مجلة كلية التجارة للبحوث العلمية
كلية التجارة
جامعة الإسكندرية

العدد الأول– المجلد السادس والخمسون – يناير ٢٠١٩

أ / أحمـد قاروصــة

مدرس مساعد بقسم الإحصاء
والرياضة والتأمين
كلية التجارة– جامعة الاسكندرية

دراسة مقارنة لبعض مقدرات إنحدار ريدج المتينة

ملخص البحث

يعتبر وجود الازدواج الخطي Multicollinearity والمشاهدات الشاذة Outliers من أهم المشاكل التي تواجه تحليل الانحدار الخطي، خاصة عند وقوع المشكلتين آنيا. وإذا كان أسلوب انحدار "ريدج" يحسن من آثار مشكلة الازدواج الخطي، كما أن طرق الانحدار المتينة أقل تأثرا بالمشاهدات الشاذة إلا أن تلك الأساليب منفردة لا يمكنها أن تعالج المشكلتين معا. وتقترح الدراسة الحالية ثلاثة مقدرات جديدة من مقدرات انحدار "ريدج المتين" لعلاج مشكلتى الازدواج الخطي والمشاهدات الشاذة آنيا. والطرق المقترحة هى مزيج من مقدر جديد مقترح من مقدرات ريدج مع ثلاثة مقدرات متينة هى مقدر "أدنى مجموع مربعات مشذبة" LeastTrimmed(LTS) Squares ومقدر MM (MM-Estimator) ومقدر "أدنى مجموع بواقي مطلقة" Least Absolute Value (LAV). وتم إجراء دراسة محاكاة للمقارنة بين الطرق المقترحة وست من الطرق المتاحة باستخدام معيار "متوسط مربعات الخطأ" للمقدر MSE. وتمت مقارنة أداء المقدرات عند توليفات مختلفة من درجات الازدواج الخطي ونسب المشاهدات الشاذة. وأظهرت النتائج تفوق المقدرات المقترحة على الطرق الأخرى في حالة وجود كل من الازدواج الخطي والمشاهدات الشاذة آنيا.

Abstract

Besides multicollinearity, outliers also constitute a problem in the multiple linear regression analysis. We propose three new estimators of the robust ridge regression in the presence of multicollinearity and outliers, which called the Ridge Least Trimmed Squares, Ridge MM and Ridge Least Absolute Value estimator .For this purpose, a simulation study is conducted in order to see the difference

between the proposed methods and the existing methods in terms of their effectiveness measured by the mean squares error. The performances of the proposed methods are examined for different percentages of outliers and different degrees of multicollinearity. The results show that the proposed estimators are better than six of the existing methods in the presence of multicollinearity and outliers

[1] بحث مشتق من رسالة دكتوراه بعنوان: مقارنة بعض المقدرات المتينة المتحيزة لعلاج مشكلتى الازدواج الخطي والمشاهدات الشاذة في نموذج الانحدار الخطي: دراسة محاكاة. إشراف: أ.د امتثال محمد حسن، و أ.د محمد علي محمد أحمد.

١ – مقدمـــــة

يمثـل نمـوذج الانحـدار الخطـي، أحـد وسـائل دراسة العلاقة بين مُتغير اسْتِجَابَة Response (Y) وعـدد مـن المتغيـرات المفسـرة Explanatory Variables (X). ويمكن كتابة النموذج علي الشكل التالي:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (١)$$

حيث المصفوفة $X_{n\times p}$ كاملة الرتبة Full R-ank، مع p من المتغيرات المفسرة غير العشوائية Nonstochastic، و $\beta_{p\times 1}$ متجه المعالم المطلوب تقديرها، و $\varepsilon_{n\times 1}$ متجـه الخطـأ العشـوائي بمتوسـط يساوي الصفر، ومصفوفة تباين وتغاير Variance Covariance Matrix هى $\mathrm{Cov}(\varepsilon) = \sigma^2 I_n$، حيـث I_n مصـفوفة الوحـدة Identity، وقيمـة σ^2 مجهولة وترمز لتباين حد الخطأ العشوائي. وحين يتم "معايرة" Standardized X وY ، تكون المصفوفة (X^tX) هى مصفوفة الارتباط بين المتغيرات المفسرة والتـي تحتوي علـى معاملات الارتباط البسيطة بين كـل زوج مـن المتغيـرات، ويمثـل المتجـه (X^tY) الارتباطات بين متغير الاستجابة وكل متغير مفسر. وعـادة مـا يعتمـد علـى طريقـة المربعـات الصـغرى العادية Ordinary Least Squares (OLS) في تقدير نموذج الانحـدار الخطـي، وذلـك نظـرا لتمتـع مقـدر OLS بـبعض الخصـائص المثلـى فـي حالـة كـون المتغيـرات المفسـرة غيـر مرتبطـة مـع بعضـها البعض، وكون حدود الخطأ العشوائي مستقلة ولها نفس التوزيع المعتدل Independent and Iden-tically Distributed Normal. ويكون مقدر المربعـات الصـغرى العاديـة $\hat{\beta}$ ، مقـدر غيـر متحيـز لمتجه المعالم β، ويأخذ الشكل التالي:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} \qquad (٢)$$

وتبـاين مقـدر المربعـات الصـغرى يكـون علـى الشكل التالي:

$$Cov(\hat{\boldsymbol{\beta}}) = \boldsymbol{\sigma}^2(\mathbf{X}^t\mathbf{X})^{-1} \qquad (٣)$$

ويعتبـر الازدواج الخطـي Multicollinearity والمشاهدات الشاذة Outliers من أهم المشاكل التي تواجه تحليل الانحدار. فكثيرا ما يحدث عند توفيق نموذج الانحدار الخطي أن تكون المتغيرات المفسرة مرتبطـة مـع بعضـها البعض. **وتعرف هذه المشكلة "بالازدواج الخطي" (أو التعدد الخطي)** ، والتي لها آثـار ضـارة، مثل تضـخم تباينـات مقدرات معـاملات المربعـات الصـغرى وأيضـا تضـخم القـيم المطلقـة لمقدرات المعاملات نفسها(Habshah and Ma-rina, 2007). ويفترض نموذج الانحدار الخطـي التقليـدي عـدم وجـود ازدواج خطـي بـين المتغيـرات المفسرة، فإذا كان الازدواج الخطـي تاما، يكون من غيـر الممكـن حسـاب معـاملات الانحـدار وتكـون أخطــائهم المعياريــة لا نهائيــة. وإذا كــان الازدواج الخطــي أقـل مـن التـام، يكـون محـدد المصـفوفة (X^tX) قريبا من الصفر ويمكن حساب المعاملات إلا أنها تكون ذات أخطاء معيارية كبيرة، مما يعني عـدم وجـود دقـة فـي تقـدير المعـاملات أى تكـون التقـديرات غيـر مسـتقرة وبالتـالي تنشـأ أخطـاء فـي الاسـتدلال. وفـي هذه الحالـة تكون المشكلة الرئيسـة مـع مقدر المربعـات الصـغري هـي التقيد بضـرورة أن يكون المقدر"غير متحيز" ولذلك تم اقتراح أساليب بديلة للتقدير تسمح بقليل من التحيز، مما يؤدي إلـى زيــادة فــي اســتقرار المصــفوفة (X^tX) ، وبالتــالي تخفيض كبير في تباين المقدرات.

وبالرغم من وجود العديد من الطرق لعلاج مشكلة الازدواج الخطي، إلا أن أكثر تلك الطرق شيوعا هو مقدر انحدار "ريدج" Ridge Regression Estimator والذي يحسن من دقة معاملات الانحدار ويوفر تقديرات مستقرة للمعاملات مع سهولة في الحساب ، ويعتبر بديل لطريقة OLS (Samkar and Alpu, 2010).

والمشكلة الثانية التي يمكن أن تواجه نموذج الانحدار الخطي هى وجود المشاهدات الشاذة. ويمكن تعريف المشاهدات الشاذة على أنها "تلك المشاهدات التي تبدو غير متسقة مع بقية البيانات" (Barnett and Lewis, 1994). وغالبا ما توجد المشاهدات الشاذة في البيانات لأسباب كثيرة، منها أخطاء في التسجيل أو أخطاء في القياس، أو أن يكون مصدر البيانات مجتمعات مختلفة. أيضا فإن التوزيعات "ثقيلة الذيل" Heavy tail عادة ما ينتج عنها مشاهدات شاذة, وهذا النوع من المشاهدات الشاذة قد يكون له تأثير ملحوظ على تقديرات معالم نموذج الانحدار (Chatterjee and Hadi, 2006). وبغض النظر عن مصدر المشاهدات الشاذة، فإن وجود مشاهدة واحدة شاذة، يمكن أن تجعل من تقديرات المربعات الصغرى عديمة الفائدة، حيث تكون التقديرات متحيزة، مع تضخم في تباينات المقدرات معاملات الانحدار، أو تضخم تباين متوسط الاستجابة المقدر، أوقد يمتد أثرها لتضخيم كليهما معا. بالاضافة إلى التأثير السلبي على الاستدلال الإحصائي المرتبط بالنموذج.

وللتغلب على مشكلة المشاهدات الشاذة، يمكن استخدام طرق التقدير المتين Robust Estimation، والتي تتميز بأنها أقل حساسية من المربعات الصغرى تجاه القيم الشاذة. وقد أشار Ronchetti (1987) إلى أن الهدف من استخدام طرق التقدير المتين هو اختيار النموذج الذي يوفق غالبية البيانات، مع الأخذ في الاعتبار أن الأخطاء قد لا تتبع التوزيع المعتدل. وجدير بالذكر أنه قد تم اقتراح العديد من أساليب تقدير الانحدار المتين، والتي أثبت بعضها نجاحا في التطبيقات العملية. على سبيل المثال: مقدر "أدنى مجموع مربعات مشذبة" Least Trimmed Squares (LTS) (Rousseeuw, 1984,1985)، ومقدر MM (Yohai, 1987)، ومقدر "أدنى مجموع بواقي مطلقة" Least Absolute Value (LAV).

إن وقوع المشكلتين آنيا يمكن أن يحدث غالبا، مثلما يمكن أن تقع كل مشكلة بصورة منفردة. وإذا كان أسلوب انحدار "ريدج" يحسن من آثار مشكلة الازدواج الخطي، كما أن طرق الانحدار المتينة أقل تأثرا بالمشاهدات الشاذة إلا أن تلك الأساليب منفردة لا يمكنها أن تعالج المشكلتين معا. ولعلاج المشكلتين آنيا اقترحت العديد من مقدرات ريدج المتينة Robust Ridge Estimators والتي هى أقل تأثرا بالمشاهدات الشاذة والازدواج الخطي، والتي تعتمد على دمج أحد مقدرات ريدج مع واحد من المقدرات المتينة. ومن هذه الدراسات ما يلي: دراسة Holland (1973) التي تعتبر أول من قدم مدخل لتقدير دالة الانحدار في ظل المشكلتين معا من خلال استخدام انحدار "ريدج" المرجح مع اختيار متين للأوزان من خلال مقدر M. كما قدم Askin and Montgomery (1980) عائلة من المقدرات تجمع بين مقدر M المتين وكل من مقدر "ريدج" ومقدر Stein والمكونات الرئيسة Principal Components . وقدم Pfaffenberger and Dielman (1984, 1985) مقدراً جديداً يجمع بين

خصائص كل من مقدر LAV المتين ومقدر ريدج. واقترح Habshah and Marina (2007) مقدر "ريدج MM" من خلال مزج مقدر MM المتين مع انحدار ريدج. كذلك قدمت دراسة Samkar and Alpu (2010) طرقاً لانحدار ريدج المتينة بالاعتماد على مقدرات M و S و MM و GM. واقترحت دراسة Zahari et al. (2012) الجمع بين مقدر انحدار "ريدج المرجح، ومقدر MM المتين لتشكيل مقدر انحدار "ريدج" المرجح MM. واقترح Pati et al. (2014) دمج مقدر "ريدج" مع مقدر الانحدار المتين LTS لانتاج مقدر "ريدج المتين". واقترح Pati et al. (2016) مقدرين من مقدرات "ريدج المرجحة" باستخدام دالة الأوزان Bisquare وبالاعتماد على مقدرى الانحدار المتين LTS و LAV. واقترح Adegoke et al. (2016) دمج مقدر "ريدج" مع بعض المقدرات المتينة وهي مقدرات: M وMM وLTS وS، وذلك لتشكيل مقدرات "ريدج المتينة".

