة Introduction

يعتبر تحليل الانحدار من اوسع الطرق الاحصائية استخداماً في مختلف العلوم حيث يحدد العلاقة بين المتغيرات على هيئة المعادلة ويستدل من تقدير معلماتها على أهمية وقوة واتجاه هذه العلاقة كما يبين تقدير الاستجابة والتنبؤ بها بما يفيد كثيراً في التخطيط والقرارات المناسبة.

من الفروض أو الشروط الواجب توفرها عند إجراء تحليل الإنحدار هو تجانس تباين قيم الخطأ العشوائي (الراوي، 1987) ففي معظم الدراسات والابحاث وخاصة التي تعتمد منها على البيانات المقطعية (Cross–Section Data) فان تشتت مشاهدات البيانات المقطعية الخاصة بالمتغير المعتمد قد تختلف اختلافاً كبيراً من مستوى الى آخر من مستويات المتغيرات المستقلة التي ربما تؤدي إلى ظهور مشكلة عدم تجانس تباين الخطأ (Heteroscedasticity) وبدورها تؤدي إلى مقدرات معلمات أنموذج خطي غير كفوءة ومتحيزة في تقديراتها لمعلمات الأنموذج فضلاً عن إختبارات المعنوية (t, F) غير المقنعة ولايمكن إعتمادها (Patrick et. al., 2003)، وبالتالي لا يمكن تقدير معلماتها باستخدام طريقة المربعات الصغرى الاعتيادية (*OLS*)، لذلك يستخدم عادةً طريقة المربعات الصغرى الموزونة (*WLS*) لمعالجة هذه المشكلة وذلك باستخدام وزن معين لجعل التباين للاخطاء متجانسة وثابتة.

من جانب آخر وبشكل عام في علم الاحصاء توجد مدرستان لتقدير معلمات التوزيعات الاحتمالية: الاولى تعرف بالمدرسة التقليدية والثانية تسمى بمدرسة بيز. لذلك تناول الجانب النظري الاسلوب التقليدي واسلوب بيز في تقدير الاوزان واستخدامها في طريقة المربعات الصغرى الموزونة لتقدير معلمات نموذج الانحدار.

2: الجانب النظري:

2–1: مفهوم تحليل الانحدار:

ان تحليل الانحدار عبارة عن أحد الادوات الاحصائية الاكثر استعمالاً لانه يعطينا طريقة سهلة لتحديد العلاقة بين المتغيرات. هذه العلاقة بين المتغيرات يمكن التعبير عنها بشكل معادلة تحتوي على المتغير المعتمد (Y_i) أو الاستجابة مع واحد أو أكثر من المتغيرات المستقلة $(X_{i1}, X_{i2},......, X_{ik})$ أو ان تحليل الانحدار هو مجموعة طرق الاحصائية التي تتعامل مع الصيغ المختلفة للنماذج الرياضية التي تصف العلاقات بين المتغيرات (الدليمي، 1988) بحيث استخدام نماذج هذه العلاقات لغرض التنبؤ والاستنتاجات الاحصائية الاخرى.

2–2: أنموذج الانحدار الخطي العام أو المتعدد:

نفرض ان لدينا المتغير المعتمد (Y_i) دالة خطية للمتغيرات المستقلة $X_{i1}, X_{i2},......, X_{ik}$ فإن نموذج الانحدار المتعدد يأخذ الصيغة التالية:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + U_i \qquad ...(1)$$

ويمكن التعبير عنها باستخدام المصفوفات كالاتي:

$$Y = X\beta + U \qquad ...(2)$$

حيث أن Y: متجه لمشاهدات المتغير المعتمد ذات بعد $(n \times 1)$

X : مصفوفة مشاهدات المتغيرات المستقلة ذات بعد $[n \times (k+1)]$

β : متجه للمعلمات المجهولة ذات بعد $[(k+1) \times 1]$

U: متجه الاخطاء العشوائية ذات بعد $(n \times 1)$

2–3: مشكلة عدم تجانس تباين الخطأ في نموذج الانحدار

ان احدى الفرضيات التي اعتمدت في تقدير معالم النموذج الخطي سواء كان بسيطاً او عاماً هي فرضية التجانس (الفرضية الثانية) الآتية:

$$V(U_i) = \sigma^2 \qquad i = 1,2..............., n$$

تعرف هذه الفرضية بفرضية تجانس تباين الخطأ (Error Homogeneity of variance Assumption) ولكن أحياننا لا تتحقق هذه الفرضية (كاظم، 2002)، ففي معظم الدراسات والابحاث وخاصة التي تعتمد منها على بيانات احصائية تأخذ شكل البيانات المقطعية (Cross–Section Data) فان تشتت مشاهدات البيانات المقطعية الخاصة بالمتغير المعتمد قد تختلف اختلافاً كبيراً من مستوى الى اخر من مستويات المتغيرات المستقلة (عبدالمنعم، 2005).

