المجلة العراقية للعلوم الإحصائية (28) 2019
ص ص [77-110]

اثر المشاهدات الشاردة وذات القوة الرافعة في بناء فترات الثقة البيزية و بوتستراب

د. مزاحم محمد يحيى الهاشمي
muzahim63@gmail.com

المستخلص:

يهدف البحث إلى مقارنة كل من حدود الثقة لبوتستراب (bootstrap confidence intervals) مع حدود الثقة البيزية (Bayesian confidence intervals) لشرائح ممهدة (smoothing splines)، فضلاً عن حدود الثقة التقليدية، وذلك لتحديد أَيٍّ من هذه الحدود هي الأَفضل في ظل وجود نقاط شاردة وذات القوة الرافعة في البيانات. تم اجراء تجارب المحاكاة على أُنموذجين. الأَول: خطي في ظل وجود بيانات ملوثة بمشاهدات شاردة، وأُخرى بمشاهدات ذات القوة الرافعة، و الانموذج الثاني اللاخطي في ظل وجود بيانات ملوثة بمشاهدات شاردة. وتم تنفيذ تجارب المحاكاة على حجوم لعينات مختلفة. واستخدام طريقة المربعات الصغرى الجزائية (Penalized Least Squares method) لتوفيق الانحدار اللامعلمي. وتم استخدام دالة الفاعلية المتقاطعة المعممة (Generalized Cross Validation function) (GCV) لاختيار قيمة التمهيد (The amount of smoothing).

الكلمات المفتاحية: المشاهدات الشاردة، المشاهدات ذات القوة الرافعة، طريقة المربعات الصغرى الجزائية، فترات الثقة البيزية، فترات الثقة للبوتستراب، فترات الثقة التقليدية.

The Effect of the Outliers and Leverage Points in the Construction of the Bayesian and Bootstrap Confidence Intervals

Abstract

The aim of this research is to compare the bootstrap confidence intervals with the Bayesian confidence intervals for smoothing splines as well as the traditional confidence intervals to determine which of these limits are best in the presence of Outliers and Leverage points in data. The simulation experiments were conducted on two models: the first was linear in the presence of data that was contaminated with outliers and the other with the Leverage points: The second model was nonlinear in the presence of data contaminated with outlying observations.

* مدرس / قسم الاحصاء والمعلوماتية / كلية علوم الحاسوب والرياضيات/ جامعة الموصل .

تاريخ استلام البحث 2018/9/26 تأريخ القبول 2018 /11/14

Simulation experiments were carried out on different samples. The Penalized Least Squares method was used to fit the Nonparametric regression. The Generalized Cross Validation function (GCV) was used to select the amount of smoothing.

Key words: Outliers, Leverage points, Penalized Least Squares, Bayesian confidence intervals, Bootstrap confidence intervals, traditional confidence intervals.

المقدمة:

لغرض تقدير معلمات المجتمع الإحصائي التي غالباً ما تكون مجهولة، يُلْجأ إلى العينات للحصول على تقديرات لمعلمات المجتمع. وللتقدير أُسلوبان، الأول: يسمى تقدير المعلمة بنقطة، والثاني يسمى تقدير المعلمة بفترة. ففي حالة تقدير المعلمة بنقطة نحصل على قيمة واحدة من العينة، وتستخدم هذه القيمة الواحدة كتقدير لمعلمة المجتمع المجهولة. أما في حالة تقدير المعلمة بفترة، فنحصل على فترة من العينة تتحدد بحدين هما: الحد الأدنى، والحد الأعلى، وإنّ هذه الفترة تحتوي على عدد غير محدد من القيم.

يتميز تقدير المعلمة بفترة بأَنه يمكن معرفة مدى دقة التقدير؛ وذلك من خلال حساب احتمال أن يكون التقدير صحيحاً، لذا فإن فترات التقدير تسمى أيضاً " فترات الثقة " Confidence intervals؛ لأنَّ هذه الفترات تعتمد إحصائياً على درجات أو مستويات ثقة معينة Confidence Levels. ويتم تحديد المستوى المطلوب من الثقة من الباحث مثل 95% أو 99 % وغيرها، بمعنى أنَّ احتمال أن تكون فترة التقدير صحيحة هو 0.95 أو 0.99.

ومن العوامل التي تؤثر في تباعد حدود الثقة هو حجم العينة، ومستوى الثقة. وعادة ما يكون لحجم العينة الأَثر الكبير في الوصول الى تقدير أفضل لمعلمة المجتمع، اذ كلما كبر حجم العينة فإنَّ ذلك سيقارب من حدود الثقة؛ لأنه يقلل من الانحراف المعياري، وهو بذلك دالة على كفاءة المقدر.

تعد دراسة القيم الشاردة (Outlying values)، فضلاً عن القيم ذات القوة الرافعة من المسائل المهمة في الإحصاء النظري والتطبيقي، لما لها من أَثر في الوصول الى دقة في النتائج. لقد أُثير الكثير من النقاش حول قبول او استبعاد هذه المشاهدات، حتى استقر الحال بقبولها في التحليل، خاصة اذا كانت هذه المشاهدات هي من صميم البيانات التي تحت الدراسة. وقد ظهرت الكثير من طرائق التحليل الحصينة، التي تختلف بشكل كبير عن الطرائق التقليدية، اذ أَخذت بنظر الاعتبار وجود هذه المشاهدات في التحليل.

إنَّ البحث عن معلمة المجتمع غير المعلومة في ظل وجود المشاهدات الشاردة وذات القوة الرافعة تعد من المسائل المهمة، لما لهذه المشاهدات من أَثر في تضليل النتائج، وخصوصاً حينما يتم التعامل مع حدود الثقة (الذي يفترض أن تشمل القيمة الحقيقية)، اذ تؤثر هذه المشاهدات في سعة فترات الثقة (كلما تقلصت فترة الثقة زادت دقة التقديرات)؛ لتغطي مشاهدات أَكبر نتيجة لتطرف هذه المشاهدات عن تجمع بقية المشاهدات (Bulk of data)، مما يؤدي الى تباعد حدود الثقة الأَمر الذي يؤدي إلى تشتت أَكبر في النتائج.

دراسات سابقة:

استنتج كل من (Wang & Wahba, 1995) (في حالة عدم وجود المشاهدات الشاردة وذات القوة الرافعة). أَنَّ فترات الثقة البيزية مشابهة لفترات الثقة لبوتستراب فيما يتعلق بمتوسط احتمال نطاق التغطية. كما توصل الباحثان إلى أَنَّ فترات الثقة للبوتستراب الأَفضل عندما تكون حجوم البيانات الصغيرة. فضلاً عن ذلك فإنَّ فترات الثقة لبوتستراب المئيني (Bootstrap percentile-interval) أَفضل من الأَنواع الأُخرى لحدود الثقة لبوتستراب.

هدف البحث:

يهدف البحث إلى مقارنة كل من حدود الثقة للبوتستراب (bootstrap confidence intervals) مع حدود الثقة البيزية (Bayesian confidence intervals) للشرائح الممهدة (smoothing splines)' فضلاً عن حدود الثقة التقليدية، لتحديد أَيٍّ من هذه الحدود حصينة تجاه القيم الشاردة وذات القوة الرافعة في البيانات.

أَهمية البحث:

إنَّ الوصول إلى الدقة في التقديرات، فضلاً عن البحث عن معلمة المجتمع غير المعلومة في ظل وجود المشاهدات الشاردة وذات القوة الرافعة تعد من المسائل المهمة في الإحصاء، لما لهذه المشاهدات من أَثر في تضليل النتائج، وخصوصاً حينما يتم التعامل مع حدود الثقة (الذي يفترض أن تشمل القيمة الحقيقية، فضلاً عن انها دالة على كفاءة المقدر)، اذ من المتوقع أَنْ تؤثر هذه المشاهدات في عرض فترات الثقة (كلما تقلصت فترة الثقة زادت دقة التقديرات)؛ لكونها تزيد من حالة التشتت، ثم في الوصول الى دقة في النتائج.

1. الجانب النظري

1.1: تقدير المربعات الصغرى الجزائية (Penalized Least Squares Estimation)

يعد الانحدار اللامعلمي (Nonparametric Regression) شكلاً من أَشكال تحليل الانحدار، اذ لا توجد أيّة صيغ مسبقة للمتغيرات التوضيحية مع متغير الاستجابة، كما لا يفترض الانحدار اللامعلمي شكل العلاقة بين المتغيرات، ولكنه قادر على استنتاج العلاقات المعقدة من البيانات (Simonoff, 2012). وبذلك يختلف الانحدار اللامعلمي عن الانحدار الخطي والانحدار اللاخطي، فضلاً عن النماذج الخطية المعممة (Generalized Linear Model) بأَنَّ النماذج الثلاثة الاخيرة تتطلب توصيف أُنموذج الانحدار، فضلاً عن أَنَّ أُنموذج الانحدار هو ذو عدد محدد من المعلمات، في حين يستخدم الانحدار اللامعلمي حينما لا تتوفر معلومات مسبقة عن أُنموذج الانحدار، او عندما لا يمكن تمثيل البيانات بأُنموذج ذي عدد محدد من المعلمات.

توفر طرائق التمهيد (Smoothing Methods) جسراً بين عدم وجود افتراضات لنماذج الانحدار اللامعلمية والافتراضات القوية لنماذج الانحدار المعلمية، وذلك من خلال ايجاد افتراضات ضعيفة نسبياً (Simonoff, 2012).

لقد أَصبحت إجراءات الانحدار الجزائية نهجاً شائعاً جداً لتقدير الدوال المعقدة. فعلى سبيل المثال الصفائح الممهدة هي الحل لمسالة التقليل (minimization)، أَو الوصول الى الحد الأَدنى في الفضاء الدالي. ففي أية مسالة تقليل، هناك العديد من الأسئلة التي يتم طرحها، مثل: هل الحل موجود؟ فإذا كانت الإجابة نعم، هل الحل فريد(unique)؟ وكيف يمكن الوصول إليه؟. فاذا كانت المشكلة تطرح في إطار إعادة انتاج فضاء هيلبرت (الذي سيتم تناوله لاحقاً)، فذلك يعني ضمان وجود الحل، وأَنه فريد، فضلاً عن أَنه يأخذ صيغة بسيطة للغاية (Nosedal, et al., 2012).

تستخدم طريقة المربعات الصغرى الجزائية لتوفيق أُنموذج الإنحدار اللامعلمي؛ وذلك بتقليل مجموع مربعات الباقي، فضلاً عن حد الجزاء. اذ يتم حساب الشرائح الممهدة للصفائح الرقيقة (thin-plate smoothing spline) (تشير الصفائح الرقيقة الى مقاربة للخصائص الفيزيائية للرقائق المعدنية التي تتضمن قدرة هذه الرقائق على التقوس والانحناء (Ruan, Zhixing, et al., 2013)) وذلك لتقريب الدوال المتعددة الممهدة للبيانات المشاهدة. وبهذه الطريقة يتم توفيق البيانات وفق أُنموذج مرن بحيث إِنَّ عدد المعلمات المعنوية يمكن أَن يكون أَكبر من نقاط التصميم الوحيدة (unique design points). لذا فانه كلما ازداد حجم العينة فإِن فضاء الأنموذج يزداد ايضاً مما يؤدي إلى اختيار أُنموذج من مجموعة كبيرة من النماذج المحتملة بحيث يحقق الاتزان بين تعقيد الأُنموذج وأَفضل توفيق(Wood, 2003).