٢- مقدرات انحدار ريدج

إن سلوك معكوس المصفوفة X^tX هو الذي يجعل من الازدواج الخطي مشكلة. ففي ظل وجود الازدواج الخطي يؤدي أى تغير نسبي صغير في المصفوفة X^tX إلى تغير نسبي كبير في المصفوفة $(X^tX)^{-1}$. بالاضافة لذلك فإنه على الأقل هناك بعض العناصر القطرية الرئيسة للمصفوفة تكون كبيرة، مما يعني أن بعض عناصر مقدرات المربعات الصغرى سيكون لها تباين كبير. بناء على ما سبق، فمن الطبيعي التعرف على (أو تحديد) درجة الازدواج الخطي بالاعتماد على "حالة" المصفوفة Condition of a matrix. ويطلق على المصفوفة **X** غير المنفردة Nonsingular أنها "معتلة الحالة" ill-conditioned إذا كانت تغيرات نسبية صغيرة في X يمكن أن تُنتج تغيرات نسبية كبيرة في X^{-1}. وبالتالي يمكن استخدام "رقم الحالة" Condition Number (CN) كمقياس لدرجة "اعتلال الحالة" ill-condition أو لاكتشاف الازدواج الخطي (Groß, 2003, p.57). ويأخذ "رقم الحالة" الشكل التالي:

$$\boldsymbol{CN}(\mathbf{X}) = \sqrt{\frac{\boldsymbol{\lambda_{max}}}{\boldsymbol{\lambda_{min}}}} \qquad (٤)$$

حيث λ_{max} وλ_{min} هما أكبر وأصغر قيمة مميزة للمصفوفة (X^tX) على الترتيب. إذا كانت جميع القيم المميزة متساوية $(\lambda_{max} = \lambda_{min})$ فإن رقم الحالة يساوي الواحد الصحيح $(CN(X) = 1)$، مما يعني عدم وجود أى ازدواج خطي ويقال للمتغيرات المفسرة أنها متعامدة، ويطلق على المصفوفة **X** أنها "تامة الحالة" Perfect Conditioned. أما إذا كانت أصغر قيمة مميزة تساوي الصفر $(\lambda_{min} = 0)$، فإن رقم الحالة يصل إلى "ما لا نهاية"، مما يعني وجود ازدواج خطي تام. بالتالي فإن مدى "رقم الحالة" يكون: $CN(\mathbf{X}) \in [1, \infty)$. وكلما زادت قيمة $CN(\mathbf{X})$ أدى ذلك إلى "سوء" الحالة. ليس هناك خط واضح المعالم يصلح بشكل عام لتمييز "الحالة السيئة" للمصفوفة. وقد اقترح Belsley et al. (1980) اعتبار أن "رقم الحالة" بين 5 و10 مؤشرا على وجود ازدواج خطي ضعيف، بينما "رقم الحالة" أكبر من 30 يعد مؤشرا على وجود ازدواج خطي قوي. وأن رقم الحالة بين 10 و30 يشير إلى إزدواج خطي من متوسط إلى قوي. وأشار Belsley (1991) إلى أن الارتباط بين المتغيرات المفسرة يمكن النظر إليه على أنه مؤشر على الازدواج الخطي، ولكن العكس ليس بالضرورة

صحيح، بمعنى أن الازدواج الخطي يمكن أن يقع بين متغيرين مفسرين دون وجود ارتباط خطي بينهما.

يمكن تصور إجراءات مختلفة لعلاج الازدواج الخطي، أحدها بناء نموذج جديد وأخذ متغيرات مفسرة جديدة في الاعتبار لا ينتج عنها نفس القدر من الازدواج الخطي. وفي نفس الوقت يرى البعض (GroB, 2003, P. 308) أن الازدواج الخطي هو مشكلة مع البيانات وليست مع النموذج وبالتالي ليس هناك داع لحذف متغير ما. ولكن يمكن استخدام طرق أخرى للعلاج تتمثل في الحصول على مشاهدات اضافية (إن أمكن) أو دمج معلومات اضافية حول المعالم (إذا كانت متاحة) إلى النموذج. كذلك يمكن أيضا تقدير توليفات خطية معينة من المعالم بدلا من تقدير متجه المعالم بالكامل. ولكن جمع مزيد من البيانات غالبا ما يكون مكلفا أو غير قابل للتطبيق العملي في كثير من الحالات. كما أن إسقاط متغير أو أكثر من النموذج قد يؤدي إلى "التحيز"، وبالتالي قد يكون الحل أسوأ عما كان عليه الوضع من قبل. وبشكل عام، يعتبر مقدر انحدار "ريدج" (Hoerl and Kennard, 1970 a, b)، هو الأسلوب الأكثر شيوعا واستخداما في التعامل مع مشكلة الازدواج الخطي. وهو ينتمي إلى مدخل لعلاج مشكلة الازدواج الخطي يركز على إيجاد مقدرات لها أقل MSE مقارنة بمقدرات المربعات الصغرى. ويهدف مقدر ريدج إلي تقليل حجم وتباين تقديرات المربعات الصغرى، عن طريق إدخال كمية ضئيلة من التحيز. ويأخذ المقدر الصيغة التالية:

$$\widehat{\boldsymbol{\beta}}_R = (\mathbf{X}^t\mathbf{X} + \mathbf{K}\,\mathbf{I})^{-1}\mathbf{X}^t\mathbf{Y} \qquad (٥)$$

حيث $k \geq 0$، تسمى "معلمة ريدج" أو معلمة التحيز Biasing Parameter والتي يحددها الباحث. ولقد تم اقتراح العديد من الطرق لاختيار تلك المعلمة وتجدر الإشارة إلى أن عملية الاختيار الأمثل لمعلمة انحدار "ريدج" لا يمكن توفيرها من الناحية النظرية. وفيما يلي نستعرض بعض الطرق المستخدمة في الدراسات السابقة لايجاد تقدير لمعلمة ريدج (k)، **بالاضافة إلى طريقة جديدة مقترحة لتقديرها.**

وبافتراض أن $\Lambda = diag(\lambda_1, \lambda_2, \ldots\ldots \lambda_p)$ هى مصفوفة قطرية Diagonal عناصرها هى القيم المميزة للمصفوفة X^tX، و $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0)$. وبافتراض أن $D_{p\times p}$ مصفوفة متعامدة أعمدتها هى المتجهات المميزة Normalized Eigen Vectors المناظرة للقيم المميزة للمصفوفة X^tX. مما يجعل $D^tD = DD^t = I_p$ ، و $X^tX = D\,\Lambda\,D^t$. وبوضع $X^* = XD$ ، $\alpha = D^t\beta$ ، يمكن التعبير عن نموذج الانحدار الخطي (١) في الشكل "الأساسي" Canonical التالي (Hoerl and Kennard, 1970 a):

$$\mathbf{Y} = \boldsymbol{X}^*\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \qquad (٦)$$

ومع $\Lambda = X^{*t}X^*$، و $Cov(Y) = \sigma^2 I_n$ ، يصبح $\hat{\alpha}$ هو مقدر OLS لمتجه المعالم α كما يلي:

$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{\Lambda}^{-1}\boldsymbol{X}^{*t}\mathbf{Y} \qquad (٧)$$

ويصبح مقدرالمربعات الصغرى لتباين حد الخطأ العشوائي هو :

$$s^2 = (Y - X^*\hat{\alpha})^t (Y - X^*\hat{\alpha})/(n-p)$$
$$s^2 = \left(Y - X\hat{\beta}\right)^t \left(Y - X\hat{\beta}\right)/(n-p) \qquad (٨)$$

٢-١ دراسة Hoerl et al. (1975)

اقترحت الدراسة استخدام المقدر التالي:

$$\hat{K}_{HKB} = \frac{p\, s^2}{\sum_{i=1}^{p} \hat{\alpha}_i^2} = \frac{p\, s^2}{\sum_{i=1}^{p} \hat{\beta}_i^2} = \frac{p\, s^2}{\hat{\beta}^t\, \hat{\beta}} \qquad (٩)$$

٢-٢ دراسة MunizandKibria(2009)

طبقت الدراسة خوارزميات الوسط الهندسي والجذر التربيعي، على المداخل التي اقترحها كل من: Khalaf and Shukur (2005) و Kibria (2003) لايجاد سبعة مقدرات جديدة، **ونختار منهم المقدرين التاليين:**

$$\hat{K}_{MK1} = \left(\prod_{i=1}^{p} m_i\right)^{\frac{1}{p}} \qquad (١٠)$$

$$\hat{K}_{MK2} = median(m_i) \qquad (١١)$$

حيث $m_i = \sqrt{\frac{s^2}{\hat{\alpha}_i^2}}$.

٢-٣ دراسة Al-Hassan (2010)

اقترحت الدراسة إدخال تعديلات Alkhamisi and Shukur (2007) على مقدر Hocking et al. (1976) **ليصبح المقدر الجديد كما يلي:**

$$\hat{K}_{ALH} = s^2 \frac{\lambda_{max} \sum_{i=1}^{p} (\lambda_i \hat{\alpha}_i)^2}{\lambda_{max} (\sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2)^2} + \frac{1}{\lambda_{max}} \qquad (١٢)$$

٢-٤ دراسة Khalaf and Iguernane (2016)

اقترحت الدراسة أخذ الجذر التربيعي لمقدر Khalaf and Shukur (2005) كما يلي:

$$\hat{K}_{KSI} = \sqrt{\frac{\lambda_{max}\, s^2}{(n-p)s^2 + \lambda_{max}\, \hat{\alpha}_{max}^2}} \qquad (١٣)$$

٢-٥ مقدر مقترح لانحدار ريدج

تقترح الدراسة الحالية شكلا جديدا لمعلمة انحدار ريدج، هو الوسط الهندسي Geometric Mean (GM) لأربعة من مقدرات ريدج السابقة وهم: مقدرا Muniz and Kibria (2009)، $\hat{K}_{MK1}$ و $\hat{K}_{MK2}$ في المعادلتين (١٠) و(١١) على الترتيب. ومقدر Al-Hassan (2010)، $\hat{K}_{ALH}$ في المعادلة (١٢)، ومقدر Khalaf and Iguernane (2016)، $\hat{K}_{KSI}$ في المعادلة (١٣). وبالتالي فإن مقدر ريدج المقترح هو:

$$\hat{K}_{new} = GM\{\hat{K}_{MK1}, \hat{K}_{MK2}, \hat{K}_{ALH}, \hat{K}_{KSI}\} = \left(\hat{K}_{MK1} \times \hat{K}_{MK2} \times \hat{K}_{ALH} \times \hat{K}_{KSI}\right)^{\frac{1}{4}} \qquad (١٤)$$

٣- مقدرات الانحدار المتين Robust Regression Estimators

يمكن تعريف المشاهدة الشاذة في اتجاه متغير الاستجابة (Y) بأنها تلك المشاهدات التي تنحرف عن علاقة الانحدار الموفقة بغالبية البيانات (Rousseeuw and Leroy, 1987). كما أن المشاهدة قد تكون شاذة في فضاء المتغيرات المفسرة (X) ويطلق عليها نقاط الرفع Leverage Points.