إذا لم تتحقق فرضية تجانس التباين فان ذلك يؤدي إلى عدم ثبوت تباين الخطأ بين المشاهدات.

$$V(U_i) \neq \sigma^2 \qquad i = 1,2..............., n$$

$$\sigma_1^2 \neq \sigma_2^2 \neq .... \neq \sigma_n^2 \neq \sigma^2$$

هذا يؤدي إلى ظهور مشكلة عدم تجانس تباين الخطأ (Heteroscedasticity) وبموجبها سوف لن تمتلك تقديرات المربعات الصغرى الاعتيادية لمعالم النموذج الخطي صفة اقل تباين ممكن بعبارة اخرى سوف لن تكون افضل تقدير خطي غير متحيز (Best linear unbiased estimator).

2–4: اختبار كشف مشكلة عدم تجانس تباين الخطأ :

يتم اكتشاف عدم ثبات تباين الخطأ بواسطة عدة طرق منها اختبار جولدفيلد–كوندات (Goldfeld–Quandt Test) حيث اقترح ترتيب البيانات حسب المتغير المستقل X ترتيباً تصاعدياً أو تنازلياً ثم بعد ذلك تقسم البيانات المرتبة الى قسمين مع حذف بعض المشاهدات من الوسط ثم اجرى تحليل الانحدار لايجاد مجموع مربعات البواقي لكل قسم وذلك للحصول على القيمة المحسوبة للإحصاءة

$$F = \frac{MSE_2}{MSE_1}$$

وتقارن مع القيمة الجدولية لرفض أو عدم رفض فرضية العدم.

2-5: معالجة مشكلة عدم تجانس تباين الخطأ العشوائي:

من أبرز الطرق المستخدمة لمعالجة المشكلة هي طريقة المربعات الصغرى الموزونة، وذلك باستخدام وزن معين (W_i) لجعل التباين الاخطاء متجانسة وثابتة (علي، (2007)).

ومن اجل دراسة هذه الظاهرة نفترض أن لدينا النموذج الآتي:

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$

نضرب طرفي النموذج بـ $\sqrt{W_i}$

$$\sqrt{W_i}Y_i = \beta_0\sqrt{W_i} + \beta_1 X_i\sqrt{W_i} + \sqrt{W_i}U_i \qquad ...(3)$$

حيث W_i معكوس التباين أي $W_i = \frac{1}{\sigma_i^2}$

ان اساس طريقة المربعات الصغرى الموزونة هو جعل مجموع مربعات الخطأ التالي اقل مايمكن:

$$Q = SSE = \sum W_i (Y_i - \beta_0 - \beta_1 X_i)^2$$

وباستخدام المشتقات الجزئية لهذه المعادلة بالنسبة لكل من (β_0, β_1)ثم جعل الناتج مساوياً للصفر وبعد التبسيط نحصل على المعادلتين الطبيعتين التالتين:

$$\beta_0 \sum W_i + \beta_1 \sum W_i X_i = \sum W_i Y_i$$

$$\beta_0 \sum W_i X_i + \beta_1 \sum W_i X_i^2 = \sum W_i X_i Y_i$$

وباستخدام المصفوفات نحصل على تقديرات (*WLS*) والتي تساوي:

$$\hat{\beta} = (X'WX)^{-1}(X'WY)$$

حيث

$$W = \begin{bmatrix} W_1 & 0 & ..... & 0 \\ 0 & W_2 & ..... & 0 \\ . & . & & . \\ & & ..... & \\ . & . & & . \\ 0 & 0 & ..... & W_n \end{bmatrix}$$

$$W_i = \frac{1}{\hat{\sigma}_i^2}$$

$$\hat{\sigma}_i^2 = S_{Y_i}^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1} \qquad ...(4)$$

2-6: مفهوم نظرية بيـز:

لقد وُضِعت نظرية بيز في أواسط القرن الثامن عشر من قبل الكاهن (*Thomas Bayes*). يركز أسلوب بيز في التقدير بمفهومه بشكل عام على توظيف معلومات أولية (*prior Information*) حول المعلمات المجهولة $\theta = \theta_1, \theta_2, ...\theta_p$ المطلوب تقديرها (أحمد، 2013) ويمكن وصفها على شكل توزيع إحتمالي يعرف بدالة الكثافة الإحتمالية الأولية (*Prior p.d.f*) ويرمز لها $f(\theta)$. أما دالة التوزيع الحالي لمشاهدات العينة قيد الدراسة فتكون فيها المتغيرات العشوائية (*Y*) لهذه المشاهدات دالة توزيعية تعتمد على θ ويرمز لها $L(\theta)$ تسمى دالة الإمكان (*Likelihood Function*).