يمكن بوساطة المربعات الصغرى الجزائية توفيق البيانات بعناية وبصورة متزنة بعيداً عن التوفيق غير الممهد (roughness of the fit) الناتجة عن وجود الكثير من العقد (Knots) (وهي التقاطعات (junctions) متعددة الحدود من الدرجة n، التي تؤدي إلى ظهور منحنٍ ممهدٍ خلال مجموعة من النقاط) في الأنموذج (رشيد و خمو، 2005)(Ruppert, et al., 2003). يعرف تقدير المربعات الصغرى الجزائية بأنه حل فريد في إيجاد السطح الذي يقلل (minimizes) المربعات الصغرى الجزائية لفئة كل السطوح التي تحقق الشروط المنتظمة الكافية (sufficient regularity conditions).

ليكن ( ) متجه المتغيرات التوضيحية ذا البعد (d) من المصفوفة $(\mathbf{X})$ ذات البعد $(n \times d)$، وليكن i متجه المتغيرات التوضيحية ذا البعد p، وان y_i هي القيم المشاهدة لمتغير الاستجابة المقترنة بكل من المتجهين $(\mathbf{x}_i, \mathbf{z}_i)$. وعلى افتراض أنَّ العلاقة بين $\mathbf{z}_i$ و y_i هي علاقة خطية، وأنَّ العلاقة بين $\mathbf{x}_i$ و y_i هي علاقة غير معلومة، فإنه يمكن توفيق البيانات باستخدام الأنموذج شبه المعلمي (Semiparametric Model) (هي نماذج إحصائية، إذ إنَّ بعض فضاء المعلمة (parameter space) لها مكون واحد أو أكثر من الابعاد غير النهائية (infinite-dimensional component)، وكالاتي (Nosedal, et al., 2012):

$$\varepsilon_i = y_i - f(\mathbf{x}_i) - \mathbf{z}_i\boldsymbol{\beta} \qquad i = 1, \dots, n \qquad (1)$$

إذ إن y_i متغير الاستجابة، وإنّ $\mathbf{x}_i$ هي متغيرات التمهيد، و f هي الدالة الممهدة (smooth function)، وإنّ $\boldsymbol{z_i}$ هي متغيرات الإنحدار، و $\boldsymbol{\beta}$ هو المتجه المعلمي (parametric vector) غير المعلوم ذو البعد p. وان $\varepsilon_i \sim N(0, \sigma^2)$ بشرط أنها مستقلة.

تستخدم طريقة المربعات الصغرى الاعتيادية لتقدير $f(\mathbf{x_i})$ و $\boldsymbol{\beta}$ بوساطة تقليل مجموع مربعات البواقي الجزائية الآتي:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(\mathbf{x_i}) - \mathbf{z_i}\boldsymbol{\beta})^2 \qquad \dots (2)$$

لكون أنَّ الفضاء الدالي (functional space) لـ $f(\mathbf{x_i})$ كبير جداً، فإنه يمكن استخدام طريقة المربعات الصغرى الجزائية التي يمكن من خلالها الحصول على التقدير الذي يوفق البيانات بشكل جيد، فضلاً عن أنَّ له درجة معينة من التمهيد، وتعرف بالآتي:

$$S_\lambda(f) = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(\mathbf{x_i}) - \mathbf{z_i}\boldsymbol{\beta})^2 + \lambda J_2(f) \qquad \dots (3)$$

إنَّ الحد الأَول من الصيغة في أعلاه يمثل دقة التوفيق، وإنّ الحد الثاني المقترن بالوزن λ يمثل الجزاء غير الممهد لـ f ،الذي يعرف في معظم الحالات بأنه التكامل لمربع المشتقة الثانية لـ f $\left(\int\{f^{"}(x)\}^2\,dx\right)$ (Nosedal, Storlie, & Christensen 2011).

تمثل المربعات الصغرى الجزائية حلاً وسطاً بين حسن المطابقة (goodness-of-fit) والتمهيد، ويتم التحكم في التوازن بين الاثنين بواسطة معلمة التمهيد λ ،التي تأخذ القيم من 0 الى ∞ (Wang, 2011). والتي تتحكم بالمفاضلة بين التمهيد وحسن المطابقة. عندما تكون λ كبيرة فإنه من الصعوبة بمكان إجراء الجزاءات للتوافيق غير الممهدة. وعلى العكس من ذلك حينما تكون λ صغيرة، فإنَّ التركيز سيكون بشكل أَكبر على حسن المطابقة.

لذا، فإنه يمكن تمثيل الصيغة (2) كتركيبة خطية (linear combination) من سلسلة من الدوال الأساسية. وبالتالي فإن التقديرات النهائية للصفائح الممهدة f تكون بالصيغة الاتية:

$$\hat{f}_{\boldsymbol{\lambda}}(\mathbf{x_i}) = \theta_0 + \sum_{j=1}^{d}\theta_j\,\mathbf{x_j} + \sum_{j=1}^{p}\delta_j\,\beta_j\left(\mathbf{x_j}\right) \qquad \ldots(4)$$

إذ إنَّ j هي الدالة الأَساسية، التي تعتمد على مكان وجود البيانات $\mathbf{x_i}$، وإنّ $\boldsymbol{\theta} = \{\theta_0, \ldots, \theta_d\}$ و $\boldsymbol{\delta} = \{\delta_0, \ldots, \delta_d\}$ هي المعلمات المطلوب تقديرها.

إذا كانت $\boldsymbol{\lambda}$ ثابتاً $(\boldsymbol{\lambda}$ تمثل معلمة انحدار الحرف في انحدار الحرف (Bates, et al., 1987)، فإنه يمكن تقدير المعلمات $(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\beta})$ بوساطة إيجاد حل للنظام $n \times n$. كما يمكن اختيار معلمة التمهيد بوساطة تقليل دالة الفاعلية المتقاطعة المعممة (Generalized Cross Validation) (GCV) ويرمز لها بـ $\boldsymbol{V}(\lambda)$. إذ تستخدم دالة الفاعلية المتقاطعة المعممة في البحث عن القيمة التقديرية $\hat{\lambda}$ المثالية لـ $\boldsymbol{\lambda}$ من البيانات. إذ إن $\widehat{\ }$ تقلل المقدار الآتي (Golub, Heath & Wahba, 1978).

$$\boldsymbol{V}(\lambda) = \frac{\boldsymbol{n^{-1}}\|(\mathbf{I}-\mathbf{A}(\boldsymbol{\lambda}))\boldsymbol{y}\|^2}{\left[\boldsymbol{n^{-1}}tr(\mathbf{I}-\mathbf{A}(\boldsymbol{\lambda}))\right]^2} \qquad \ldots(5)$$

إذ إنَّ $\boldsymbol{A}(\boldsymbol{\lambda})$ هي مصفوفة القبعة (hat matrix) ذات بعد $n \times n$ لمصفوفة انحدار الحرف، وتسمى مصفوفة التمهيد (smoothing matrix) كما تسمى أَيضاً (influence matrix)، وتعرف بالآتي (Bates, et al., 1987):

$$\boldsymbol{A}(\boldsymbol{\lambda}) = \boldsymbol{X}(\boldsymbol{X^T X} + \boldsymbol{n\lambda I})^{-1}\boldsymbol{X^T} \qquad \ldots(6)$$

على فرض أَنَّ $\mathcal{R}_m$ هو فضاء للدوال التي مشتقاتها الجزئية من الرتبة الكلية m، وعلى فرض أَنَّ أُنموذج البيانات هو الآتي:

$$y_i = f(\mathbf{x}_i) + \varepsilon_i \ , \ i = 1, \dots, n \qquad \dots (7)$$

وان $f \in \mathcal{R}_m$.

واذا كانت λ ثابتاً. فإنه لغرض تقدير f يتطلب تقليل دالة المربعات الصغرى الجزائية

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(\mathbf{x_i}) - \mathbf{z_i}\boldsymbol{\beta})^2 + \lambda J_m(f) \qquad \dots (8)$$

إنَّ $J_m(f)$ λ هو حد الجزاء لفرض التمهيد على f إذ إنه يتناقص كلما زاد التمهيد.

هناك عدة طرائق لتعريف $J_m(f)$. بالنسبة لشرائح التمهيد ذات الصفائح الرقيقة، عندما تكون $\mathbf{X}$ ذات بعد d، فإنه يمكن تعريف $J_m(f)$ على انها (Wahba, 1990)

$$J_m(f) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{m!}{\alpha_1! \dots \alpha_1!} \left(\frac{\partial^m f}{\partial_{x_1}\alpha_1 \dots \partial_{x_d}\alpha_d} \right)^2 d_{x_1} \dots d_{x_d} \qquad \dots (9)$$

إذ إنّ i $m = \sum_{i=1}^{n}$. وفقاً للتعريف في أعلاه، فإن بعض دوال $J_m(\ )$ سيكون لها جزاء مقداره الصفر. ويسمى الفضاء الذي يمتد بوساطة مجموعة متعددي الحدود (polynomials) ذات المساهمات الصفرية الجزاء الفضاء متعدد الحدود (polynomial space). إنّ البعد للفضاء متعدد الحدود M هو دالة من البعد d والرتبة m للجزاء التمهيدي.

عندما $d = 2$ و $m = 2$، فإن $J_m(\ )$ تعرف بالآتي:

$$J_2(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(f_{x_1x_1}^2 + 2f_{x_1x_2}^2 + f_{x_2x_2}^2\right) d_{x_1}\, d_{x_2} \qquad \dots (10)$$

إذ إنّ

$$f_{x_1x_1}^2 = \left(\frac{\partial^2 f}{\partial_{x_1^2}}\right)^2 ; f_{x_1x_2}^2 = \left(\frac{\partial^2 f}{\partial_{x_1}\partial_{x_2}}\right)^2 ; f_{x_2x_2}^2 = \left(\frac{\partial^2 f}{\partial_{x_2^2}}\right)^2$$

وإن M هي الحجم لـ $3 = (\{1, x_1, x_2\})$.

بشكل عام فإن m و d يجب أَن تستوفيا الشرط بأَنّ $2m - d > 0$. لإجل التبسيط، سيتم التعامل مع $m = 2$ للصيغ والمعادلات في أَدناه.