وللمشاهدات الشاذة تأثير كبير على قيم معاملات نموذج الانحدار الموفق بطريقة OLS، بالاضافة إلى تأثيرها على الأخطاء المعيارية Standard Errors للمقدرات والاستدلال عن معالم النموذج. وُيستخدم الانحدار المتين لعلاج مشكة المشاهدات الشاذة حيث أن أقل تأثرا بتلك المشاهدات، وتعتمد طرق الانحدار المتين على مقدرات تخفض من أوزان المشاهدات الشاذة في توفيق نموذج الانحدار. وبالرغم من تعدد الطرق المتينة التي قُدمت لتقدير نموذج الانحدار، إلا أنه لا توجد طريقة – أو مقدر أو مدخل– واحدة تُعدُ الأفضل من جميع الجوانب. ويمكن تقييم ومقارنة مقدرات الانحدار المتين باستخدام عدد من المعايير المختلفة. وعلى الرغم من أن بعض من هذه بعض المعايير أكثر أهمية من البعض الأخر، فإن المقدر النموذجي يجب أن يحقق الخصائص الايجابية لجميع المعايير. ومن أهم تلك المعايير:

- **نقطةالتحطم Breakdown Point**

يمكن النظر لنقطة التحطم لمقدر ما علي أنها أقل نسبة من المشاهدات الشاذة يكون لها تأثير كبير علي المقدر محل الاهتمام (Hampel et al., 1986, p. 41). وقد تكون نقطة التحطم منخفضة حتي تصل إلى الصفر (أو يشار إليها بأنها تساوي $\frac{1}{n}$) وهو ما يعني أنه بوجود مشاهدة واحدة شاذة لن يكون للمقدر قدرة على اعطاء معلومات مفيدة عن المعلمة، وعلى الجانب الآخر يمكن أن تصل نقطة التحطم لحدها الأقصى وهو 50%، أي أن المقدر يمكن الاستفادة منه بالرغم من أن نصف المشاهدات شاذة.

- **التأثيرالمحدود**

Bounded Influence

هى خاصية مميزة لطريقة متينة، تشير إلى قدرتها على التحكم في مقدار التأثير لنقطة شاذة في اتجاه X (نقطة رفع). وتعتبر طريقة المربعات الصغرى الأكثر تأثرا بنقاط الرفع، ولكن بعض الطرق المتينة أيضا لها تأثير غير محدود. ويتم دراسة ما إذا كان المقدر له تأثير محدود أم لا من دراسة "دالة التأثير" Influence Function (Hampel, 1974) والتي تصف تأثير مشاهدة إضافية x_i على المقدر.

- **الكفاءة النسبية**

Relative Efficiency

ويعبر عنها كنسبة مئوية وتشير إلى الدرجة التي يؤدي بها المقدر مثل طريقة المربعات الصغرى في حالة كون الأخطاء تتبع التوزيع المعتدل. وتحسب الكفاءة بقسمة متوسط مربعات الخطأ لمقدر المربعات الصغرى على نظيره للمقدر. ومن المرغوب أن تكون للمقدر كفاءة كبيرة (قد تقترب من ٩٠%).

وفيما يلي بعض الدراسات السابقة في مقدرات الانحدار المتين التي تحقق بعض أو كل المعايير السابقة:

٣-١ مقدر أدنى مجموع بواقي مطلقة LeastAbsoluteValue(LAV)

ويمكن تعريفه على الصورة:

$$\min_{\hat{\beta}} \sum_{i=1}^{n} |r_i| \qquad (١٥)$$

وقد يطلق عليه اسم L1-norm أو LAD أو LAR (Rousseeuw and Leroy ,1987)

ويهدف المقدر إلى تدنية مجموع البواقي المطلقة $\sum_{i=1}^{n}|y_i - x_i^t \beta|$، وعادة يتم حله باستخدام البرمجة الخطية ، ويعتبر هذا المقدر مناسبا عندما يكون توزيع الأخطاء له "ذيول ثقيلة جدا" heavy very tails، أو غير متماثل أو عندما تكون العينة كبيرة جدا (Birkes and Dodge, 1993). وعلى الرغم من أن هذا المقدر متين تجاه القيم الشاذة في اتجاه Y، إلا أنه في حالة وجود مشاهدة ذات قيمة رفع عالية High Leverage فإنها تجذب خط الانحدار تجاهها أو تجعله يمر بها. لذلك يمكن لنقطة شاذة واحدة أن تطيح تماما بالمقدر، وبالتالي يكون للمقدر نقطة تحطم تساوي الصفر (أو $\frac{1}{n}$) (Wilcox, 2010). بالاضافة لذلك فإن المقدر يتمتع بكفاءة نسبية ٦٤% ، كما أن للمقدر تباين أكبر من تباين مقدر المربعات الصغرى بمقدار 1.57 ضعف (أو $\frac{\pi}{2}$).

٣-٢ مقدر "أدنى مجموع مربعات مشذبة Least Trimmed Squares (LTS)

قدم Rousseeuw (1984,1985) مقدر (LTS) كعلاج للبطء الشديد في معدل تقارب مقدر LMS. وفي ظل هذا المقدر، يتم تدنية مجموع أول عدد h من مربعات البواقي المرتبة تصاعديا، ويمكن صياغة دالة الهدف كما يلي:

$$\min \sum_{i=1}^{h} (r_i^2)_{i:n} \qquad (١٦)$$

حيث $r_{(1)}^2, r_{(2)}^2, \dots, r_{(n)}^2$ هي مربعات البواقي المرتبة تصاعدياً، وبالتالي فإن هذا المدخل مشابه لطريقة المربعات الصغرى فيما عدا أن المجموع لا يشمل أكبر مربعات البواقي، مما يسمح للتوفيق أن يتفادى المشاهدات الشاذة. وبشكل عام يمكن أن توضع قيمة h بين n و $n/2$. وقد اقترح Rousseew and Leroy, (1987) أن تكون قيمتها $h = [n(1-\alpha) + 1]$ ، حيث α هي نسبة "الشذب" Trimmed (النسبة المحتملة للمشاهدات الشاذة في العينة)، وتختار هذه النسبة بحيث تحقق الكفاءة النسبية للمقدر. وبصفة عامة، اذا كانت $h = [\frac{n+p+1}{2}]$، فإن نقطتي تحطم المقدرين LTS و LMS تتساويان. وعلى الرغم من أن المقدر له معدل تقارب مثل مقدرات M يساوي $n^{-\frac{1}{2}}$، وله نقطة تحطم عالية (50%)، إلا أنه يعاب عليه أنه أيضاً يتطلب جهداً حسابياً ضخماً مما يجعله أحيانا غير قابل للتطبيق العملي، خاصة إذا كان حجم العينة كبيرا. كما يعاب عليه انخفاض كفائته النسبية والتي تساوي٨%. لذلك قدم Rousseeuw and Driessen (2006) خوارزم (FAST– LTS)، الذي يتميز بالسرعة والقدرة علي التعامل مع البيانات كبيرة الحجم.

٣-٣مقدر MM–estimator

اقترح Yohai (1987) مقدرات MM للمرة الأولى، ومنذ ذلك الحين وهى تزداد انتشارا واستخداما، لكونها تجمع بين نقطة التحطم العالية، والكفاءة النسبية الجيدة التي تصل إلى ٩٥%، وأيضا لأنها مقاومة للمشاهدات الشاذة في اتجاه كل من متغير الاستجابة والمتغيرات المفسرة. ويرجع الاسم MM إلى حقيقة استخدام مقدر M في أكثر من مرحلة لحساب التقديرات النهائية. **ويتم الحصول على مقدر MM في ثلاث مراحل كما يلي:**

المرحلة الأولى: إيجاد المقدر المبدئي $\widehat{\boldsymbol{\beta}}_0$ لمتجه المعالم $\boldsymbol{\beta}$، والبواقي المبدئية المناظرة $r_i^{(0)}$. هذا المقدر المبدئي يجب أن يكون مقدر ذو

نقطة تحطم عالية وليس من الضروري أن يكون كفء. وعادة ما يُستخدم مقدر S أو مقدر LMS كمقدر مبدئي.

المرحلة الثانية: يتم إيجاد مقدر M لتشتت البواقي $(\hat{\sigma}_{(0)})$ باستخدام التقدير المبدئي وذلك بحل معادلات M لمعلمة التشتت التالية:

$$\frac{1}{n}\sum_{i=1}^{n}\rho_0\left(\frac{r_i^{(0)}}{\sigma}\right)=0.5 \qquad (١٧)$$

المرحلة الثالثة: وأخيرا، يتم الحصول على النهاية الصغرى المطلقة لما يلي:

$$L(\beta)=\sum_{i=1}^{n}\rho_1\left(\frac{r_i}{\hat{\sigma}_{(0)}}\right)\mathbf{x}_i=0 \qquad (١٨)$$

فإذا كان ψ_1 هي مشتقة ρ_1، فإن مقدر MM لمتجه معالم الانحدار يكون هو أى حل للمعادلة التالية:

$$\sum_{i=1}^{n}\psi_1(r_i/\hat{\sigma}_{(1)})\mathbf{x}_i=0 \qquad (١٩)$$

والذي يحقق المتباينة التالية:

$$L(\hat{\beta}_{MM})\leq L(\hat{\beta}_0)$$

وأوضح Yohai (1987) أنه إذا تحققت المتباينة السابقة, فإن نقطة تحطم $\hat{\beta}_{MM}$ (مقدر MM) أكبر من أو تساوي فإن نقطة تحطم $\hat{\beta}_0$ (المقدر المبدئي). ومن الواضح أن مقدر MM يعتمد على دالتين مختلفتين في البواقي هما: ρ_0 و ρ_1 وذلك لتحديد نقطة التحطم والكفاءة. ويجب أن تكون $\rho_1(x)\leq\rho_0(x)$، حيث:

$$\rho_1(r)=\rho\left(\frac{r}{c_1}\right),\quad \rho_0(r)=\rho\left(\frac{r}{c_0}\right)$$

ولكى تكون $\rho_1\leq\rho_0$ ، يجب أن تُختار $c_0\leq c_1$. وأوضح Maronna et al. (2006) أن القيم الأكبر من c_1 تؤدي إلى زيادة الكفاءة.

٤ – مقدرات ريدج المتينة

على الرغم من أن أداء مقدر انحدار ريدج يكون جيدا في حالة وجود الازدواج الخطي، إلا أنه غير متين في ظل وجود المشاهدات الشاذة. وحيث أن مقدرات ريدج ومقدرات الانحدار المتينة غير قادرة على التعامل مع مشكلتى المشاهدات الشاذة والازدواج الخطي آنيا، بالتالي يكون من الأفضل مزج مقدر ريدج مع بعض المقدرات المتينة لانتاج مقدرات انحدار "ريدج المتينة". وهناك العديد من الدراسات السابقة التي قدمت بعض مقدرات انحدار ريدج المتينة. والتي سنتناول ثلاثة منها. كما نقدم ثلاثة مقدرات أخرى مقترحة تعتمد على دمج مقدر انحدار ريدج باستخدام معلمة ريدج المقترحة (k_{new}) مع ثلاثة مقدرات متينة هى LTS وMM وLAV كما يلي:

٤–١مقدر "Ridge LTS" (RLTS0)

اقترح Pati et al. (2014) دمج مقدر "ريدج" مع مقدر الانحدار المتين LTS لانتاج مقدر "ريدج المتين" RLTS يتشكل مقدر انحدار Ridge LTS (ويتم الاشارة إليه بالرمز RLTS0) من دمج مقدر ريدج مع المقدر المتين LTS.

ويأخذ المقدر الصيغة التالية:

$$\hat{\boldsymbol{\beta}}_{RLTS0}=(\mathbf{X^tX}+\mathbf{K}^*_{\mathbf{LTS}}\ \mathbf{I})^{-1}\mathbf{X^t}Y \qquad (٢٠)$$

وتأخذ $\mathbf{K}^*_{\mathbf{LTS}}$ الصيغة التالية

$$\mathbf{K}^*_{\mathbf{LTS}}=\frac{ps^2_{LTS}}{\hat{\boldsymbol{\beta}}^{\mathbf{t}}_{\mathbf{LTS}}\ \hat{\boldsymbol{\beta}}_{\mathbf{LTS}}} \qquad (٢١)$$

حيث $\hat{\beta}_{LTS}$ هو مقدر LTS المتين لمتجه معالم الانحدار β، و s^2_{LTS}

$$s^2_{LTS} = \frac{(Y - X\,\widehat{\beta}_{LTS})^t (Y - X\,\widehat{\beta}_{LTS})}{n-p} \quad (٢٢)$$

ومـن الجـدير بالـذكر أن K^*_{LTS} هـي نفسـها k_{HKB} فـي المعادلـة (٩) مـع اختلافـين همـا: أولا استخدام مقدر LTS في تقدير متجه المعالم β بدلا من طريقة OLS، وثانيا يتم تعديل s^2 في المعادلة (٨) باستخدام تقديرات LTS بدلا من طريقة OLS. وتهدف هذه التغييرات إلى تخفيض تأثير المشاهدات الشاذة علي القيمة المختارة لمعلمة ريدج.