ان مقدر بيز ما هو الا الجزء الغني بالمعلومات في دالة الكثافة الإحتمالية اللاحقة (*posterior p.d.f*) التي يمكن الحصول عليها من دمج دالة الكثافة الإحتمالية الأولية للمعلمات مع دالة الإمكان للمشاهدات.

2-7: تقدير بيز لتباين المتغير المعتمد (σ^2)

سيتم تقدير بيز لـ (σ^2) بإستخدام التوزيع الاحتمالي الأولي المعلوماتي لـ (σ^2) ويفترض ان (σ^2) تتوزع توزيع معكوس كاما (*Inverse Gamma*) وفق الدالة الإحتمالية الآتية(P. Richard Hahn et al., 2020) :

$$f(\sigma^2)=\frac{(\frac{v_0S_0^2}{2})^{\frac{v_0}{2}}}{\Gamma(\frac{v_0}{2})}(\sigma^2)^{-(\frac{v_0}{2}+1)}\exp\left[\frac{-v_0S_0^2}{2\sigma^2}\right] \quad , \quad \sigma^2>0 \quad , \quad v_0>0$$

where: $v_0S_0^2=\sum(Y_{i0}-\bar{Y}_0)^2$

$v_0=n_0-1$

وبإستخدام التوزيع الاحتمالي الأولي غير المعلوماتي (*Jeffreys*) لـ (θ) : $f(\theta|\sigma^2)\propto\frac{1}{\sigma}$

وللحصول على دالة كثافة إحتمالية أولية مشتركة للمعلمات (β,σ^2) يتم إستخدام الصيغة الآتية:

$$f(\beta,\sigma^2)\propto f(\sigma^2)\,f(\theta|\sigma^2)$$

$$f(\beta,\sigma^2)\propto\left[\frac{1}{\sigma}\right](\sigma^2)^{-(\frac{v_0}{2}+1)}\exp\left[\frac{-v_0S_0^2}{2\sigma^2}\right]$$

اما دالة الإمكان (Box and Tiao, 1992) فتكون:

$$L(\beta,\sigma^2)\propto\left[\sigma^2\right]^{-\frac{n}{2}}\exp\left[\frac{-1}{2\sigma^2}\sum(Y-\theta)^2\right] \qquad -\infty<Y<\infty \quad , \quad \sigma^2>0$$

$$L(\beta,\sigma^2)\propto\left[\sigma^2\right]^{-\frac{n}{2}}\exp\frac{-1}{2\sigma^2}\left[\sum(Y-\bar{Y})^2+n(\theta-\bar{Y})^2\right]$$

$$L(\beta,\sigma^2)\propto\left[\sigma^2\right]^{-\frac{n}{2}}\exp\frac{-1}{2\sigma^2}\left[vS^2+n(\theta-\bar{Y})^2\right]$$

where: $vS^2=\sum(Y_i-\bar{Y})^2$

$v=n-1$

من خلال نظرية بيز نحصل على دالة الكثافة الإحتمالية اللاحقة المشتركة للمعلمات (θ,σ^2) وكالآتي:

$$f(\theta,\sigma^2|Y)\propto f(\theta,\sigma^2)\,L(\theta,\sigma^2)$$

$$f(\theta,\sigma^2|Y)\propto\left[\frac{1}{\sigma}\right](\sigma^2)^{-(\frac{n+v_0}{2}+1)}\exp\frac{-1}{2\sigma^2}[v_0S_0^2+v\,S^2+n(\theta-\bar{Y})^2]$$

$$f(\theta,\sigma^2|Y)\propto\left[\frac{1}{\sigma}\right]\exp\frac{-1}{2\sigma^2}n(\theta-\bar{Y})^2(\sigma^2)^{-(\frac{n+v_0}{2}+1)}\exp\frac{-1}{2\sigma^2}[fS_B^2] \qquad ...(5)$$

Where: $fS_B^2=\sum(Y_{i0}-\bar{Y}_0)^2+\sum(Y_i-\bar{Y})^2$

$f=v_0+v=n+n_0-2$

$$\therefore\quad \hat{\sigma}_B^2=S_B^2=\frac{\sum(Y_{i0}-\bar{Y}_0)^2+\sum(Y_i-\bar{Y})^2}{n+n_0-2} \qquad ...(6)$$

إذن الصيغة (5) تمثل نواة التوزيع الطبيعي− معكوس كاما (*Normal−Inverse Gamma*).