وقد برهن Duchon (1977) على ان f_λ لها الصيغة الاتية (لمزيد من التفاصيل، انظر Duchon (1977) [3]:

$$f_\lambda(\mathbf{x}_i) = \theta_0 + \sum_{j=1}^{d}\theta_j \mathbf{x}_{ij} + \sum_{j=1}^{d}\delta_j \boldsymbol{E}_2\left(\mathbf{x}_i - \mathbf{x}_j\right) \qquad \dots (11)$$

إذ إنَّ $_2(\mathbf{x}_i - \mathbf{x}_j) = \frac{1}{2^3\pi}\|\mathbf{x}_i - \mathbf{x}_j\|^2 \log\left(\|\mathbf{x}_i - \mathbf{x}_j\|\right)$ $\|.\|$ هي المعيار الاقليدي (Euclidean norm)) عندما $d = 2$.

إذا كانت $\mathbf{K}$ تحتوي على العناصر $_{ij} = {}_2(\mathbf{x}_i - \mathbf{x}_j)$ ، وإن $\mathbf{T}$ تحتوي على العناصر $\mathbf{T}_{ij} = (\mathbf{X}_{ij})$، فإن الهدف هو في إيجاد المعلمات β ، θ و δ التي تقلل المقدار الآتي

$$S_\lambda(\beta, \theta, \delta\ ) = \frac{1}{n}\ \|y - \mathbf{T}\theta - \mathbf{K}\delta - \mathbf{Z}\beta\|^2 + \lambda\ \delta^T\ \mathbf{K}\delta \qquad \ldots(12)$$

يمكن الحصول على حل وحيد إذا كانت المصفوفة $\mathbf{T}$ ذات رتبة كاملة (full rank)، وإن $\boldsymbol{\delta}$ يجب ان تستوفي الشرط الآتي $\mathbf{T}^T\boldsymbol{\delta} = 0$.

إذا كانت $\alpha = \begin{pmatrix} \\ \beta \end{pmatrix}$ ، وان $\mathbf{X} = (\mathbf{T}\ \mathbf{Z})$، فإن الصيغة لـ S_λ ستكون على الآتي:

$$\frac{1}{n}\ \|\boldsymbol{y} - \mathbf{X}\alpha - \mathbf{K}\delta\|^2 + \lambda\ \delta^T\ \mathbf{K}\delta \qquad \ldots(13)$$

يمكن الحصول على المعلمات α و δ وذلك بحل المعادلات الآتية:

$$(\mathbf{K} + \boldsymbol{n}\lambda\mathbf{I}_n)\boldsymbol{\delta} + \mathbf{X}\boldsymbol{\alpha} = \boldsymbol{y} \qquad \ldots(14)$$

$$\mathbf{X}^{\boldsymbol{T}}\boldsymbol{\delta} = \mathbf{0} \qquad \ldots(15)$$

يمكن حساب α و δ وذلك بتحلل QR[1] (QR decomposition) للمصفوفة $\mathbf{X}$، وعلى الآتي (Wang, Yuedong)

$$\mathbf{X} = (\mathbf{Q_1} \quad \mathbf{Q_2})\begin{pmatrix}\mathbf{R}\\ \mathbf{0}\end{pmatrix} \qquad \ldots(16)$$

إذ إن $\mathbf{Q_1}$ مصفوفة ذات بعد $n \times p$ و $\mathbf{Q_2}$ مصفوفة ذات بعد $n \times (n-p)$ و $\mathbf{R}$ هي مصفوفة مثلثية عليا ذات بعد $p \times p$، وإنّ $(\mathbf{Q_1} \quad \mathbf{Q_2})$ مصفوفة متعامدة (orthogonal matrix)، وان $\mathbf{R}$ هي مصفوفة مثلثية عليا، وإنّ $\mathbf{X}^{\boldsymbol{T}}\boldsymbol{Q_2} = \mathbf{R}^{\boldsymbol{T}}\boldsymbol{Q_1^T Q_2} = \mathbf{0}$.

بما أنَّ $\boldsymbol{Q_1^T\delta} = \mathbf{0}$ كما في المعادلة (15)، فإن δ يجب ان تكون ضمن فضاء العمود لـ $\mathbf{Q_2}$. لذا فإنَّ δ يمكن صياغتها على أنها $\delta = \mathbf{Q_2}\gamma$ للمتجه γ. وبالتعويض عن $\delta = \mathbf{Q_2}\gamma$ في المعالة السابقة وبالضرب بـ $\boldsymbol{Q_2^T}$، نحصل على

---

1 يمكن تحلل أَية مصفوفة مربعة A ، وذلك بأَنّ A = QR، اذ ان Q هي مصفوفة متعامدة (orthogonal matrix) وان R هي المصفوفة المثلثية العليا (upper triangular matrix).

$$\boldsymbol{Q_2^T(K + n\lambda I)Q_2\gamma = Q_2^T y} \quad \ldots(17)$$

او

$$\boldsymbol{Q_2\gamma = \left[Q_2^T(\mathbf{K} + n\lambda\mathbf{I})\right]^{-1}Q_2^T y} \quad \ldots(18)$$

$$\delta = \mathbf{Q_2}\gamma = \mathbf{Q_2}\left[\boldsymbol{Q_2^T}(\mathbf{K} + n\lambda\mathbf{I})\mathbf{Q_2}\right]^{-1}\boldsymbol{Q_2^T y} \quad \ldots(19)$$

يمكن الحصول على α بحل المعادلة الآتية:

$$\mathbf{R}\alpha = \boldsymbol{Q_1^T}\left[\mathbf{y} - (\mathbf{K} + n\lambda\mathbf{I})\delta\right] \quad \ldots(20)$$

وإذا تم افتراض أَن الأَثر (trace) لـ $A(\lambda)$ هي درجات الحرية للانموذج، وأنّ الأَثر لـ $(\mathbf{I} - \mathbf{A}(\lambda))$ هي درجات الحرية للخطأ، فإن σ^2 يمكن أَن تأخذ الصيغة الآتية:

$$\hat{\sigma}^2 = \frac{RSS(\lambda)}{tr\ (\mathbf{I}-\mathbf{A}(\lambda))} \quad \ldots(21)$$

إذ إن RSS(λ) هو مجموع مربعات البواقي (Residuals Sum of Squares).

2.1: حدود الثقة:

تعد حدود الثقة من المقاييس الإحصائية التي لا يقتصر الاستفادة منها على تقدير معلمة المجتمع، بل تتعدى ذلك الى إعطاء معلومات عن دقة التقدير. حينما تكون حدود الثقة متباعدة فهي دالة على الكثير من عدم اليقين بقيمة معلمة المجتمع، وعندما تكون حدود الثقة متقاربة فتدل على أَن التقدير محدد بدقة كبيرة (Reed, Pileggi & Winkelman, 1973) (Terry & Kelley, 2012).

2.1.1: بناء حدود الثقة للبوتستراب (Wang & Wahba, 1995)

لتكن $\{\mathrm{x}_i\}_{i=\mathbf{1},n}$ نقاط تصميم ثابتة (fixed design points)، ولتكن $\hat{f}_{\hat{\lambda}}$ و $\hat{\sigma}^2$ تقديرات لكل من f و σ^2 من البيانات. وعلى افتراض أَنَّ $\hat{f}_{\hat{\lambda}}$ هي دالة حقيقية لـ f، فإنه يمكن اتباع الخطوات الآتية لغرض بناء حدود الثقة للبوتستراب:

1. توليد عينة بوتستراب بحسب الصيغة الآتية:

$$y_i^* = \hat{f}_{\hat{\lambda}}(\mathrm{x}_i) + \epsilon_i^*$$

إذ إنَّ $^* = (\epsilon_1^*, \ldots, \epsilon_n^*)^T \sim N(0, \hat{\sigma}^2 I_{n\times n})$.

2. إيجاد تقدير للصفائح الممهدة $\hat{f}_{\hat{\lambda}}^{*}$ بالاستناد على عينة البوتستراب.
3. افرض أنَّ $f^{*}(\mathrm{x}_i)$ تمثل متغيراً عشوائياً لقيم بوتستراب التقديرية عند x_i.
4. إعادة الخطوات السابقة B من المرات.
5. إيجاد $D_i^{*} = \left(f_{\lambda}^{*}(\mathbf{x}_i) - \hat{f}_{\hat{\lambda}}(\mathbf{x}_i)\right)$ والقيم $\chi_{\alpha/2}$ و $\chi_{1-\alpha/2}$، إذ إنَّ i^{*}.

يلاحظ من الخطوات في أعلاه أنه في كل x_i، لدينا B من تقديرات البوتستراب لـ $\hat{f}_{\hat{\lambda}}(\mathrm{x}_i)$. وهي بذلك تحقق B من $(\mathrm{x}_i)^{*}$.

2.1.2: بناء فترات الثقة البيزية:

اقترح (Wahba, 1983) أُنموذج الشريحة (Spline model) كأُنموذج بيزي (Bayesian model)، واقترح الصيغة الآتية لفترات الثقة البيزية لتمهيد تقديرات الشرائح:

$$\hat{f}_{\lambda}(\mathbf{x}_i) = \pm z_{\alpha/2}\sqrt{\hat{\sigma}^2 a_{ii}(\lambda)}$$

إذ إنّ $\mathrm{a_{ii}}(\lambda)$ هو العنصر i للمصفوفة $\mathrm{A}(\lambda)$، وإنَّ $\mathrm{z}_{\alpha/2}$ هي الربيعي $1-\alpha/2$ للتوزيع الطبيعي القياسي. وتفسر فترات الثقة على أنها فترات "عبر الدوال" (across the function) بدلا من الفترات النقطية (pointwise intervals).

2 الجانب العملي:

ينظر الى المحاكاة على أَنها الملاذ الأَخير الذي يتم توظيفه في حالة صعوبة الحصول على حلول ممكنة باستخدام منهجيات الحل التحليلي. إنَّ الأَساليب الحديثة المتقدمة في المحاكاة، نتيجة لتوفر البرامجيات، فضلاً عن تطور التقنيات. جعل من المحاكاة واحدة من أَكثر الطرائق انتشاراً، والادوات الاكثر قبولاً في الوصول إلى معلومات قيمة عن النماذج المستخدمة وعن المتغيرات الأكثر أهمية، فضلاً عن كيفية تفاعل هذه المتغيرات مع بعضها البعض.

تم اجراء تجارب المحاكاة على أُنموذجين، الاول: خطي في ظل وجود بيانات ملوثة بمشاهدات شاردة، الآخر بمشاهدات ذات القوة الرافعة، والانموذج الثاني اللاخطي في ظل وجود بيانات ملوثة بمشاهدات شاردة. تم تنفيذ تجارب المحاكاة على حجوم لعينات مختلفة، وهي على الآتي: $n = 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, 10, 5,4$. وفيما يتعلق بتوصيف النماذج فقد تم إجراء تجارب المحاكاة بموجب الآتي:

1. أُنموذج الانحدار الخطي:

   a. توليد بيانات ملوثة بمشاهدات شاردة.