٤-٢ مقدر "Ridge MM" (RMM0)

اقترح Habshah and Marina (2007) مقدر "ريدج MM" من خلال مزج مقدر MM المتين مع انحدار ريدج. ويأخذ هذا المقدر الشكل التالي وكما سـبق نحصـل علـى مقـدر Ridge MM (ويـتم الاشارة إليه بالرمز RMM0) من دمج مقدر ريدج مـع المقـدر المتـين MM. **ويأخـذ المقـدر الصـيغة التالية:**

$$\widehat{\beta}_{MM0} = (X^tX + K^*_{MM}\ I)^{-1}X^tY \quad (٢٣)$$

وتأخذ K^*_{MM} الصيغة التالية

$$K^*_{MM} = \frac{ps^2_{MM}}{\widehat{\beta}^t{}_{MM}\,\widehat{\beta}_{MM}} \quad (٢٤)$$

حيث $\widehat{\beta}_{MM}$ هو مقدر MM المتين لمتجه معالم الانحدار β، و s^2_{MM} هو:

$$s^2_{MM} = \frac{(Y - X\,\widehat{\beta}_{MM})^t (Y - X\,\widehat{\beta}_{MM})}{n-p} \quad (٢٥)$$

٤-٣مقدر "Ridge LAV" (RLAV0)

قدم Pfaffenberger and Dielman (1984, 1985) مقدراً جديداً (RLAV) يجمع بين خصائص كل من مقدر LAV المتين ومقدر ريدج. ويأخـذ مقـدر RLAV الصـيغة التاليـة وكمـا سـبق يتكون مقدر انحدار Ridge LAV (ويتم الاشارة إليـه بـالرمز RLAV0) مـن مـزج مقـدر ريـدج مـع المقدر المتين LAV. ويأخذ المقدر الصيغة التالية:

$$\widehat{\beta}_{RLAV0} = (X^tX + K^*_{LAV}\ I)^{-1}X^tY \quad (٢٦)$$

وتأخذ K^*_{LAV} الصيغة التالية

$$K^*_{LAV} = \frac{ps^2_{LAV}}{\widehat{\beta}^t{}_{LAV}\,\widehat{\beta}_{LAV}} \quad (٢٧)$$

حيث $\widehat{\beta}_{LAV}$ هو مقدر LAV المتين لمتجـه معالم الانحدار β، و s^2_{LAV}

$$s^2_{LAV} = \frac{(Y - X\,\widehat{\beta}_{LAV})^t (Y - X\,\widehat{\beta}_{LAV})}{n-p} \quad (٢٨)$$

٤-٤ مقدر "Ridge LTS" المقترح (RLTS1)

يتشكل مقدر انحدار Ridge LTS المقترح (ويتم الاشارة إليه بالرمز RLTS1) من دمج مقدر ريدج باستخدام معلمة ريدج الجديدة k_{new} مع المقدر المتين LTS. **ويأخذ المقدر الصيغة التالية:**

$$\widehat{\beta}_{RLTS1} = (X^tX + K^*_{new(LTS)}\ I)^{-1}X^tY \quad (٢٩)$$

وتأخذ K^*_{LTS1} الصيغة التالية

$$K^*_{new(LTS)} = (\widehat{K}^*{}_{MK1} \times \widehat{K}^*{}_{MK2} \times \widehat{K}^*{}_{ALH} \times \widehat{K}^*{}_{KSI})^{\frac{1}{4}} \quad (٣٠)$$

وتختلف $K^*_{new(LTS)}$ عن k_{new} في استخدام الأولـى لطريقة LTS المتينـة فـي ايجـاد المقـدرات $\widehat{\beta}$

(ومن ثم $\widehat{\alpha}$) و s^2، بدلا من طريقة OLS، وبالتالي المقدرات المتينة لمعلمة ريدج: $\widehat{K}^*_{MK1}$ و $\widehat{K}^*_{MK2}$ و $\widehat{K}^*_{ALH}$ و $\widehat{K}^*_{KSI}$ ، فــي المعــادلات (١٠و ١١ و١٢ و١٣).

٤-٥ مقــدر "Ridge MM" المقتـــرح (RMM1)

يتشكل مقدر انحدار Ridge MM المقترح (ويتم الاشارة إليه بالرمز RMM1) من دمج مقدر ريدج باستخدام معلمة ريدج الجديدة k_{new} مع المقدر المتين MM. ويأخذ المقدر الصيغة التالية:

$$\widehat{\boldsymbol{\beta}}_{RMM1} = \left(\mathbf{X}^{t}\mathbf{X} + \mathbf{K}^{*}_{\mathbf{new(MM)}}\ \mathbf{I}\right)^{-1} \mathbf{X}^{t}\boldsymbol{Y} \qquad (٣١)$$

وتأخذ $\mathbf{K}^*_{\mathbf{MM1}}$ الصيغة التالية

$$\mathbf{K}^{*}_{\mathbf{new(MM)}} = \left(\widehat{\mathbf{K}}^{*}_{MK1} \times \widehat{\mathbf{K}}^{*}_{MK2} \times \widehat{\mathbf{K}}^{*}_{\mathbf{ALH}} \times \widehat{\mathbf{K}}^{*}_{\mathbf{KS}I}\right)^{\frac{1}{4}} \qquad (٣٢)$$

وتختلف $K^*_{new(MM)}$ عن k_{new} في استخدام الأولى لطريقة MM المتينة في ايجاد المقدرات $\widehat{\beta}$ (ومن ثم $\widehat{\alpha}$) و s^2، بدلا من طريقة OLS، وبالتالي المقدرات المتينة لمعلمة ريدج: $\widehat{K}^*_{MK1}$ و $\widehat{K}^*_{MK2}$ و $\widehat{K}^*_{ALH}$ و $\widehat{K}^*_{KSI}$ ، فــي المعــادلات (١٠ و ١١ و١٢ و١٣).

٤-٦ مقدر "Ridge LAV" المقترح (RLAV1)

يتشكل مقدر انحدار Ridge LAV المقترح (ويتم الاشارة إليه بالرمز RLAV1) من دمج مقدر ريدج باستخدام معلمة ريدج الجديدة k_{new} مع المقدر المتين LAV. ويأخذ المقدر الصيغة التالية:

$$\widehat{\boldsymbol{\beta}}_{RLAV1} = \left(\mathbf{X}^{t}\mathbf{X} + \mathbf{K}^{*}_{\mathbf{new(LAV)}}\ \mathbf{I}\right)^{-1} \mathbf{X}^{t}\boldsymbol{Y} \qquad (٣٣)$$

وتأخذ $\mathbf{K}^*_{\mathbf{LAV1}}$ الصيغة التالية

$$\mathbf{K}^{*}_{\mathbf{new(LAV)}} = \left(\widehat{\mathbf{K}}^{*}_{MK1} \times \widehat{\mathbf{K}}^{*}_{MK2} \times \widehat{\mathbf{K}}^{*}_{\mathbf{ALH}} \times \widehat{\mathbf{K}}^{*}_{\mathbf{KS}I}\right)^{\frac{1}{4}} \qquad (٣٤)$$

وتختلــف $K^*_{new(LAV)}$ عــن k_{new} فــي استخدام الأولى لطريقة LAV المتينة في ايجاد المقدرات $\widehat{\beta}$ (ومن ثم $\widehat{\alpha}$) و s، بدلا من طريقة OLS، وبالتالي المقدرات المتينة لمعلمة ريدج: $\widehat{K}^*_{MK1}$ و $\widehat{K}^*_{MK2}$ و $\widehat{K}^*_{ALH}$ و $\widehat{K}^*_{KSI}$ ، في المعادلات (١٠ و ١١ و١٢ و١٣).

٥- دراسة المحاكاة

تهدف دراسة المحاكاة إلى المقارنة بين مقدرات ريدج المتينة المقترحة (٣ مقدرات) وبعض المقدرات البديلة (٦ مقدرات). وقد تم تصميم دراسة المحاكاة بحيــث تســمح بوجــود كــلا مــن الازدواج الخطــي والمشـاهدات الشـاذة آنيا. **وبالتــالي تشــتمل دراســة المحاكاة على تسعة مقدرات هم:**

١) مقـدر المربعـات الصـغرى العاديـة (OLS) فـي معادلة (٢)
٢) مقـدر الانحـدار المتـين (MM)، المعـرف فـي المعادلة (١٩).
٣) مقدر انحدار ريدج (Ridge) في معادلـة (٥)، وباستخدام k في معادلة (٩).
٤) مقــدر انحــدار ريــدج المتــين (RLTS0) فــي معادلة (٢٠)، مع k^* من معادلة (٢١).
٥) مقدر انحدار ريدج المتين (RMM0) في معادلة (٢٣)، مع k^* من معادلة (٢٤).
٦) مقــدر انحــدار ريــدج المتــين (RLAV0) فــي معادلة (٢٦)، مع k^* من معادلة (٢٧).
٧) مقدر انحدار ريدج المتين المقترح (RLTS1) في معادلة (٢٩)، مع k^*_{new} من معادلة (٣٠).

٨) مقدر انحدار ريدج المتين المقترح (RMM1) في معادلة (٣١)، مع k_{new}^{*} من معادلة (٣٢).

٩) مقدر انحدار ريدج المتين المقترح (RLAV1) في معادلة (٣٣)، مع k_{new}^{*} من معادلة (٣٤).

بافتراض نموذج الانحدار الخطي في (١)، يتم أولا توليد المصفوفة X وفقا لطريقة **McDonald andGalarneau (1975)**، بحيث تكون عناصر المصفوفة $X^tX = \{\omega_{ij}\}$ على الشكل التالي:

$$\omega_{ij} = \begin{cases} 1 & i = j \\ \rho & i \neq j \end{cases} \qquad (٣٥)$$

ويتم توليد متجهات المصفوفة X كمايلي:

$$X_j = (1-\rho)^{1/2} Z_j + \rho^{1/2} Z_{p+1} \qquad j = 1, 2, \dots, p \qquad (٣٦)$$

حيث $Z_1, Z_2, \dots, Z_{p+1}$ متجهات من المتغيرات المعتدلة المعيارية المستقلة، و (ρ) هو معامل الارتباط البسيط بين كل زوج من المتغيرات المفسرة. ويتم اختيار متجه معالم الانحدار (β) بحيث يكون هو المتجه المميز المناظر لأكبر قيمة مميزة للمصفوفة X^tX بشرط أن $\beta^t\beta = 1$ ، أى طوله يساوي الوحدة. حيث يكون MSE يكون عند حده الأدنى وذلك كما أشار Newhouse and Oman (1971) . كما يتم توليد المشاهدات الشاذة (أو الشاذة في y) من خلال حد الخطأ العشوائي ε. فإذا كانت نسبة المشاهدات الشاذة هى α (حيث $\alpha = 0\%, 5\%, 10\%, 15\%$)، تقوم الدراسة بتوليد حدود الخطأ العشوائي من توزيع "معتدل ملوث" Contaminated Normal له دالة الكثافة التالية:

$$f(\varepsilon) = (1-\alpha)N(0,1) + \alpha\, N(0,100) \qquad (٣٧)$$

حيث $(1-\alpha)$ تمثل نسبة حدود الخطأ غير الشاذة. وهذا يعني أن أغلبية حدود الخطأ تأتي من توزيع معتدل معياري، ونسبة صغيرة تأتي من من توزيع معتدل وسطه الصفر وتباينه 100. وهكذا فالأخطاء العشوائية لها احتمال $(1-\alpha)$ أن تأتي من توزيع معتدل معياري، واحتمال (α) أن تأتي من توزيع معتدل بمتوسط صفر وتباين 100.