بإجراء عملية تكامل الدالة (5) بالنسبة (θ) نحصل على الدالة الإحتمالية الحدية اللاحقة للمعلمة (σ^2) أو بهذه الطريقة (Chib et. al., 2008):

$$f(\sigma^2|Y)=\frac{f(\theta,\sigma^2|Y)}{f(\theta|Y,\sigma^2)}$$

$$=\frac{N\,IG[(\hat{\theta}),V(\hat{\theta}),\frac{f}{2},\frac{f\,S_B^2}{2}]}{N[(\hat{\theta}),V(\hat{\theta})]}\sim IG[\frac{f}{2},\frac{f\,S_B^2}{2}]$$

إذن $(\sigma^2|Y)$ لها توزيع معكوس كاما (*Inverse Gamma*).

8.2: الطريقة المقترحة:

تمَ إقتراح إستخدام مقدر تباين التوزيع النهائي لمعكوس كاما في الصيغة (6) لتقدير أوزان بيز كما في الصيغة الآتية:

$$WB_i = \frac{1}{\hat{\sigma}_B^2} = \frac{n+n_0-2}{\sum(Y_{i0}-\bar{Y}_0)^2+\sum(Y_i-\bar{Y})^2} \qquad ...(7)$$

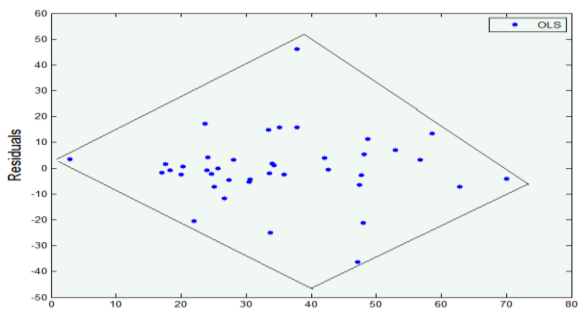
أوزان بيز WB_i المقدرة في الصيغة (7) سوف تستخدم كأوزان في طريقة المربعات الصغرى الموزونة (BWLS) لتقدير معلمات أنموذج الإنحدار الخطي الذي يعاني من مشكلة عدم تجانس تباين قيم الخطأ العشوائي.

3: الجانب التطبيقي:

لغرض تطبيق الطريقة المقترحة لأسلوب المربعات الصغرى الموزونة بأوزان بيز (*BWLS*) ومقارنتها مع الطريقة التقليدية (*WLS*) في معالجة مشكلة عدم تجانس تباين الخطأ العشوائي وتقدير النموذج الملائم للبيانات تناول البحث نوعان من البيانات تمثل الأولى دراسة محاكاة (محاكاة تجريبية) والثانية بيانات حقيقية (تطبيق عملي).

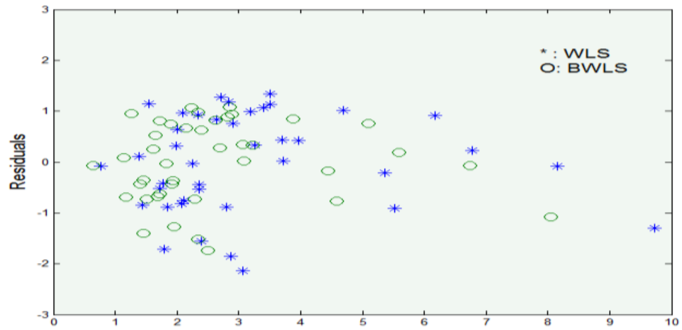
1.3: دراسة محاكاة:

للمقارنة بين طريقة أوزان بيز والطريقة التقليدية في معالجة مشكلة عدم تجانس تباين الخطأ العشوائي تمَ إجراء محاكاة بإستخدام لغة ماتلاب من خلال برنامج صمم لهذا الغرض على إفتراض حجم عينة مقدارها (n=40) مشاهدة قسمت إلى جزئين يتألف كل جزء منها (20) مشاهدة بحيث يُولد الخطأ العشوائي للجزء الأول من التوزيع الطبيعي بمعدل صفر وتباين متجانس يساوي واحد (توزيع طبيعي قياسي) في حين يتوزع الخطأ العشوائي للجزء الثاني التوزيع الطبيعي بمعدل صفر وإنحراف معياري مختلف لعدة حالات مقداره ($\sigma_1=15$، $\sigma_2=20$ و $\sigma_3=25$)، لذلك فإن العينة الكلية لديها مشكلة عدم تجانس تباين الخطأ العشوائي مع فرض شرط وجود عدم التجانس للبيانات المولدة من خلال حساب قيمة-F (وإهمال البيانات التي لاتعاني من هذه المشكلة) ومقارنتها مع قيمة-F المجدولة تحت مستوى المعنوية 5% ودرجات حرية ($v_1=16$ و $v_2=16$) بعد ترتيب البيانات تصاعدياً مع حذف $\left(\frac{n}{5}\right)$ من البيانات الوسطى المولدة. وتمَ إفتراض وجود ثلاث متغيرات مستقلة x_1، x_2 و x_3 مع متجه معلمات مختلفة β مقدارها (10، 5، 10 و15) و(20، 10، 20 و−10) وتم رسم التجربة الأولى للقيم المقدرة $\hat{y}$ ضد البواقي المقدرة بإستخدام طريقة المربعات الصغرى الإعتيادية كما يوضحه الشكل الآتي:



الشكل (1): الرسم البياني لقيم البواقي ضد القيم $\hat{y}$ لطريقة (OLS)

يبين الشكل (1) وجود زيادة ونقصان في تباين الخطأ يدل على عدم تجانس تباين الخطأ أي أن تباين البواقي غير ثابت. كما تمَ رسم نفس التجربة الأولى للقيم المقدرة $\hat{y}$ ضد البواقي المقدرة بإستخدام طريقة المربعات الصغرى الموزونة التقليدية والمقترحة بإستخدام أوزان بيز كما يوضحه الشكل الآتي:



الشكل (2): الرسم البياني لقيم البواقي ضد القيم $\hat{y}$ لطريقة (WLS) و(BWLS)

يبين الشكل (2) عدم وجود زيادة أو نقصان في تباين الخطأ والذي يدل على تجانس تباين الخطأ أي أن تباين البواقي ثابت.

لغرض المقارنة بين الطريقة المقترحة (BWLS) والتقليدية (WLS) تم تكرار التجربة (1000) مرة ولمختلف الحالات المفترضة وحساب عدد حالات نجاح معالجة مشكلة عدم تجانس تباين الخطأ العشوائي ومعدلات (RMSE، R^2 ، دالة الترجيح، AIC وBIC) ولخصت النتائج في الجدول الآتي:

الجدول (1): بعض معايير نتائج تجارب المحاكاة للطريقة التقليدية والمقترحة

الحالة	الطريقة	عدد حالات نجاح المعالجة	RMSE	R^2	دالة الترجيح	معيار AIC	معيار BIC
1	قيم المعلمات المفترضة عند $\sigma = 20$			10	5	10	15
	OLS	------	13.824	0.4484	-159.28	326.56	333.32
	WLS	745	1.2079	**0.6461**	-59.899	127.80	134.55
	BWLS	**781**	**1.1552**	0.5438	**-59.264**	**126.53**	**133.28**
2	قيم المعلمات المفترضة عند $\sigma = 20$			20	10	20	-10
	OLS	------	13.899	0.5610	-159.43	326.86	333.62
	WLS	531	1.1629	**0.6932**	-58.872	125.74	132.50
	BWLS	**538**	**0.9986**	0.6177	**-53.605**	**115.21**	**121.97**
3	قيم المعلمات المفترضة عند $\sigma = 25$			10	5	10	15
	OLS	------	17.270	0.3518	-168.19	344.37	351.66
	WLS	853	**1.3335**	**0.5291**	**-63.658**	**135.32**	**142.07**
	BWLS	**876**	1.3436	0.4374	-65.039	138.08	144.83
4	قيم المعلمات المفترضة عند $\sigma = 25$			20	10	20	-10
	OLS	------	17.316	0.4635	-168.25	344.49	351.25
	WLS	**698**	1.2606	**0.5977**	-61.766	131.53	138.29
	BWLS	691	**1.1603**	0.5210	**-59.392**	**126.78**	**133.54**
5	قيم المعلمات المفترضة عند $\sigma = 15$			10	5	10	15
	OLS	------	10.440	0.4695	-148.00	304.01	310.76
	WLS	654	1.3156	**0.6799**	-63.716	135.43	142.19
	BWLS	**788**	**1.1684**	0.5704	**-59.843**	**127.69**	**134.44**
6	قيم المعلمات المفترضة عند $\sigma = 15$			20	10	20	-10
	OLS	------	10.356	0.6924	-147.68	303.36	310.11
	WLS	311	1.0394	**0.7950**	-54.808	117.62	124.37
	BWLS	**352**	**0.8138**	0.7361	**-45.590**	**99.180**	**105.94**