- المتغير التوضيحي: تم توليد مشاهدات المتغير التوضيحي وفق التوزيع المنتظم $\mathrm{U}(0,1)$.
- الخطأ العشوائي: تم توليد 90% من الأَخطاء وفق التوزيع الطبيعي القياسي $N(0,1)$، و 10% وفق التوزيع الطبيعي $N(0,10)$ التي تعد على أَنها مشاهدات شاردة، ويصطلح عليها حالة نقل المتوسط (Mean shift).

b. توليد بيانات ملوثة بمشاهدات ذات القوة الرافعة (القيم الشاردة في فضاء المتغير التوضيحي)، فقد تم توليد البيانات وفقاً للآتي:

- المتغير التوضيحي: تم توليد 15% من مشاهدات المتغير التوضيحي وفق التوزيع الطبيعي $N(0,10)$.
- الخطأ العشوائي: تم توليد 75% من الأَخطاء وفق التوزيع الطبيعي القياسي $N(0,1)$، و 25% وفق التوزيع الطبيعي $N(0,10)$.

2. أُنموذج الانحدار اللاخطي:

- المتغير التوضيحي: تم توليد مشاهدات المتغير التوضيحي وفق التوزيع المنتظم $\mathrm{U}(0,1)$.
- تم توليد 90% من الاخطاء وفق التوزيع الطبيعي القياسي $N(0,1)$، 10% وفق التوزيع الطبيعي $N(0,10)$، التي تعد على أَنها مشاهدات شاردة.
- الانموذج اللاخطي وصيغته هي $y = 5\sin(3x) + e$.

ابتداءً يتم توفيق أُنموذج التقدير الخطي واللاخطي للبيانات التي يتم توليدها والمتضمنة للمشاهدات الشاردة، فضلاً عن المشاهدات ذات القوة الرافعة باستخدام طريقة المربعات الصغرى الجزائية.

و بالاعتماد على التقديرات التي يتم الحصول عليها من توفيق أُنموذجي التقدير الخطية و اللاخطي، تم توليد 1200 عينة بوتستراب (لمزيد من التفاصيل عن حجوم العينات في البوتستراب، انظر Efron & Tibshirani, 1994). بما أَنَّ بعضاً من عينات البوتستراب يمكن أَنْ تتسبب في مشاكل معينة، لذا فإن هذه العينات يتم فرزها واستبعادها من قبل دالة الفاعلية المتقاطعة المعممة. لذا فإن الـ 200 عينة من البوتستراب هي لأَجل تعويض النقص في العينات التي يتم استبعادها. ويتم توفيق جميع المتغيرات ($y_1, y_2, \dots, y_{1200}$) كمتغيرات استجابة، ثم يتم توفيق جميع عينات البوتستراب آنياً. وبالتالي إيجاد $\hat{i}^*$ والقيم $\chi_{\alpha/2}$ و $\chi_{1-\alpha/2}$.

يتضمن الانموذج قيد الدراسة متغير استجابة مستمراً، ومتغيراً توضيحياً الذي عُدَّ على أَنه متغير تمهيدي (smoothing). إنّ رتبة الاشتقاق في الجزاء (Order of Derivative in the Penalty) هي (2) ولجميع التجارب، وهي ناتج حساب الصيغة الآتية $\max(2, \text{int}\left(\frac{d}{2}\right) + 1)$، إذ إن d هي عدد المتغيرات التمهيدية (Smoothing Variables). وإن بُعُد الفضاء المتعدد الحدود (Dimension of Polynomial Space) هو (2) وهو يُمثل كلاً من الحد الثابت والمتغير التوضيحي. وفيما يتعلق بالانحراف المعياري (Standard Deviation) فيتم ايجاده من خلال الصيغة الاتية:

$$\text{SD} = \sqrt{\left(\frac{\text{RSS}(\boldsymbol{\lambda})}{\text{tr}\,(\boldsymbol{I} - \boldsymbol{A})}\right)}$$

1.2 تحليل تجارب المحاكاة:

تم تحليل نتائج تجارب المحاكاة للانموذج الخطي بحجوم عينات مختلفة في ظل وجود المشاهدات الشاردة، فضلاً عن المشاهدات ذات القوة الرافعة، وللانموذج اللاخطي بحجوم عينات مختلفة في ظل وجود المشاهدات الشاردة. وتم تفسير النتائج وفقاً للرسوم البيانية، الجزاء التمهيدي (Smoothing Penalty)، الانحراف المعياري، فضلاً عن دالة الفاعلية المتقاطعة المعممة.

الأَشكال 1، 3، و 5 هي الرسوم البيانية لما ورد في الجداول (1) الى (3) على التوالي، وذلك لغرض زيادة الإيضاح، ولتسهيل تفسير النتائج.

الأَشكال (2)، (4) و (6) تمثل المنحنيات التقديرية وحدود الثقة البيزية والبوتستراب والتقليدية باستخدام طريقة المربعات الصغرى الجزائية، ويلاحظ من هذه الرسوم دقة التقدير ولجميع تجارب المحاكاة.

1.3 مناقشة النتائج:

a. الانموذج الخطي في ظل وجود المشاهدات الشاردة:

من الشكل (2) يلاحظ أَنَّ حدود الثقة بوتستراب هي الأَفضل من حدود الثقة البيزية والتقليدية في ظل وجود المشاهدات الشاردة ولحد حجم العينة 10، ثم تصبح حدود الثقة البيزية هي الافضل. كما يلاحظ أَنَّ حدود الثقة البيزية أَفضل من التقليدية ولجميع حجوم العينات.

يلاحظ من الشكل (1) والجدول (1) بأَنَّ الـ GCV تراوحت بين 0 حينما حجم العينة 4، والـ 61.8 حينما حجم العينة 60، كما يلاحظ تناغم تذبذب الـ SD مع GCV. فيما يتعلق بالجزاء التمهيدي، فيلاحظ أَنَّ قيمته قد ارتفعت بشكل كبير عند حجم العينة 20، وبشكل أَكبر عند حجم العينة 60. أَما مايتعلق بـ (Loglambda) فهو ناتج اللوغاريتم الطبيعي لحاصل ضرب معلمة التمهيد بحجم العينة$\log_{10}(n\lambda)$.

b. الانموذج الخطي في ظل وجود المشاهدات ذات القوة الرافعة:

من الشكل (4) يلاحظ أَنَّ حدود الثقة بوتستراب هي الأَفضل من حدود الثقة البيزية والتقليدية في ظل وجود المشاهدات الشاردة ولجميع حجوم العينات. كما يلاحظ أَنَّ حدود الثقة البيزية أَفضل من التقليدية ولجميع حجوم العينات.

يلاحظ من الشكل (2) والجدول (2) بأَنَّ الـ GCV تراوحت بين 57 حينما حجم العينة 90، والـ 1719 حينما حجم العينة 40. فيما يتعلق بالجزاء التمهيدي، فيلاحظ أَنه مستقر عند الصفر ولجميع حجوم العينات.

c. الانموذج اللاخطي في ظل وجود المشاهدات الشاردة:

من الشكل (6) يلاحظ أَنّ حدود ثقة البوتستراب أَفضل من حدود الثقة البيزية والتقليدية ولجميع التجارب. كما يلاحظ أَنَّ هناك تشابهاً كبيراً بين حدود الثقة البيزية والتقليدية لجميع التجارب ماعدا حينما يكون عدد المشاهدات 4، اذ تتفوق حدود الثقة البيزية على كل من حدود الثقة بوتستراب والتقليدية.

يلاحظ من الشكل (3) والجدول (3) بأَنَّ الـ GCV تراوحت بين 3.07−0.7 لاحجام العينات من 100 الى حجم العينة 5، إلا أَنها قفزت الى 54.3 عند حجم العينة 4، وبالرجوع الى الشكل (6) يلاحظ أَنَّ تشتت النقاط كان كبيراً، مما أَدى الى عدم تمثيل النقاط بشكل جيد، وهذا انعكس بدوره على الـ SD، وكما هو موضح ذلك في الشكل (5). أَمّا فيما يتعلق بالجزاء التمهيدي فيلاحظ أَنَّ قيمته قد ارتفعت بشكل كبير عند حجم العينة 60 وبشكل اكبر عند حجم العينة 20، وعند الرجوع الى الشكل (5)، ويلاحظ أَنَّ المنحني يعاني من تذبذب عند حجم العينة 20، وتذبذب أَكبر عند حجم العينة 60.

1.4 الاستنتاجات:

1. تفوق حدود الثقة بوتستراب على حدود الثقة البيزية والتقليدية لتجارب المحاكاة ولحجوم العينات أَكبر، او يساوي الـ 10 حينما تكون البيانات ملوثة بالنقاط الشاردة للانموذج الخطي.

2. أظهرت حدود الثقة بوتستراب تفوقاً على حدود الثقة البيزية والتقليدية لتجارب المحاكاة ولجميع العينات حينما تكون البيانات ملوثة بالنقاط ذات القوة الرافعة للانموذج الخطي.
3. تقدم حدود الثقة بوتستراب على حدود الثقة البيزية والتقليدية لتجارب المحاكاة وبحجوم العينات اكبر، او يساوي الـ 5 حينما تكون البيانات ملوثة بالنقاط الشاردة للانموذج اللاخطي.
4. لوحظ أنَّ حدود الثقة البيزية تتفوق على حدود الثقة التقليدية لجميع حجوم العينات ولجميع تجارب البحث.