وقد تم تنفيذ المحاكاة باستخدام أربعة عوامل، من المتوقع أن يكون لها تأثير على متوسط مربعات الخطأ. **وهذه العوامل والمستويات هي:**

١- معامل الارتباط (ρ) بين كل زوج من المتغيرات المفسرة، مع أربعة مستويات هم: 0.10 و 0.80 و 0.90 و 0.99. وهى تعكس تدرج تصاعدي لمشكلة الازدواج الخطي بداية من عدم وجود المشكلة عند ارتباط 0.10 وصولا إلى ازدواج خطي شديد عند ارتباط 0.99.

٢- نسبة المشاهدات الشاذة مع أربعة مستويات هي: 0% و 5% و 10% و 15%.

٣- عدد المتغيرات المفسرة (p) مع مستويين هما 4 و 8.

٤- حجم العينة (n) مع مستويين هما 20 و 40.

وبالتالي يكون إجمالي توليفات دراسة المحاكاة 64 توليفة $(4 \times 4 \times 2 \times 2)$.

ولكل توليفة من توليفات المحاكاة تم توليد المصفوفة X ومن ثم معايرتها بحيث تكون X^tX في شكل مصفوفة ارتباط. ثم يتم توليد متجه عشوائي من توزيع معتدل Normal $N(0, \sigma^2)$ ليمثل حد الخطأ العشوائي.

ويتم إيجاد المتجه Y من نموذج الانحدار الخطي في المعادلة (١). عند كل توليفة يتم توليد حد الخطأ العشوائي ومن ثم المتجه Y 10,000 مرة (معاودة) Replications. ويتم حساب متوسط مربعات الخطأ للمقدر $\hat{\beta}_R$ كما يلي:

$$MSE(\widehat{\boldsymbol{\beta}}_{\mathbf{R}}) = \frac{\mathbf{1}}{\mathbf{10000}} \sum_{i=1}^{10000} \left[(\widehat{\boldsymbol{\beta}}_{\mathbf{R}} - \boldsymbol{\beta})^{\mathbf{t}} (\widehat{\boldsymbol{\beta}}_{\mathbf{R}} - \boldsymbol{\beta}) \right] \qquad (٣٨)$$

▪ نتائج دراسة المحاكاة

تم وضع نتائج المحاكاة في الجدول (١)، والذي يعطي قيم MSE للمقدرات عند توليفات المحاكاة والتي يبلغ عددها 64 توليفة، عند القيم المختلفة من ρ و O و p و n . كذلك تم اعطاء رتب Ranks للمقدرات تصاعديا داخل كل توليفة وفقا لقيم MSE (في جدول ٢)، بحيث يأخذ المقدر ذو أقل MSE الرتبة 1 وهكذ حتى تكون رتبة المقدر الذي له أكبر MSE هى 9. أيضا تستخدم الدراسة الأعمدة البيانية لبعض التوليفات حيث تكون بعض عوامل المحاكاة ثابتة في حين تتغير باقي العوامل، وذلك لاستخلاص النتائج عند كل عامل من عوامل المحاكاة. وأيضا تستخدم الدراسة الصناديق البيانية لقيم MSE للمقدرات عند جميع التوليفات، ولرتب المقدرات، وذلك بغرض استخلاص نتائج عامة من دراسة المحاكاة.

٥-١ أداء المقدرات وفقا لدرجة الارتباط بين المتغيرات المفسرة ρ:

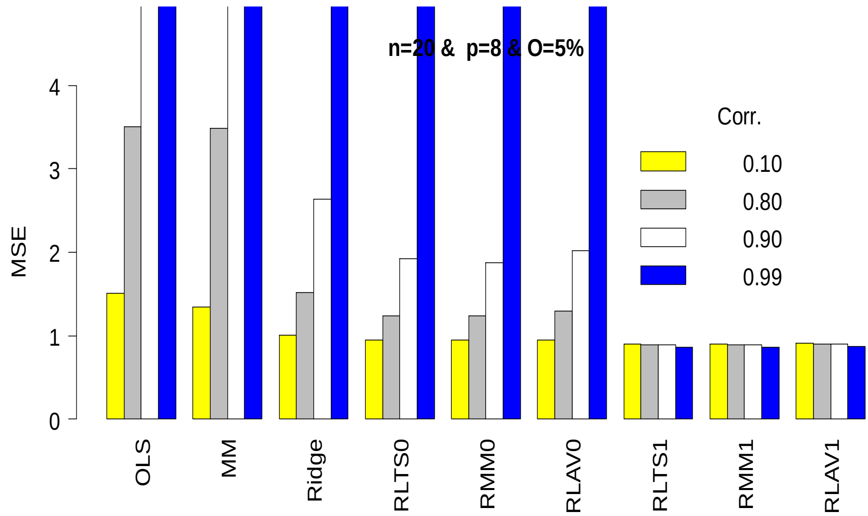
من جدول (١) وبالاستعانة بالشكلين (١) و(٢) يمكن أن نستنتج أن زيادة درجة الارتباط بين المتغيرات المفسرة له تأثير سلبي على المقدرات السابقة (غير المقترحة) يتمثل في ارتفاع قيم MSE لتلك المقدرات، ويزداد هذا الارتفاع بشكل كبير جدا عند درجة ارتباط 0.99 (أى عند ازدواج خطي شديد). كما نلاحظ عدم التأثير السلبي لزيادة درجة الارتباط على المقدرات المقترحة، بل بالعكس قد يصاحب ذلك انخفاض، ولو طفيف، في قيم MSE وبشكل عام نجد أن المقدرات المقترحة تتفوق دائما على باقي المقدرات في حالة وجود ازدواج خطي (عند درجة ارتباط 0.80 أو 0.90 أو 0.99) عند جميع التوليفات المناظرة، وأن هذا التفوق يكون بارزا عند درجة ارتباط 0.99. وفي حالة عدم وجود ازدواج خطي (عند $\rho = 0.10$) نجد أن المقدرات المقترحة ليست دائما هى الأفضل، ففي بعض التوليفات يكون مقدر MM المتين هو الأفضل. وعند عدم الازدواج الخطي وكذلك المشاهدات الشاذة مع حجم عينة 40 وعدد متغيرات 4 يكون أفضل ثلاثة مقدرات هي: OLS يليها مقدر MM ثم مقدر RLTS0 (وهى حالة وحيدة استثنائية، حيث لا وجود للمقدرات المقترحة ضمن أفضل ثلاث مقدرات).

جدول (١): قيم MSE للمقدرات عن جميع توليفات المحاكاة

ρ	O	n	p	OLS	MM	Ridge	RLTS0	RMM0	RLAV0	RLTS1	RMM1	RLAV1
0.10	0	20	4	0.8131	0.8396	0.8151	0.8084	0.8140	0.8117	0.8057	0.8083	0.8070
0.10	0	20	8	1.1179	1.3832	0.8783	0.8708	0.8753	0.8728	0.8369	0.8382	0.8376
0.10	0	40	4	0.8139	0.8222	0.8471	0.8390	0.8463	0.8412	0.8398	0.8428	0.8410
0.10	0	40	8	0.9749	1.0120	0.8717	0.8718	0.8713	0.8708	0.8565	0.8570	0.8564
0.10	0.05	20	4	0.9793	0.8913	0.8916	0.8784	0.8809	0.8773	0.8671	0.8671	0.8669
0.10	0.05	20	8	1.5027	1.3429	1.0079	0.9469	0.9479	0.9496	0.9017	0.9006	0.9044
0.10	0.05	40	4	0.9349	0.8790	0.9267	0.9284	0.9332	0.9274	0.9201	0.9218	0.9200
0.10	0.05	40	8	1.0969	0.9606	0.9643	0.9460	0.9466	0.9455	0.9406	0.9391	0.9407
0.10	0.10	20	4	1.0169	0.8914	0.9456	0.9310	0.9330	0.9307	0.9251	0.9251	0.9260
0.10	0.10	20	8	1.4262	1.1789	1.0381	0.9714	0.9706	0.9745	0.9529	0.9509	0.9573
0.10	0.10	40	4	0.9842	0.9050	0.9679	0.9661	0.9681	0.9644	0.9591	0.9586	0.9583
0.10	0.10	40	8	1.1548	0.9541	0.9945	0.9655	0.9660	0.9646	0.9624	0.9599	0.9632
0.10	0.15	20	4	1.0710	0.9269	0.9776	0.9569	0.9595	0.9584	0.9495	0.9490	0.9511
0.10	0.15	20	8	1.5462	1.2892	1.0902	1.0024	1.0000	1.0064	0.9790	0.9761	0.9835
0.10	0.15	40	4	1.0247	0.9278	0.9805	0.9761	0.9773	0.9735	0.9661	0.9661	0.9656
0.10	0.15	40	8	1.1657	0.9588	1.0117	0.9807	0.9807	0.9794	0.9803	0.9789	0.9822
0.80	0	20	4	1.4401	1.5337	0.9222	0.9416	0.9245	0.9346	0.7430	0.7431	0.7431
0.80	0	20	8	2.0618	2.6239	1.0711	1.0520	1.0634	1.0625	0.8043	0.8062	0.8058
0.80	0	40	4	1.1354	1.1701	0.8734	0.8894	0.8750	0.8833	0.7918	0.7919	0.7914
0.80	0	40	8	1.5341	1.6465	0.9560	1.0086	0.9633	0.9840	0.8352	0.8256	0.8301
0.80	0.05	20	4	1.3326	1.1201	0.9588	0.9025	0.8984	0.9054	0.8350	0.8349	0.8362
0.80	0.05	20	8	3.5059	3.4864	1.5133	1.2405	1.2379	1.2935	0.8919	0.8885	0.9026
0.80	0.05	40	4	1.1096	0.9588	0.9419	0.9110	0.9097	0.9111	0.8894	0.8895	0.8894
0.80	0.05	40	8	1.7087	1.2560	1.0857	0.9786	0.9492	0.9727	0.9043	0.8961	0.9034
0.80	0.10	20	4	1.5166	1.1420	1.0721	0.9576	0.9480	0.9670	0.8911	0.8900	0.8941
0.80	0.10	20	8	3.6091	2.8962	1.5557	1.1996	1.1731	1.2280	0.9354	0.9284	0.9431
0.80	0.10	40	4	1.1916	0.9598	0.9900	0.9420	0.9420	0.9424	0.9271	0.9273	0.9270
0.80	0.10	40	8	2.0267	1.1910	1.2037	0.9838	0.9583	0.9856	0.9333	0.9268	0.9351
0.80	0.15	20	4	1.9957	1.2527	1.2111	0.9908	0.9885	1.0138	0.9149	0.9152	0.9185
0.80	0.15	20	8	6.2917	3.9141	2.1876	1.4482	1.4215	1.5090	0.9625	0.9592	0.9717
0.80	0.15	40	4	1.4955	1.0140	1.0893	0.9590	0.9558	0.9643	0.9417	0.9416	0.9421
0.80	0.15	40	8	1.8058	1.0905	1.1537	0.9691	0.9598	0.9762	0.9523	0.9488	0.9567

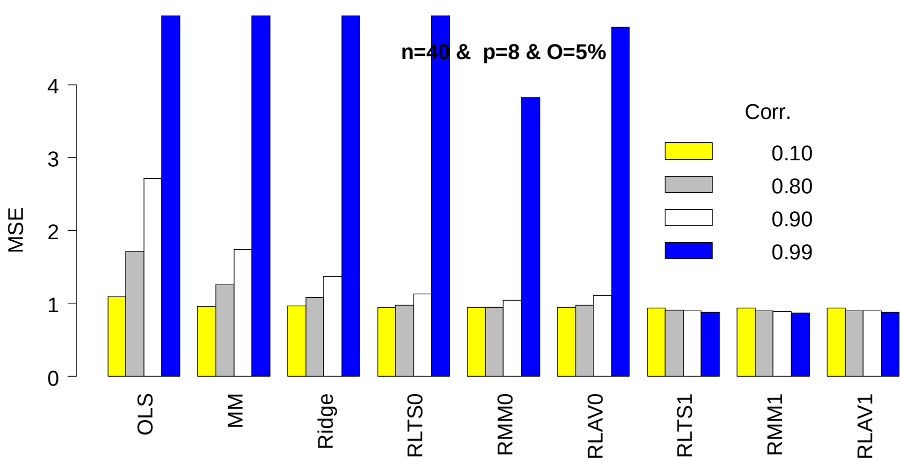
تابع جدول (١): قيم MSE للمقدرات عن جميع توليفات المحاكاة