من خلال الجدول (1) نلاحظ أن عدد حالات نجاح معالجة مشكلة عدم تجانس تباين الخطأ العشوائي (قيمة–F المحسوبة أقل من قيمتها الجدولية) كانت الأفضلية للطريقة المقترحة ولكل الحالات المفترضة ماعدا الحالة الرابعة (قيم المعلمات المفترضة عند $\sigma = 25$) كانت الأفضلية للطريقة التقليدية بمقدار (698) مقابل

(691) حالة نجاح للطريقة المقترحة، كما أن هنالك إنخفاض في عدد حالات النجاح بشكل عام للطريقتان مع زيادة قيمة المعلمات المفترضة وإنخفاض مقادير الإنحراف المعياري المفترض.

حسب المعيار جذر متوسط الخطأ التربيعي (RMSE) كانت الأفضلية للطريقة المقترحة لكل الحالات ما عدا الحالة الثالثة (قيم المعلمات المفترضة عند $\sigma = 25$) كانت الأفضلية للطريقة التقليدية بمقدار (1.3335) مقابل (1.3436) للطريقة المقترحة، وكلا الطريقتان كانت لهما قيم أفضلية واقل بكثير مقارنةً مع طريقة (OLS).

حسب المعيار معامل التحديد R^2 كانت الأفضلية للطريقة التقليدية لكل الحالات مقارنة مع الطريقة المقترحة، وكلا الطريقتان كانت لهما قيم أفضلية (وأقل بكثير) مقارنةً مع طريقة (OLS).

حسب المعيار دالة الترجيح، (AIC) و(BIC) كانت الأفضلية للطريقة المقترحة لكل الحالات ما عدا الحالة الثالثة (قيم المعلمات المفترضة عند $\sigma = 25$) كانت الأفضلية للطريقة التقليدية، وكلا الطريقتان كانت لهما قيم أفضلية (واقل بكثير) مقارنةً مع طريقة (OLS).

2.3: تطبيق بيانات حقيقية:

لغرض توضيح كيفية معالجة مشكلة عدم تجانس تباين الخطأ العشوائي باستخدام الاسلوب التقليدي واسلوب بيز (الطريقة المقترحة) لتقدير الاوزان في طريقة المربعات الصغرى الموزونة المستخدمة في تقدير معلمات نموذج الانحدار واجراء مقارنة بين الاسلوبين على بيانات حقيقية تم استخدام البيانات من المصدر (Montgomery and Peck, 1982) والتي تمثل المعدل الشهري لمبيعات سلعة ما (Y_i) في كل من (30) متجرا ومقدار نفقات الاعلانات السنوية عليها (X_i)، حيث تم تقسيم البيانات الى جزئين الجزء الاول يتكون من (20) مشاهدة كبيانات حالية (Likelihood) والجزء الثاني يتكون من (10) مشاهدات كبيانات أولية (Prior).

اختبار مشكلة عدم تجانس تباين الخطأ:

تم استخدام اختبار كولد–فليد لاختبار مشكلة عدم تجانس التباين أي إختبار الفرضية الآتية:

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

جدول (2): قيم *MSE* و *Cal. F* لإختبار كولد–فليد

	M	*Cal*	*Tab. F*(6,6,0.(
Sampl	260974	4	4
Sampl	1129580		

بما ان قيمة *F*– المحسوبة أكبر من *قيمتها المجدولة* لذلك ترفض فرضية العدم وتقبل الفرضية البديلة أي يوجد مشكلة عدم تجانس التباين. وعلى هذا الأساس سيتم معالجة مشكلة عدم التجانس من خلال مايلي:

1– الطريقة التقليدية (*WLS*):

تم استخدام الصيغة رقم (4) لتقدير الاوزان والتي تساوي مقلوب تباين الخطأ العشوائي في طريقة المربعات الصغرى الموزونة لتقدير معلمات نموذج الانحدار، بعد أن تم تقدير معلمات النموذج الآتي:

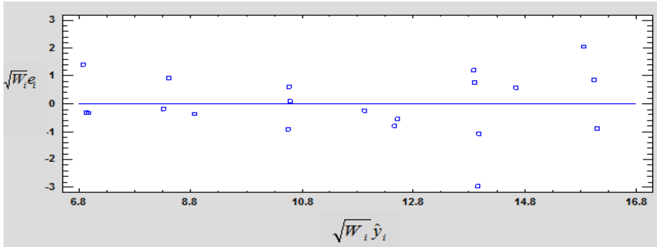
$$\hat{Y}_i = 4883486 + 8.232 X_i$$

جدول (3): تحليل التباين للنموذج المقدر باستخدام الاسلوب التقليدي لتقدير الاوزان

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*	R^2
Model	41545816772	1	41545816772	777.29	0.000	0.9774
Residual	962087830.9	18	53449323.94			
Total	42507904603	19				

يبين جدول رقم (3) تلخيص النتائج من خلال تحليل التباين للنموذج المقدر وقيمة (R^2) باستخدام الاسلوب التقليدي حيث تبين معنوية النموذج المقدر.

كشف مشكلة عدم تجانس تباين الخطأ للنموذج المعالج بالاسلوب التقليدي:



الشكل (3): الرسم البياني لقيم البواقي ضد القيم $\sqrt{W_i}\hat{y}_i$ لطريقة (WLS)

يبين الشكل (3) عدم وجود زيادة أو نقصان في تباين الخطأ والذي يدل على تجانس تباين الخطأ للنموذج المقدر باستخدام الاسلوب التقليدي (WLS).

2- الطريقة المقترحة (*BWLS*):

تم استخدام الصيغة رقم (7) لتقدير الاوزان والتي تساوي مقلوب تباين الخطأ العشوائي باستخدام اسلوب بيز في طريقة المربعات الصغرى الموزونة لتقدير معلمات نموذج الانحدار، والتي تم الحصول من خلالها على النموذج المقدر الاتي:

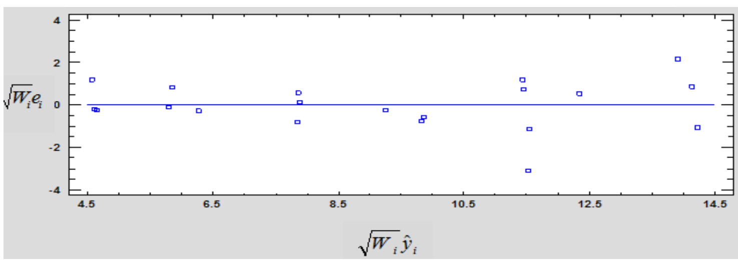
$$\hat{Y}_i = 48029.13 + 8.304X_i$$

جدول رقم (5): تحليل التباين للنموذج باستخدام اسلوب بيز في تقدير الاوزان

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*	R^2
Model	41210847802.93	1	41210847802.93			
Residual	961708928.39	18	53428273.8	771.33	0.000	0.9772
Total	42172556731.32	19				

يبين جدول رقم (5) تلخيص النتائج للنموذج المقدر باستخدام اوزان بيز حيث تبين معنوية النموذج المقدر.

كشف مشكلة عدم تجانس تباين الخطأ للنموذج المعالج باستخدام اسلوب بيز (*BWLS*):



الشكل (4): الرسم البياني لقيم البواقي ضد القيم $\hat{y}_i$ لطريقة (*BWLS*)

يبين الشكل (4) عدم وجود زيادة أو نقصان في تباين الخطأ والذي يدل على تجانس تباين الخطأ للنموذج المقدر باستخدام الطريقة المقترحة (*BWLS*).

المقارنة بين اسلوب بيز (الطريقة المقترحة) والطريقة التقليدية:

جدول رقم (7): قيم MSE و R^2 للنموذجين

R^2	MSE	الطريقة المستخدمة
0.9774	53449323.9	الاسلوب التقليدي لتقدير الاوزان في طريقة المربعات الصغرى الموزونة (WLS)
0.9772	53428273.8	اسلوب بيز لتقدير الاوزان في طريقة المربعات الصغرى الموزونة ($BWLS$)

يبيّن جدول رقم (7) أن قيمة متوسط مجموع مربعات الخطأ (MSE) لإسلوب بيز أقل من الاسلوب التقليدي والذي يؤكد على أفضلية اسلوب بيز (الطريقة المقترحة) في حساب الاوزان لطريقة المربعات الصغرى الموزونة في تقدير معلمات نموذج الانحدار ومعالجة مشكلة عدم تجانس تباين الخطأ العشوائي. ولكن حسب معامل التحديد (R^2) كانت الأفضلية للطريقة التقليدية بشكل طفيف جداً مقارنةً مع الطريقة المقترحة.