المصادر

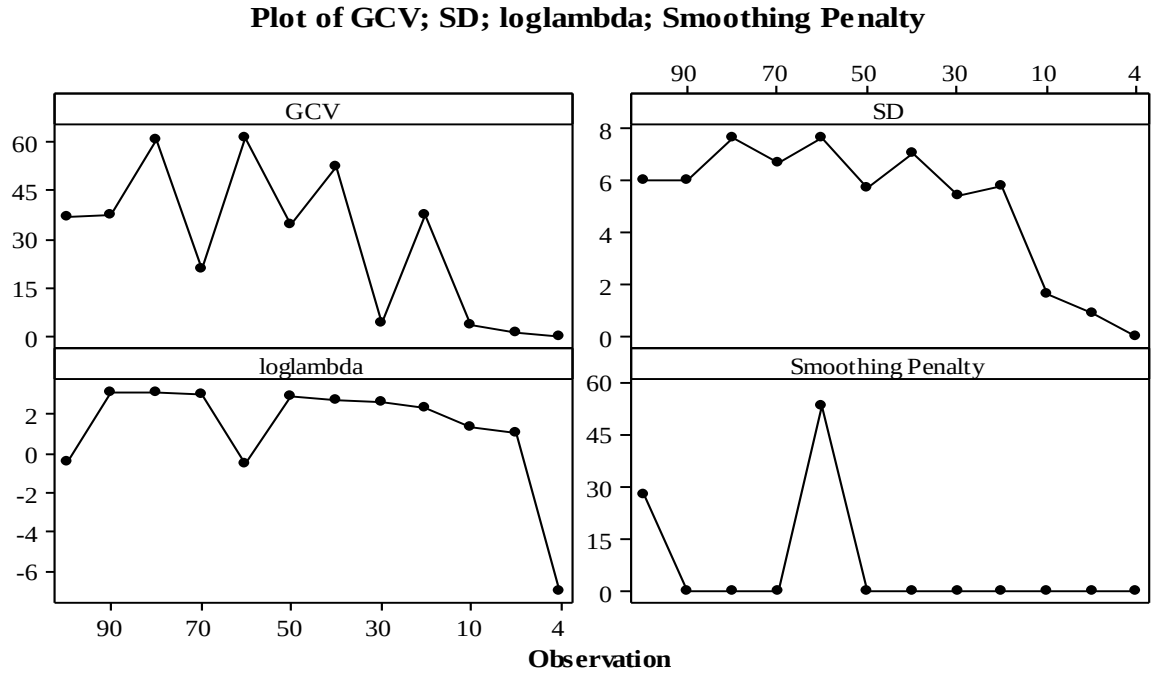
1. رشيد، ظافر حسين وخلود يوسف خمو، (2005)م " مقارنة الطرائق الشرائحية لتقدير منحنى الانحدار اللامعلمي" ، المجلة العراقية للعلوم الاحصائية، العدد 8، ص- ص: 40-62.
2. Bates, D. M., Lindstrom, M. J., Wahba, G., & Yandell, B. S. (1987). Gcvpack–routines for generalized cross validation: Gcvpack–routines for generalized. Communications in Statistics-simulation and Computation, 16(1), 263-297.
3. Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive theory of functions of several variables* (pp. 85-100). Springer, Berlin, Heidelberg.
4. Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics, 21(2), 215-223.
5. *Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC press.*
6. *Nosedal-Sanchez, A., Storlie, C. B., Lee, T. C., & Christensen, R. (2012). Reproducing kernel Hilbert spaces for penalized regression: A tutorial. The American Statistician, 66(1), 50-60.*
7. Nosedal-Sanchez, A., Storlie, C. B., Lee, T. C., & Christensen, R. (2012). Reproducing kernel Hilbert spaces for penalized regression: A tutorial. *The American Statistician*, *66*(1), 50-60.
8. Reed, A. H., Cannon, D. C., Pileggi, V. J., & Winkelman, J. W. (1973). Use of confidence intervals to assess precision of normal range estimates. *Clinical biochemistry*, *6*, 29-33.
9. Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). Semiparametric regression (No. 12). Cambridge university press.

10. Simonoff, J. S. (2012). *Smoothing methods in statistics*. Springer Science & Business Media.
11. *Ruan, Z., Guo, H., Liu, H., & Yan, S. (2013). Glacier surface velocity estimation in the West Kunlun Mountain range from L-band ALOS/PALSAR images using modified synthetic aperture radar offset-tracking procedure. Journal of Applied Remote Sensing, 8(1), 084595.*
12. *Wahba, G. (1983). Bayesian" confidence intervals" for the cross-validated smoothing spline. Journal of the Royal Statistical Society. Series B (Methodological), 133-150.*
13. Wahba, G. (1990). *Spline models for observational data*. Society for industrial and applied mathematics.
14. Wang, Y., & Wahba, G. (1995). Bootstrap confidence intervals for smoothing splines and their comparison to Bayesian confidence intervals. Journal of Statistical Computation and Simulation, 51(2-4), 263-279.
15. Wang, Y. (2011). Smoothing splines: methods and applications. CRC Press.
16. Wang, Y. (2011). Smoothing splines: methods and applications. Chapman and Hall/CRC.
17. Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95-114.
18. Terry, L., & Kelley, K. (2012). Sample size planning for composite reliability coefficients: Accuracy in parameter estimation via narrow confidence intervals. *British Journal of Mathematical and Statistical Psychology*, *65*(3), 301.

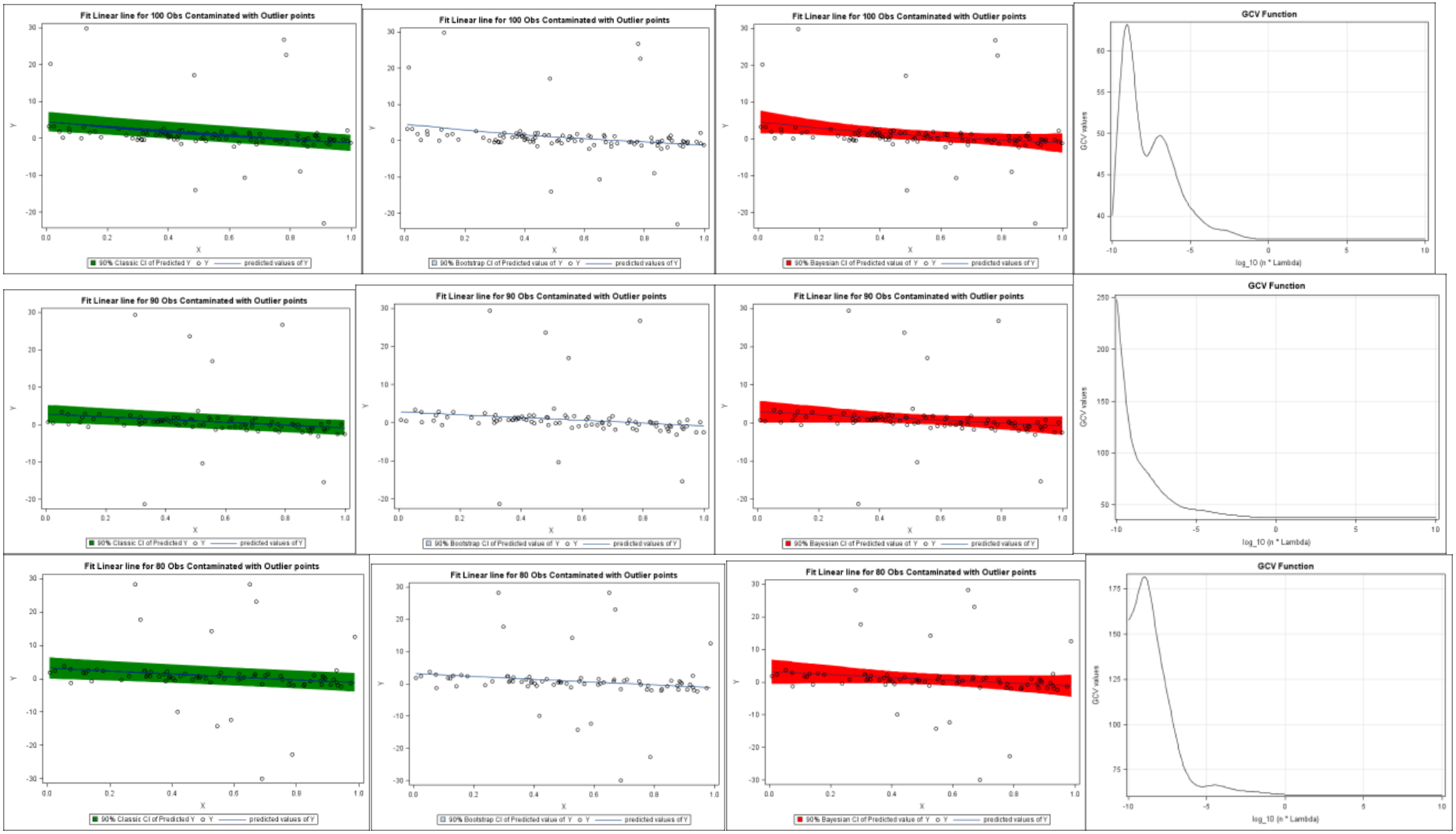
الملحق

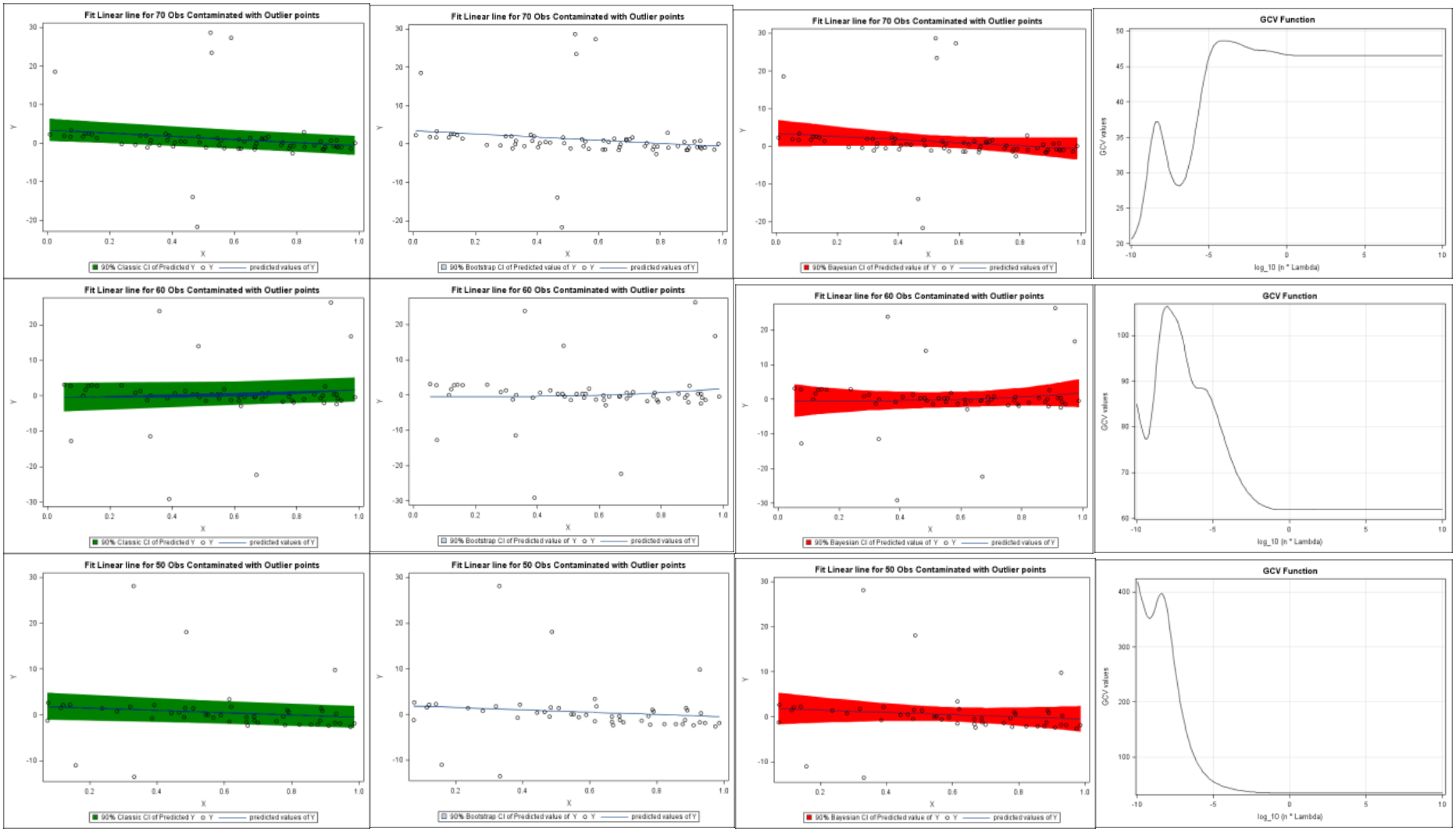
الجدول (1): يوضح الإحصاءات للتقديرات النهائية للانموذج الخطي في ظل وجود المشاهدات الشاردة، ولحجوم عينات مختلفة