ρ	O	n	p	OLS	MM	Ridge	RLTS0	RMM0	RLAV0	RLTS1	RMM1	RLAV1
0.90	0	20	4	2.3173	2.5059	1.1781	1.2231	1.1836	1.2059	0.7367	0.7363	0.7363
0.90	0	20	8	3.5649	5.2811	1.4676	1.4603	1.4647	1.4459	0.8170	0.8176	0.8163
0.90	0	40	4	1.5806	1.6597	0.9971	1.0448	1.0029	1.0288	0.7869	0.7856	0.7860
0.90	0	40	8	2.5868	2.8703	1.2361	1.3790	1.2589	1.3209	0.8390	0.8218	0.8299
0.90	0.05	20	4	2.9438	2.1303	1.4212	1.1846	1.1471	1.1979	0.8173	0.8162	0.8176
0.90	0.05	20	8	7.6452	6.7724	2.6388	1.9234	1.8752	2.0233	0.8911	0.8859	0.8993
0.90	0.05	40	4	1.9043	1.3318	1.1654	1.0041	0.9723	1.0053	0.8741	0.8736	0.8742
0.90	0.05	40	8	2.7137	1.7354	1.3698	1.1288	1.0440	1.1072	0.9014	0.8898	0.8989
0.90	0.10	20	4	2.2739	1.4348	1.2382	1.0125	0.9999	1.0320	0.8753	0.8750	0.8768
0.90	0.10	20	8	6.8453	4.7746	2.5264	1.6339	1.5858	1.7448	0.9350	0.9287	0.9466
0.90	0.10	40	4	1.8046	1.1174	1.1591	0.9665	0.9563	0.9763	0.9195	0.9194	0.9198
0.90	0.10	40	8	4.1364	1.7489	1.7910	1.1265	1.0513	1.1483	0.9265	0.9196	0.9292
0.90	0.15	20	4	2.4647	1.4503	1.3422	1.0421	1.0265	1.0628	0.9134	0.9127	0.9150
0.90	0.15	20	8	6.2829	5.2804	2.4266	1.6390	1.5885	1.7486	0.9813	0.9752	0.9960
0.90	0.15	40	4	1.8776	1.0835	1.2002	0.9686	0.9629	0.9815	0.9396	0.9395	0.9401
0.90	0.15	40	8	2.1617	1.1658	1.2567	0.9848	0.9689	0.9951	0.9519	0.9478	0.9564
0.99	0	20	4	31.9975	36.4880	10.3844	11.7096	10.5583	11.0726	0.6895	0.6880	0.6882
0.99	0	20	8	76.1518	100.4933	22.7745	22.3102	22.6176	22.5651	0.7740	0.7742	0.7740
0.99	0	40	4	10.8216	11.7394	3.7121	4.2601	3.7845	4.1103	0.7617	0.7600	0.7611
0.99	0	40	8	18.1969	20.7242	5.5040	6.8381	5.7033	6.2788	0.8223	0.8033	0.8125
0.99	0.05	20	4	27.7675	18.3207	8.9515	6.0748	5.4423	6.1074	0.7879	0.7871	0.7877
0.99	0.05	20	8	99.1518	81.5700	27.0804	16.1172	15.7365	17.5419	0.8598	0.8557	0.8658
0.99	0.05	40	4	6.4479	3.3951	2.4899	1.6007	1.4443	1.6304	0.8635	0.8633	0.8637
0.99	0.05	40	8	29.5995	14.9201	8.5971	5.0335	3.8246	4.7861	0.8821	0.8701	0.8783
0.99	0.10	20	4	46.5820	22.4820	15.3776	7.4852	6.4750	8.1684	0.8451	0.8447	0.8463
0.99	0.10	20	8	43.1532	27.8946	12.5218	6.3934	6.0303	6.9070	0.9208	0.9152	0.9297
0.99	0.10	40	4	9.6421	3.0543	3.4026	1.4433	1.3083	1.5487	0.9069	0.9070	0.9065
0.99	0.10	40	8	19.4912	5.9537	5.9555	2.2297	1.7881	2.3087	0.9162	0.9095	0.9168
0.99	0.15	20	4	22.0120	8.7210	7.1229	2.9219	2.7006	3.2994	0.8910	0.8907	0.8923
0.99	0.15	20	8	100.4827	64.7210	29.8878	13.7794	12.9802	15.7823	0.9427	0.9382	0.9566
0.99	0.15	40	4	13.9585	3.2976	4.7770	1.4681	1.3235	1.6832	0.9324	0.9327	0.9326
0.99	0.15	40	8	35.5075	7.8809	10.3974	2.3971	1.9656	2.7794	0.9304	0.9277	0.9331



شكل (١): أداء المقدرات وفقا لدرجة الارتباط بين المتغيرات المفسرة عند

$$(n = 20, p = 8, \ O = 0.05)$$

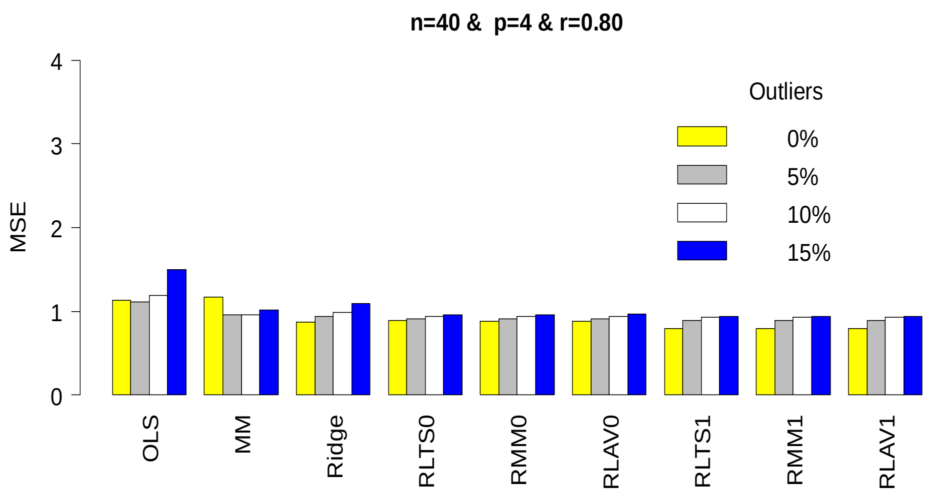


شكل (٢): أداء المقدرات وفقا لدرجة الارتباط بين المتغيرات المفسرة عند

$$(n = 40, p = 8, \ O = 0.05)$$

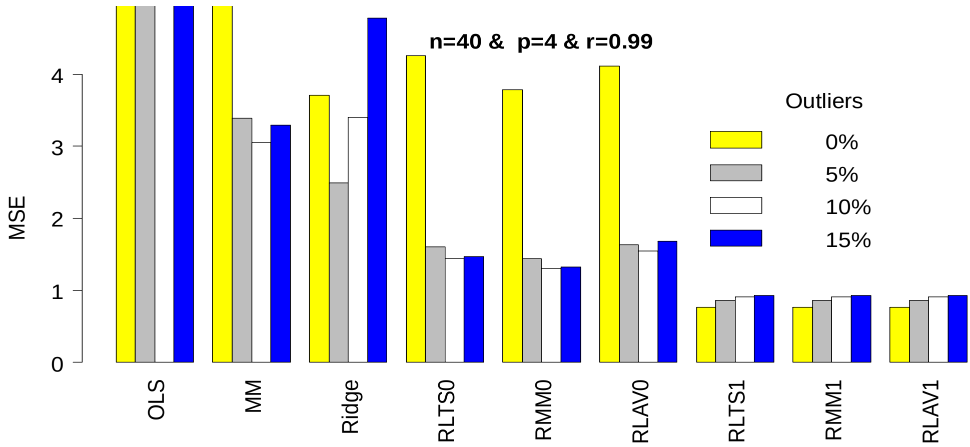
٥-٢ أداء المقدرات وفقا لنسبة المشاهدات الشاذة O

يتضح من الجدول (١) ومن الشكلين (٣) و (٤) : في حالة عدم وجود ازدواج خطي، أو وجود ازدواج خطي ضعيف نجد أن زيادة نسبة المشاهدات الشاذة تؤدي إلى ارتفاع في قيم MSE لجميع المقدرات وتكون المقدرات المقترحة ليست أفضل المقدرات المقارنة، حيث يكون مقدر MM هو الأفضل. وفي حالة عدم وجود كل من الازدواج الخطي والمشاهدات الشاذة يكون مقدر OLS هو الأفضل. أما عند درجة ارتباط 0.90 و 0.99 نجد أن زيادة نسبة المشاهدات الشاذة لا تؤثر بشكل سلبي على مقدرات ريدج المتينة السابقة، بينما تتأثر المقدرات المقترحة قليلا بشكل سلبي نتيجة زيادة نسبة المشاهدات الشاذة. ورغم ذلك فإن المقدرات المقترحة تظل هى الأفضل عند النسب المختلفة من المشاهدات الشاذة لأن مستوى قيم MSE للمقدرات المقترحة يكون أدنى من نظيره للمقدرات السابقة.



شكل (٣): أداء المقدرات وفقا لنسبة المشاهدات الشاذة O

$$(n = 40, p = 4\ , \rho = 0.80)$$

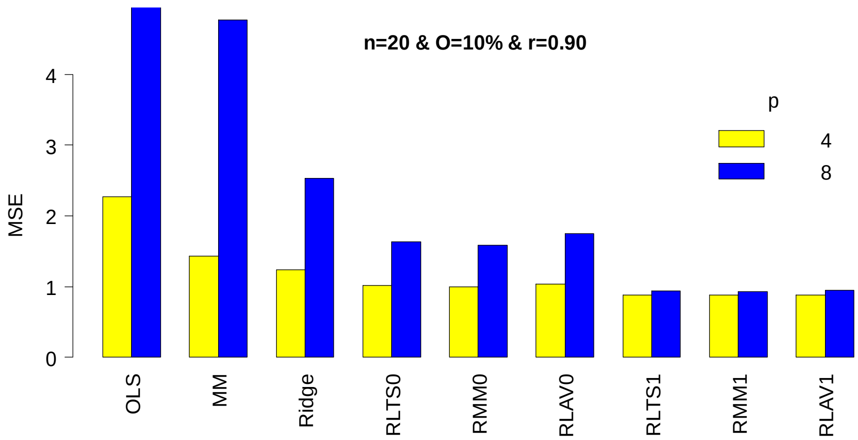


شكل (٤): أداء المقدرات وفقا لنسبة المشاهدات الشاذة O

$$(n = 40, p = 4\ , \rho = 0.99)$$

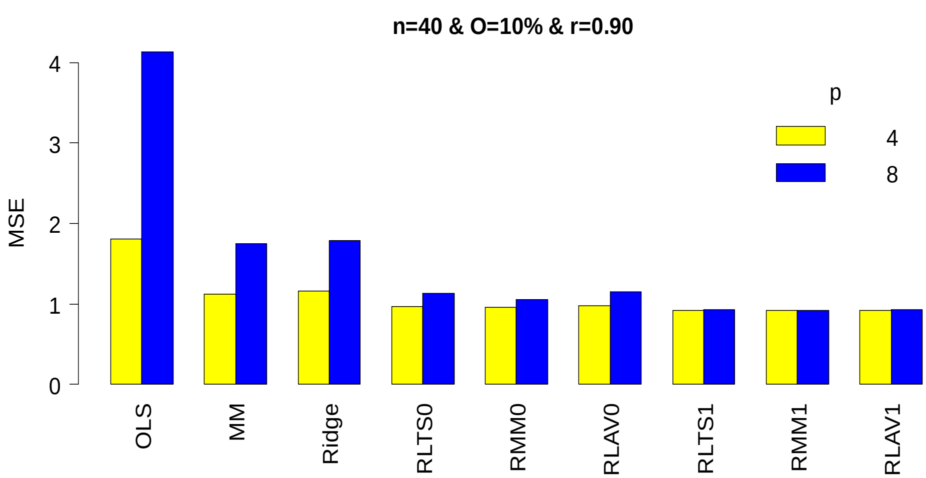
٥-٣ أداء المقدرات وفقا لعدد المتغيرات المفسرة p

من جدول (١) ومن الشكلين (٩) و(١٠) نجد أنه: زيادة عدد المتغيرات المفسرة يؤثر بشكل سلبي على أداء جميع المقدرات وهو ما يتمثل في ارتفاع قيم MSE لتلك المقدرات. وأن هذا الارتفاع في MSE يكون بدرجة أقل عند العينات الكبيرة عما هو في العينات الصغيرة. كما أن المقدرات المقترحة تكاد لا تتأثر سلبا بزيادة عدد المتغيرات عند حجم العينة 40. كذلك نجد أن تفوق المقدرات المقترحة على باقي المقدرات يكون أكثر وضوحا عند عدد متغيرات 8. بشكل عام المقدرات المقترحة أفضل من باقي المقدرات عند معظم التوليفات من عدد متغيرات 4 أو 8، ويكون الاستثناء عند بعض التوليفات من درجة ارتباط 0.10 (عدم وجود ازدواج خطي) مع نسب مشاهدات شاذة 10% و 15%، حيث يكون مقدر MM هو الأفضل. ويكون أسوأ أداء للمقدرات المقترحة عند التوليفة من 4 متغيرات وعدم وجود كل من الازدواج والمشاهدات الشاذة مع حجم عينة 40.