4: الإستنتاجات (Conclusions):

من خلال الجانب التجريبي (المحاكاة) والتطبيقي تمَ إستنتاج مايلي:

1- عدد حالات معالجة مشكلة عدم تجانس تباين قيم الخطأ العشوائي للطريقة المقترحة كانت أكثر من الطريقة التقليدية لجميع حالات المحاكاة ماعدا الحالة الرابعة (التي كانت أيضاً متقاربة).

2- كانت الأفضلية للطريقة المقترحة حسب المعيار RMSE، دالة الترجيح، AIC و BIC لكل حالات المحاكاة مقارنةً مع الطريقة التقليدية، ماعدا الحالة الثالثة كانت الأفضلية للطريقة التقليدية.

3- كانت الأفضلية للطريقة التقليدية حسب المعيار R^2 لكل حالات المحاكاة مقارنةً مع الطريقة المقترحة.

4- الجانب التطبيقي على بيانات حقيقية أكد نفس النتائج التي تمَ الحصول عليها من الجانب التجريبي.

5: التوصيات (Recommendations):

1- إعتماد الطريقة المقترحة في معالجة مشكلة عدم تجانس تباين قيم الخطأ العشوائي.

2- إجراء دراسات مستقبلية لمعالجة مشكلة عدم تجانس تباين الخطأ العشوائي بإستخدام أوزان بيز لبيانات لها توزيعات رياضية غير التوزيع الطبيعي مثل توزيع كاما وويبيل وغيرها.

3- إستخدام طرائق بيز أخرى مثل بيز المتسلسل وغيرها في حساب أوزان بيز لمعالجة مشكلة عدم تجانس تباين الخطأ العشوائي.

Reference

1. Ahmed, Kamran Hassan: (2013) "Using the Bayes Method to Estimate the Household Consumption Expenditure Function in Erbil Governorate", PhD thesis in Statistics, College of Administration and Economics, Salahaddin University, Erbil, pp. 37-45.
2. Al-Dulaimi, Muhammad Munajid and Kazem, Amori Hadi: (1988) “Introduction to Linear Regression Analysis”, University of Baghdad, pp. 292-305.
3. Al-Rawi, Khasha’ Mahmoud: 1987, “The Introduction to Regression Analysis,” University of Mosul, pp. 146-158.
4. Abdel Moneim, Tharwat Muhammad: (2005) “The Decline”, The Anglo Egyptian Bookshop, Cairo, pp. 173-185
5. Ali, Taha Hussein and Tara Ahmed Hassan, (2007) “Estimating the Logistic Model Using Sequential Base Weights”, Journal of Economic and Administrative Sciences\University of Baghdad, Vol. 13, No. 11.
6. Kazem, Amori Hadi and Muslim, Bassem Shailbeh: (2002) “Advanced Economic Measurement Theory and Practice”, Dar Al-Kutub and Documents, University of Baghdad.
7. Box, G.E.P, & Tiao, G. (1992): "Bayesian Inference in Statistical Analysis", John Wiley & Sons, New York.
8. Chib, S., Griffiths, W., Koop, G. & Terrell, D. (2008): "Bayesian Econometrics" Howard House, Wagon Lane, Bingley BD16 1WA, UK.
9. Montgomery, D. C. and E. A. Peck, 1982, "Introduction to Linear Regression Analysis", Wiley, New York.
10. Patrick J. Rosopa, Meline M. Schaffer and Amber N. Schroeder, (2003): "Managing Heteroscedasticity in General Linear Models", Psychological Methods, American Psychological Association, Vol. 18, No. 3, 335–351.
11. P. Richard Hahn, Jared S. Murray, and Carlos M. Carvalho, (2020): "Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, and Heterogeneous Effects": Bayesian) .

Using Bayes Weights To Remedy The Heterogeneity Problem Of Random Error Variance In Linear Models

Taha Hussein Ali Al-Zubaidi . Kamran Hassan Ahmed & Jimin Abu Bakr Omar

Department of Statistics and Informatics, College of Business and Economics, Salahaddin University, Erbil, Iraq

Abstract : In this research, it was suggested to use the Informative Bayes method in calculating the Bayes weights and use them to treat the of heterogeneity problem when estimating the linear regression model parameters using the weighted least squares method (BWLS). And compare it with the classical method through an experimental side to simulate the generated data from a normal distribution and for several different cases as well as an applied side of real data. The results of the research provided the preference of the proposed method on the classical method by relying on some statistical criteria through a program designed for this purpose in the language of MATLAB.

Keyword: regression Analysis, contrast inhomogeneity problem, biz style Weighted least squares method.