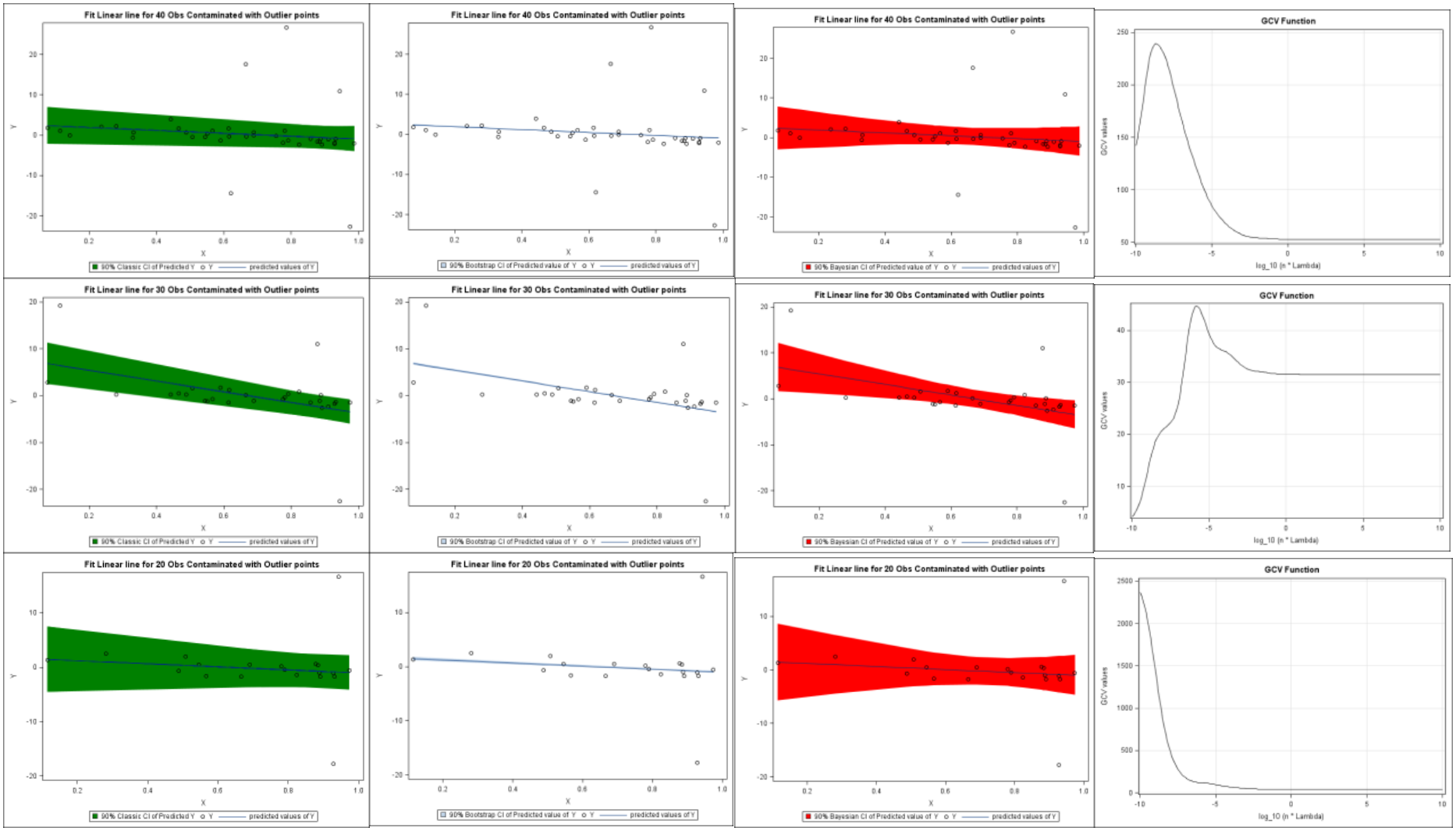
NO of observations	loglambda	SMOOTHING PENALTY	Residual	DF	SD	GCV
100	-0.3187	27.7405	3547.9786	2.3713	6.0284	37.224314
90	3.2303	0.0000	3232.5208	2.0001	6.0608	37.568138
80	3.1777	0.0000	4645.6375	2.0001	7.7175	61.086662
70	3.1484	0.0000	3073.5613	2.0001	6.7231	20.636316
60	-0.4702	53.3494	3430.1515	2.3026	7.7104	61.823631
50	2.9526	0.0000	1596.5690	2.0001	5.7673	34.647872
40	2.8188	0.0000	1905.6808	2.0001	7.0816	52.789193
30	2.7059	0.0000	824.0716	2.0001	5.4251	4.087004
20	2.4289	0.0000	612.3235	2.0001	5.8325	37.797773
10	1.4288	0.0000	21.9219	2.0001	1.6554	3.425383
5	1.1420	0.0000	2.1475	2.0001	0.8461	1.193091
4	-6.9773	0.0000	0.0000	3.5107	0.0000	0.0000

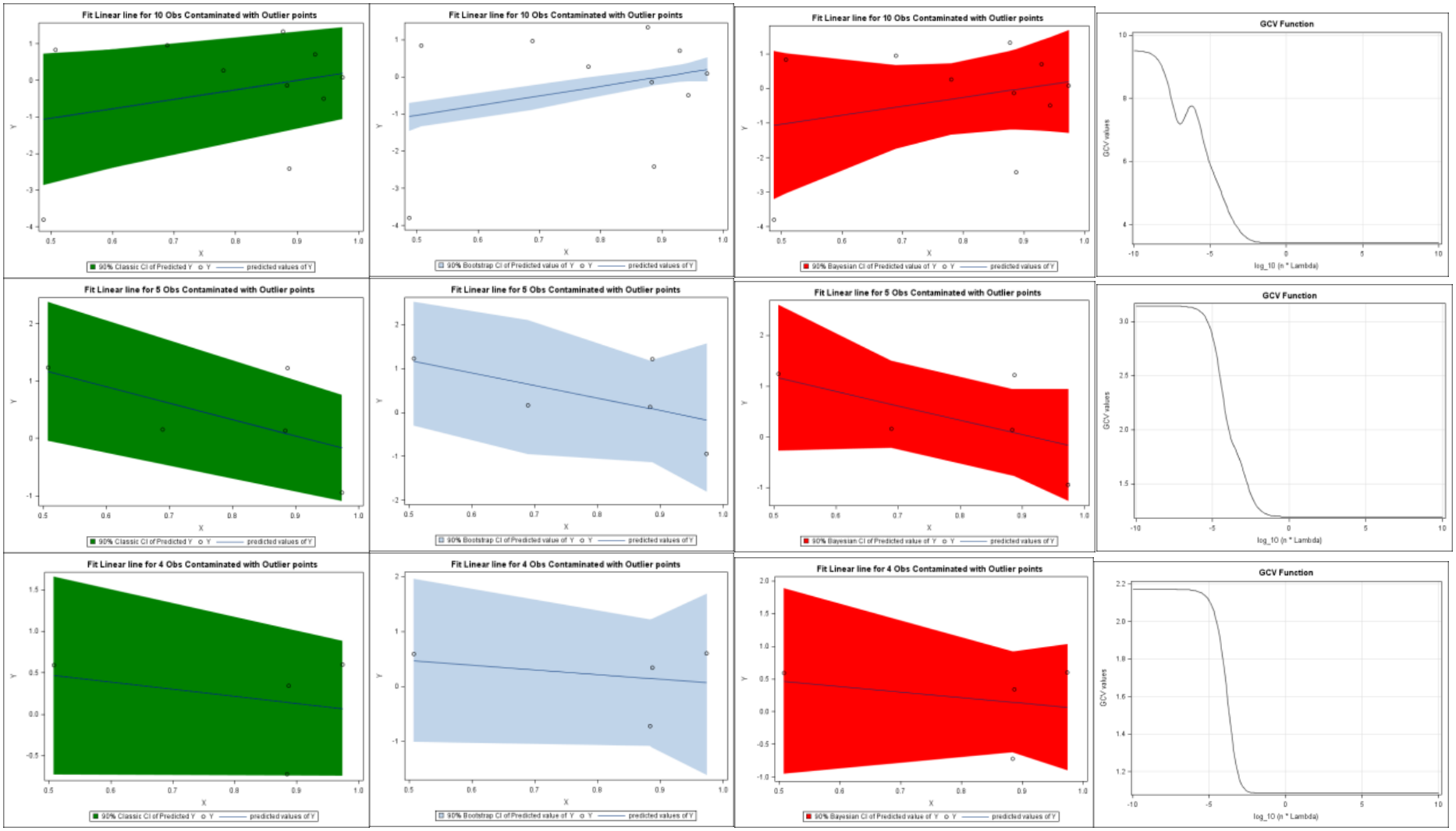


الشكل (1): يوضح الرسوم البيانية لإحصاءات التقديرات النهائية الواردة في الجدول (1) للانموذج الخطي في ظل وجود المشاهدات الشاردة، ولحجوم عينات مختلفة


Fit Linear line for 100 Obs Contaminated with Outlier points
90% Classic CI of Predicted Y
Y
predicted values of Y
Fit Linear line for 100 Obs Contaminated with Outlier points
90% Bootstrap CI of Predicted value of Y
Y
predicted values of Y
Fit Linear line for 100 Obs Contaminated with Outlier points
90% Bayesian CI of Predicted value of Y
Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 90 Obs Contaminated with Outlier points
90% Classic CI of Predicted Y
Y
predicted values of Y
Fit Linear line for 90 Obs Contaminated with Outlier points
90% Bootstrap CI of Predicted value of Y
Y
predicted values of Y
Fit Linear line for 90 Obs Contaminated with Outlier points
90% Bayesian CI of Predicted value of Y
Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 80 Obs Contaminated with Outlier points
90% Classic CI of Predicted Y
Y
predicted values of Y
Fit Linear line for 80 Obs Contaminated with Outlier points
90% Bootstrap CI of Predicted value of Y
Y
predicted values of Y
Fit Linear line for 80 Obs Contaminated with Outlier points
90% Bayesian CI of Predicted value of Y
Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)


Fit Linear line for 70 Obs Contaminated with Outlier points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 70 Obs Contaminated with Outlier points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 70 Obs Contaminated with Outlier points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 60 Obs Contaminated with Outlier points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 60 Obs Contaminated with Outlier points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 60 Obs Contaminated with Outlier points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 50 Obs Contaminated with Outlier points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 50 Obs Contaminated with Outlier points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 50 Obs Contaminated with Outlier points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)