شكل (٥): أداء المقدرات وفقا لعدد المتغيرات p

$$(n = 20, O = 10\%, \rho = 0.90)$$



شكل (٦): أداء المقدرات وفقا لعدد المتغيرات p

$$(n = 40, O = 10\%, \rho = 0.90)$$

٥-٤ أداء المقدرات وفقا لحجم العينة n

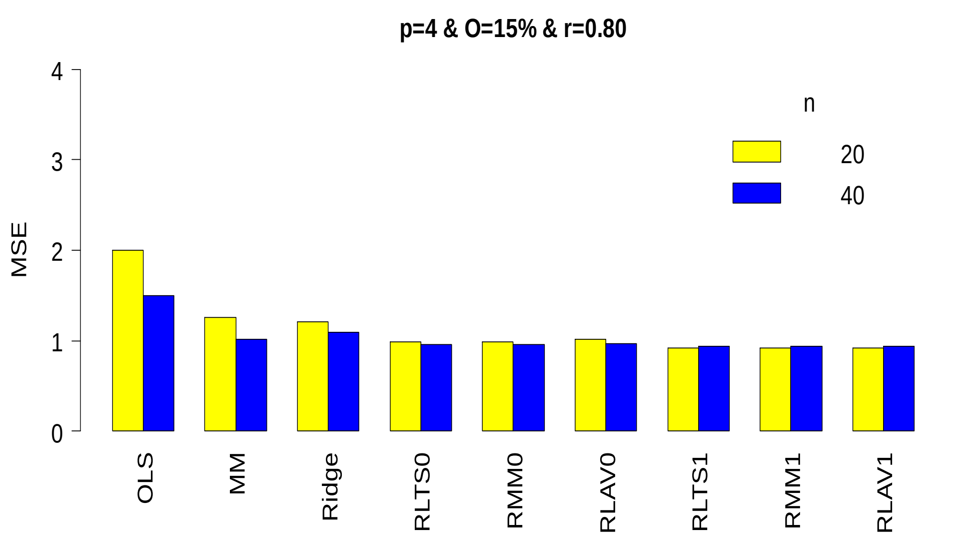
من جدول (١) ومن الشكلين (٨) و(٩) (على سبيل المثال) اللذين يعرضان أداء المقدرات وفقا لحجم العينة عند التوليفتين

$(n = 20, O = 10\%, \rho = 0.90)$ و

$(n = 40, O = 10\%, \rho = 0.90)$

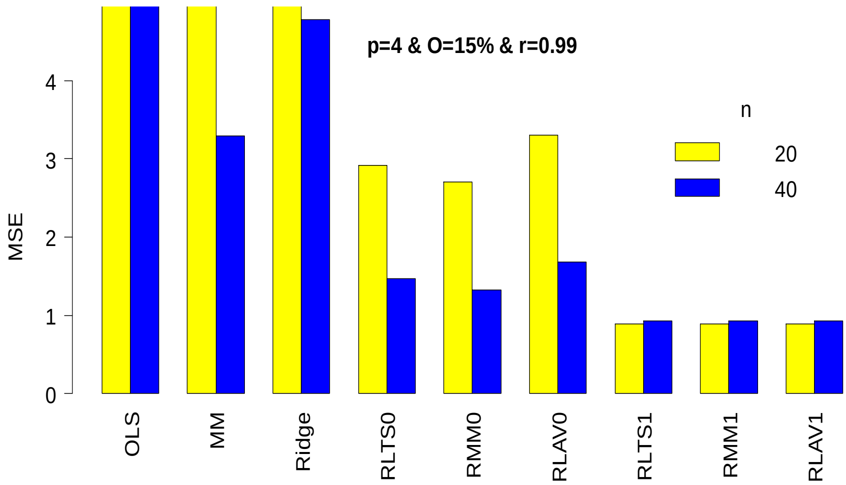
حيث تتغير درجة الارتباط، نجد أنه في حالة عدم وجود ازدواج خطي أو وجود ازدواج خطي ضعيف لا يكون لزيادة حجم العينة تأثير ملحوظ على المقدرات. وعند درجات الارتباط الأعلى نجد أن زيادة حجم العينة تؤثر بشكل إيجابي على المقدرات السابقة، بينما نجد أن المقدرات المقترحة لا تتأثر تقريبا بزيادة حجم العينة. وفي نفس الوقت تكون المقدرات المقترحة هى الأفضل عند جميع أحجام العينات، وتكون أكثر تفوقا عن حجم العينة الصغير.

وبالتالي نجد أن المقدرات المقترحة أفضل من باقي المقدرات عند جميع توليفات حجم العينة فيما عدا بعض التوليفات التي يكون فيها مقدر MM المتين هو الأفضل، وذلك عند حجم عينة 20 وعدم وجود ازدواج خطي مع نسب مشاهدات شاذة **10%** و**15%** مع عدد متغيرات 4. وهناك استثناء ثان عند حجم عينة 40 وعدم وجود ازدواج خطي، وأى نسبة للمشاهدات الشاذة. واستثناء أخير أن يكون مقدر OLS أفضل المقدرات عند عدم وجود كل من الازدواج الخطي والمشاهدات الشاذة مع عينة كبيرة.



شكل (٧): أداء المقدرات وفقا لحجم العينة n

$(p = 4, O = 15\%, \rho = 0.80)$

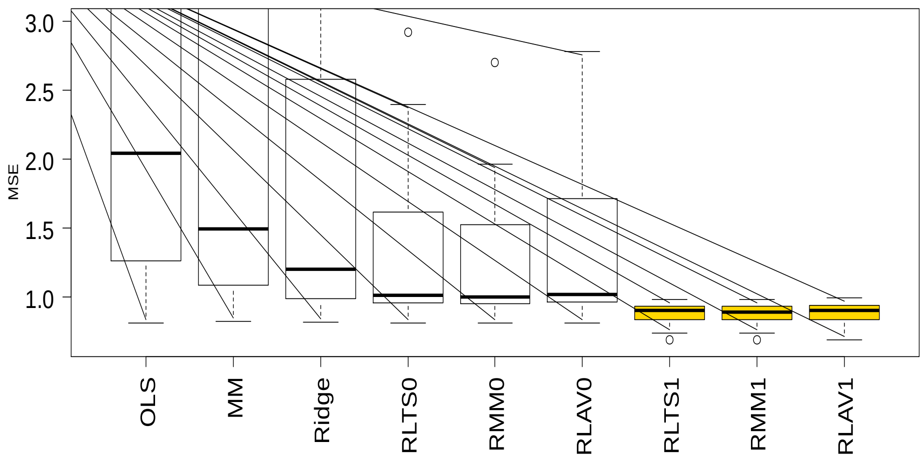


شكل (٨): أداء المقدرات وفقا لحجم العينة n

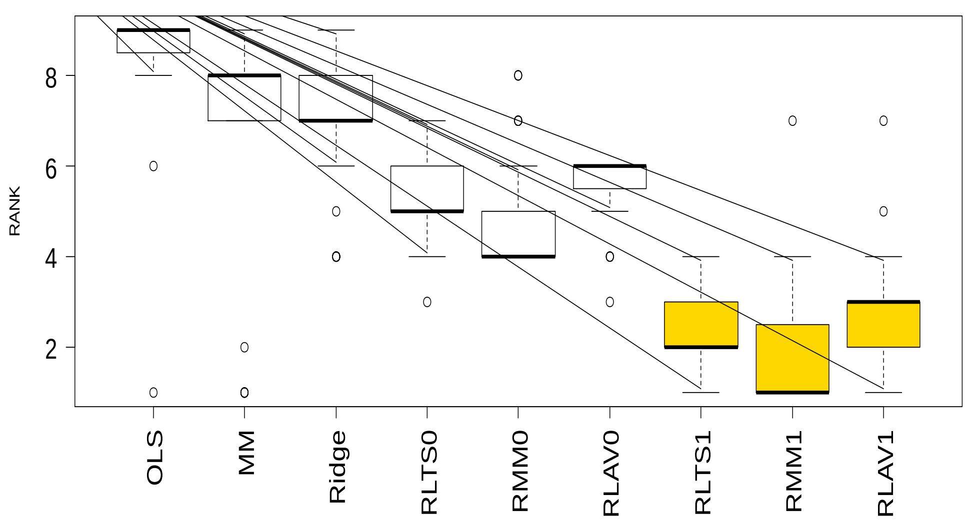
$$(p = 4, O = 15\%, \rho = 0.99)$$

٥-٥ مقارنة بين المقدرات

من الواضح من نتائج المحاكاة وفقا لكل عامل من عوامل المحاكاة تفوق المقدرات المقترحة عند معظم توليفات المحاكاة. ويمكن أن نصل إلى نفس النتيجة عند مقارنة قيم MSE لجميع المقدرات عند جميع توليفات المحاكاة، وذلك من خلال الشكل (١٠). ويعرض هذا الشكل قيم MSE من خلال الصناديق البيانية، وقد تم قطع المحور الرأسي عند القيمة 3 (بسبب وجود قيم شاذة كبيرة في قيم MSE تصل إلى 100.49) حتي يمكن المقارنة بين المقدرات بشكل عام. ويتضح من الشكل تفوق المقدرات المقترحة الثلاثة ولكن قد يكون من الصعب المقارنة بينهم بسبب تقارب قيم MSE. لذلك يمكن أيضا ترتيب المقدرات وفقا لقيم MSE بشكل تصاعدي عند كل توليفة من توليفات المحاكاة واعطاء "**رتبة**" **لكل مقدر**، بحيث المقدر الذي له أقل MSE داخل التوليفة يأخذ المرتبة "1"، وهكذا حتى يأخذ المقدر صاحب أكبر MSE المرتبة "9". ويعطي جدول (٢) تلك الرتب عند جميع التوليفات، وقد تم تظليل خلايا أفضل ثلاث مقدرات عند كل توليفة. وتم تمثيل تلك الرتب باستخدام الصناديق البيانية في الشكل (١٠). وبناءً على ماسبق يمكن ترتيب المقدرات بحيث يكون أفضل المقدرات هو هو مقدر RMM1 يليه مقدر RLTS1 ثم مقدر RLAV1.