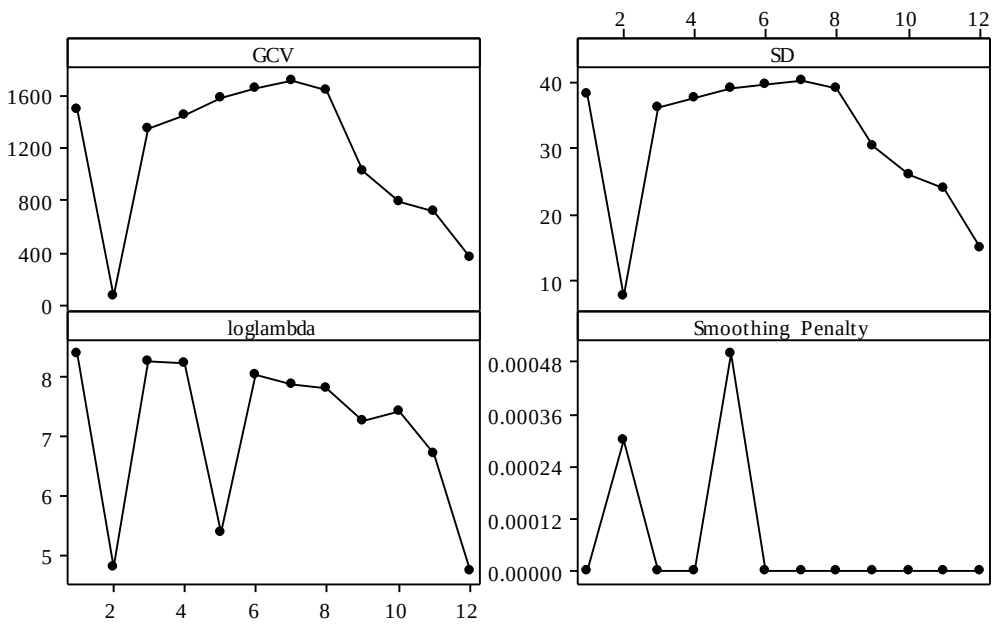

Fit Linear line for 40 Obs Contaminated with Outlier points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 40 Obs Contaminated with Outlier points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 40 Obs Contaminated with Outlier points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 30 Obs Contaminated with Outlier points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 30 Obs Contaminated with Outlier points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 30 Obs Contaminated with Outlier points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 20 Obs Contaminated with Outlier points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 20 Obs Contaminated with Outlier points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 20 Obs Contaminated with Outlier points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)


Fit Linear line for 10 Obs Contaminated with Outlier points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 10 Obs Contaminated with Outlier points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 10 Obs Contaminated with Outlier points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 5 Obs Contaminated with Outlier points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 5 Obs Contaminated with Outlier points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 5 Obs Contaminated with Outlier points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 4 Obs Contaminated with Outlier points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 4 Obs Contaminated with Outlier points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 4 Obs Contaminated with Outlier points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)

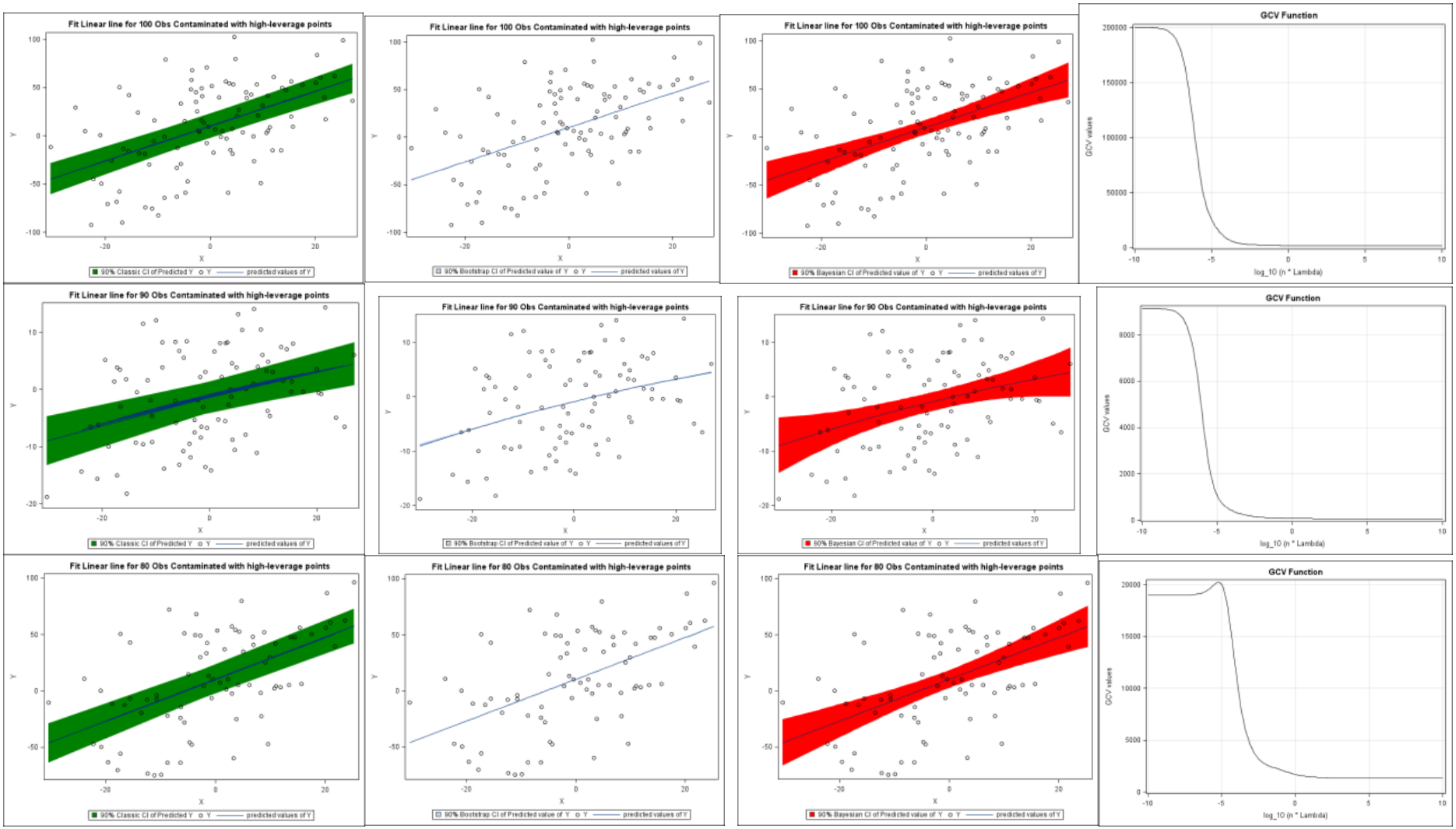
الشكل (1): يوضح الرسوم البيانية لكل من حدود الثقة بوتستراب والبيزية والكلاسيكية، فضلاً عن الرسم البياني لدالة الصلاحية التقاطع المعممة للانموذج الخطي في ظل وجود المشاهدات الشاردة ولحجوم عينات مختلفة.

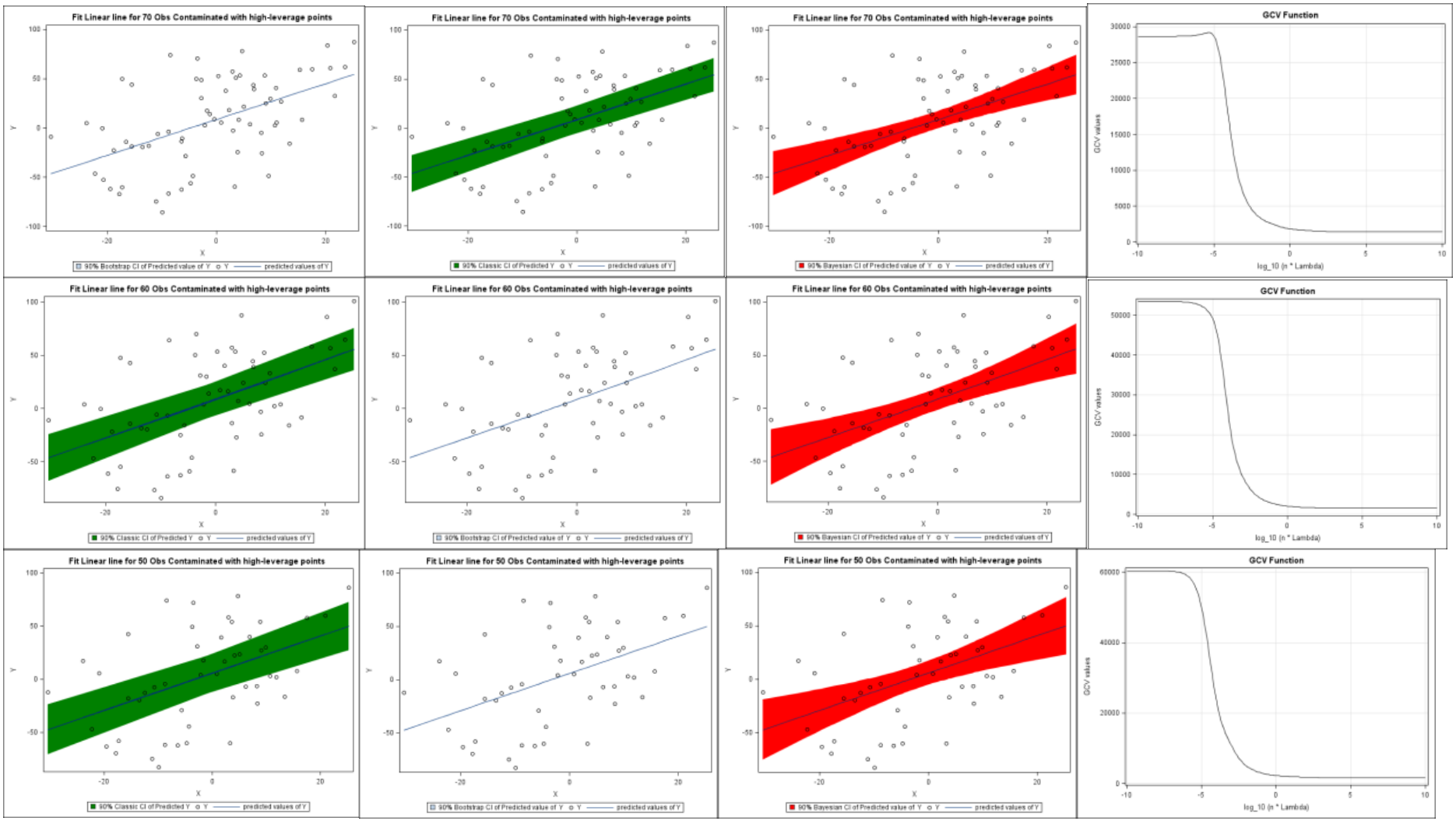
الجدول (2): يوضح الاحصاءات للتقديرات النهائية للانموذج الخطي في ظل وجود المشاهدات ذات القوة الرافعة ولحجوم عينات مختلفة