شكل (٩): قيم MSE للمقدرات عند جميع التوليفات



شكل (١٠): رتب المقدرات عند جميع التوليفات

جدول (٢): رتب المقدرات عند كل توليفة من توليفات المحاكاة

r	prop	n	p	OLS	MM	Ridge	RLTS0	RMM0	RLAV0	RLTS1	RMM1	RLAV1
0.1	0	20	4	6	9	8	4	7	5	1	3	2
0.1	0	40	4	1	2	9	3	8	6	4	7	5
0.1	0	20	8	8	9	7	4	6	5	1	3	2
0.1	0	40	8	8	9	6	7	5	4	2	3	1
0.1	0.05	20	4	9	7	8	5	6	4	2.5	2.5	1
0.1	0.05	40	4	9	1	5	7	8	6	3	4	2
0.1	0.05	20	8	9	8	7	4	5	6	2	1	3
0.1	0.05	40	8	9	7	8	5	6	4	2	1	3
0.1	0.1	20	4	9	1	8	6	7	5	2.5	2.5	4
0.1	0.1	40	4	9	1	7	6	8	5	4	3	2
0.1	0.1	20	8	9	8	7	5	4	6	2	1	3
0.1	0.1	40	8	9	1	8	6	7	5	3	2	4
0.1	0.15	20	4	9	1	8	5	7	6	3	2	4
0.1	0.15	40	4	9	1	8	6	7	5	3.5	3.5	2
0.1	0.15	20	8	9	8	7	5	4	6	2	1	3
0.1	0.15	40	8	9	1	8	5.5	5.5	3	4	2	7
0.8	0	20	4	8	9	4	7	5	6	1	2.5	2.5
0.8	0	40	4	8	9	4	7	5	6	2	3	1
0.8	0	20	8	8	9	7	4	6	5	1	3	2
0.8	0	40	8	8	9	4	7	5	6	3	1	2
0.8	0.05	20	4	9	8	7	5	4	6	2	1	3
0.8	0.05	40	4	9	8	7	5	4	6	1.5	3	1.5
0.8	0.05	20	8	9	8	7	5	4	6	2	1	3
0.8	0.05	40	8	9	8	7	6	4	5	3	1	2
0.8	0.1	20	4	9	8	7	5	4	6	2	1	3
0.8	0.1	40	4	9	7	8	4.5	4.5	6	2	3	1
0.8	0.1	20	8	9	8	7	5	4	6	2	1	3
0.8	0.1	40	8	9	7	8	5	4	6	2	1	3
0.8	0.15	20	4	9	8	7	5	4	6	1	2	3
0.8	0.15	40	4	9	7	8	5	4	6	2	1	3
0.8	0.15	20	8	9	8	7	5	4	6	2	1	3
0.8	0.15	40	8	9	7	8	5	4	6	2	1	3
0.9	0	20	4	8	9	4	7	5	6	3	1.5	1.5
0.9	0	40	4	8	9	4	7	5	6	3	1	2
0.9	0	20	8	8	9	7	5	6	4	2	3	1
0.9	0	40	8	8	9	4	7	5	6	3	1	2
0.9	0.05	20	4	9	8	7	5	4	6	2	1	3
0.9	0.05	40	4	9	8	7	5	4	6	2	1	3
0.9	0.05	20	8	9	8	7	5	4	6	2	1	3
0.9	0.05	40	8	9	8	7	6	4	5	3	1	2
0.9	0.1	20	4	9	8	7	5	4	6	2	1	3
0.9	0.1	40	4	9	7	8	5	4	6	2	1	3
0.9	0.1	20	8	9	8	7	5	4	6	2	1	3
0.9	0.1	40	8	9	7	8	5	4	6	2	1	3
0.9	0.15	20	4	9	8	7	5	4	6	2	1	3
0.9	0.15	40	4	9	7	8	5	4	6	2	1	3
0.9	0.15	20	8	9	8	7	5	4	6	2	1	3
0.9	0.15	40	8	9	7	8	5	4	6	2	1	3
0.99	0	20	4	8	9	4	7	5	6	3	1	2
0.99	0	40	4	8	9	4	7	5	6	3	1	2
0.99	0	20	8	8	9	7	4	6	5	1.5	3	1.5
0.99	0	40	8	8	9	4	7	5	6	3	1	2

0.99	0.05	20	4	9	8	7	5	4	6	3	1	2
0.99	0.05	40	4	9	8	7	5	4	6	2	1	3
0.99	0.05	20	8	9	8	7	5	4	6	2	1	3
0.99	0.05	40	8	9	8	7	6	4	5	3	1	2
0.99	0.1	20	4	9	8	7	5	4	6	2	1	3
0.99	0.1	40	4	9	7	8	5	4	6	2	3	1
0.99	0.1	20	8	9	8	7	5	4	6	2	1	3
0.99	0.1	40	8	9	7	8	5	4	6	2	1	3
0.99	0.15	20	4	9	8	7	5	4	6	2	1	3
0.99	0.15	40	4	9	7	8	5	4	6	1	3	2
0.99	0.15	20	8	9	8	7	5	4	6	2	1	3
0.99	0.15	40	8	9	7	8	5	4	6	2	1	3

٦- الخلاصة

يعتبر وجود الازدواج الخطي والمشاهدات الشاذة من أكثر المشاكل تأثيرا في تحليل الانحدار الخطي. وعلى الرغم من تعدد الطرق المتاحة لعلاج كل مشكلة من المشكلتين بشكل منفصل، إلا أنه في الواقع عادة ما تحدث المشكلتان آنيا. وكطريقة لعلاج المشكلتين يتم دمج أحد المقدرات المتينة مع أحد مقدرات انحدار ريدج لتشكيل مقدر انحدار ريدج المتين. وقد تم تصميم دراسة محاكاة للمقارنة بين أداء ثلاثة من مقدرات انحدار ريدج المتينة المقترحة وستة مقدرات أخرى، باستخدام معيار متوسط مربعات الخطأ للمقدرات MSE بالاضافة إلى رتب المقدرات عند كل توليفة من توليفات المحاكاة. وتشير نتائج المحاكاة إلى تفوق المقدرات الثلاثة المقترحة في معظم توليفات الازدواج الخطي والمشاهدات الشاذة. وقد تفوق مقدر MM المتين على بعض المقدرات المقترحة في بعض حالات عدم وجود ازدواج خطي مع نسب مختلفة من المشاهدات الشاذة. وأكدت دراسة المحاكاة على أنه عند وجود الازدواج الخطي مع أى نسبة من المشاهدات الشاذة، تحتل الطرق المقترحة المراتب الثلاث الأولى كأفضل المقدرات، ويمكن القول أن أفضل المقدرات المقترحة هو مقدر RMM1 يليه مقدر RLTS1 ثم مقدر RLAV1.

المراجــــع

[1], Adegoke, A.S., Adewuyi, E., Ayinde, K., and Lukman, A.F. (2016), "A comparative study of some robust Ridge and Liu estimators," Science World Journal, 11, 4, 16- 20.

[2], Al-hassan, Y. (2010), "Performance of new ridge regression estimators," Journal of the Association of Arab Universities for Basic and Applied Science, 9, 23-26.

[3], Alkhamisi, M. A., and Shukur, G. (2007) "A Monte Carlo study of recent ridge parameters," Communications in Statistics—Simulation and Computation, 36, 535–547.

[4], Askin, R. G., and Montgomery, D. C. (1980), "Augmented robust estimators," Technometrics, 22, 333-341.

[5], Barnett, V., and Lewis, T. (1994), Outliers in statistical data, 3rd ed., Wiley: Chicester, UK.

[6], Belsley, D.A. (1991), Conditioning diagnostics, collinearity and weak data, Wiley: New York.

[7], Belsley, D.A., Kuh, E., and Welsch, R.E. (1980), Regression diagnostics: identifying influential data and sources of collinearity, Wiley: New York.

[8], Birkes, D., and Dodge, Y. D. (1993), Alternative methods of regression, Wiley: New York.

[9], Chatterjee, S., and Hadi, A. S. (2006), Regression Analysis by Example, Wiley: New Jersey.

[10], Groß, J. (2003), Linear regression, Springer -Verlag Berlin Heidelberg.

[11], Habshah, M., and Marina, Z. (2007), "A simulation study on ridge regression estimators in the presence of outliers and multicollinearity," Journal Teknologi, 47, 59-74.

[12], Hampel, F. R. (1974), "The Influence Curve and Its Role in Robust Estimation," Journal of the American Statistical Association, 69, 346 , 383-393

[13], Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), Robust statistics: The approach based on influence functions, Wiley: New York.

[14], Hocking, R. R., Speed, F. M., and Lynn, M. J. (1976), "A class of biased estimators in linear regression," Technometrics, 18, 4, 425-437.

[15], **Hoerl, A. E., and Kennard, R. W. (1970 a),** "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, 12, 55-67.

[16], **Hoerl, A. E., and Kennard, R. W. (1970 b),** "Ridge regression: Applications to non-orthogonal problems," Technometrics, 12, 69-82.

[17], **Hoerl, A. E., Kennard, R.W. and Baldwin, K.F., (1975),** "Ridge regression: some simulations," Communications in Statistics, 4, 105–123.

[18], **Holland, P. W. (1973),** "Weighted ridge regression: Combining ridge and robust regression methods," NBER Working Paper Series, Working Paper No. 11, 1-19.

[19], **Khalaf, G., and Iguernane, M. (2016)** "Multicollinearity and a ridge parameter estimation approach," Journal of Modern Applied Statistical Methods, 15, 2, 400-410.

[20], **Khalaf, G., and Shukur, G. (2005),** "Choosing ridge parameter for regression problem," Communications in Statistics – Theory and Methods, 34, 1177–1182.

[21], **Kibria, B. M. G. (2003),** "Performance of some new ridge regression estimators" Communications in Statistics-Simulation and Computation, 32, 419-435.

[22], **Maronna, R.A., Martin, R.D., and Yohai, V.J. (2006),** Robust statistics theory and methods, Wiley: New York.

[23], **Muniz, G., and Kibria, B.G. (2009),** "On some ridge regression estimators: An empirical comparisons,"Communications in Statistics - Simulation and computation, 38, 621-630.

[24], **Newhouse, J.P., and Oman, S.D. (1971),** "An evaluation of ridge estimators," Rand Corporation, P-716-PR, 1-28.

[25], **Pati, K. D., Adnan, R., Rasheed, B. A., and Alias, M.J. (2016),** "Estimation parameters using bisquare weighted robust ridge regression brlts estimator in the presence of multicollinearity and outliers," AIP Conference Proceedings 1750, 060-028; doi: 10.1063/1.4954633

[26], **Pati, K. D., Adnan, R., and Rasheed, B. A. (2014),** "Ridge least trimmed squares estimators in presence of multicollinearity and outliers," Nature and Science, 12,12, 1-8.

[27], Pfaffenberger, R. C., and Dielman, T. E. (1984), "A modified ridge regression estimator using the least absolute value criterion in the multiple linear regression model," Proceedings of the American Institute for Decision Sciences, Toronto, 791-793.

[28], Pfaffenberger, R. C., and Dielman, T. E. (1985), "A comparison of robust ridge estimators," Business Economics Section Proceedings of the American Statistical Association, 631-635.

[29], Ronchetti, E. M. (1987), "Robust C(α)-type tests for linear models," Sankhya. A, 49, 1-16.

[30], Rousseeuw, P. J. (1984), "Least median of squares regression," Journal of the American Statistical Association, 79, 871-880.

[31], Rousseeuw, P. J. (1985), "Multivariate estimation with high breakdown point," in Mathematical Statistics and Applications, B (ed. by W. Grossmann, G. Pflug, I. Vincze, and W. Wertz), Dordrecht, Netherlands: Reidel Publishing Company, 283 - 297.

[32], Rousseeuw, P. J., and Leroy, A. M. (1987), "Robust regression and outlier detection," Wiley: New York.

[33], Rousseeuw, P.J., and Van Driessen, K. (2006), "Computing LTS regression for large data sets," Data Mining and Knowledge Discovery, 12, 29–45.

[34], Samkar, H., and Alpu, O. (2010), "Ridge regression based on some robust estimators," Journal of Modern Applied Statistical Methods, 9, 2, Article 17.

[35], Wilcox, R. R. (2010), "Fundamentals of modern statistical methods," 2nd. Edition, Springer, London.

[36], Yohai, V. J. (1987), "High breakdown point and high efficiency robust estimates for regression," Ann Statist. , 15, 642–656.

[37], Zahari S. M., Zainol M. S., and AL-banna M. I. (2012), "Weighted ridge MM-estimator in robust ridge regression with multicollinearity," Mathematical Models and Methods in Modern Science, 124-129.