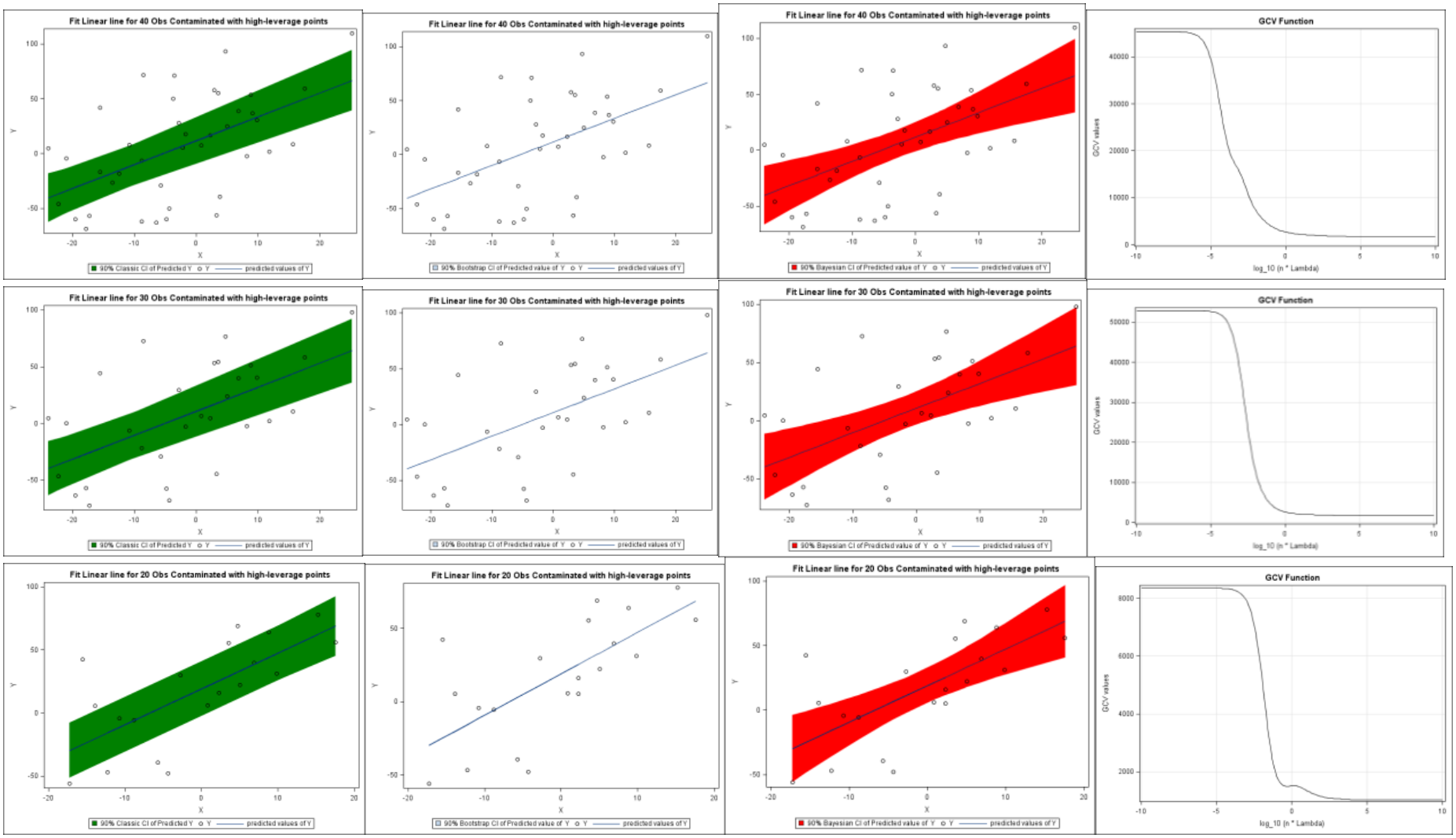
NO of observations	loglambda	SMOOTHING PENALTY	Residual	DF	SD	GCV
100	8.4041	0.0000	143428.561	2.0001	38.2565	1493.42637
90	4.8019	0.0003	4929.3933	2.3314	7.4985	57.566274
80	8.2682	0.0000	102752.408	2.0001	36.2952	1349.71252
70	8.2258	0.0000	95948.9183	2.0001	37.5635	1452.51794
60	5.3884	0.0005	88187.2184	2.0741	39.0181	1576.92778
50	8.0470	0.0000	76176.7415	2.0001	39.8374	1653.14797
40	7.8807	0.0000	62057.5531	2.0001	40.4116	1719.05042
30	7.8290	0.0000	43030.7973	2.0001	39.2023	1646.59402
20	7.2660	0.0000	15605.2598	2.0001	30.2979	1025.96014
15	7.4396	0.0000	8871.0250	2.0001	26.1227	787.372401
10	6.7249	0.0000	4579.9019	2.0001	23.9269	715.623698
5	4.7226	0.0000	652.0314	2.0001	14.7429	362.248500

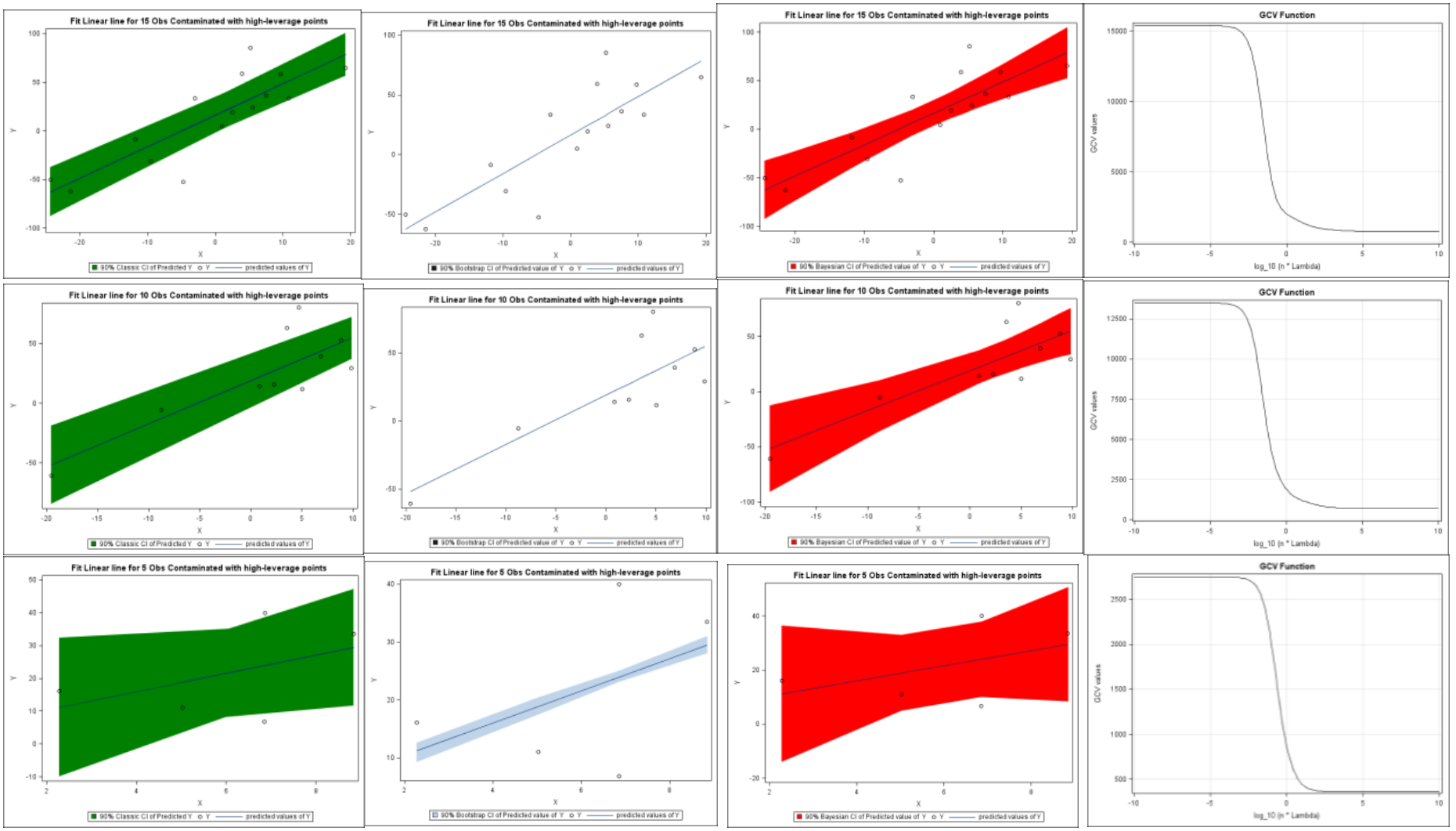


الشكل (3): يوضح الرسوم البيانية لاحصاءات التقديرات النهائية الواردة في الجدول (2) للانموذج الخطي في ظل وجود المشاهدات ذات القوة الرافعة ولحجوم عينات مختلفة


Fit Linear line for 100 Obs Contaminated with high-leverage points
90% Classic CI of Predicted Y
Y
predicted values of Y
Fit Linear line for 100 Obs Contaminated with high-leverage points
90% Bootstrap CI of Predicted value of Y
Y
predicted values of Y
Fit Linear line for 100 Obs Contaminated with high-leverage points
90% Bayesian CI of Predicted value of Y
Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 90 Obs Contaminated with high-leverage points
90% Classic CI of Predicted Y
Y
predicted values of Y
Fit Linear line for 90 Obs Contaminated with high-leverage points
90% Bootstrap CI of Predicted value of Y
Y
predicted values of Y
Fit Linear line for 90 Obs Contaminated with high-leverage points
90% Bayesian CI of Predicted value of Y
Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 80 Obs Contaminated with high-leverage points
90% Classic CI of Predicted Y
Y
predicted values of Y
Fit Linear line for 80 Obs Contaminated with high-leverage points
90% Bootstrap CI of Predicted value of Y
Y
predicted values of Y
Fit Linear line for 80 Obs Contaminated with high-leverage points
90% Bayesian CI of Predicted value of Y
Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)


Fit Linear line for 70 Obs Contaminated with high-leverage points
95% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 70 Obs Contaminated with high-leverage points
95% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 70 Obs Contaminated with high-leverage points
95% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 60 Obs Contaminated with high-leverage points
95% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 60 Obs Contaminated with high-leverage points
95% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 60 Obs Contaminated with high-leverage points
95% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 50 Obs Contaminated with high-leverage points
95% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 50 Obs Contaminated with high-leverage points
95% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 50 Obs Contaminated with high-leverage points
95% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)

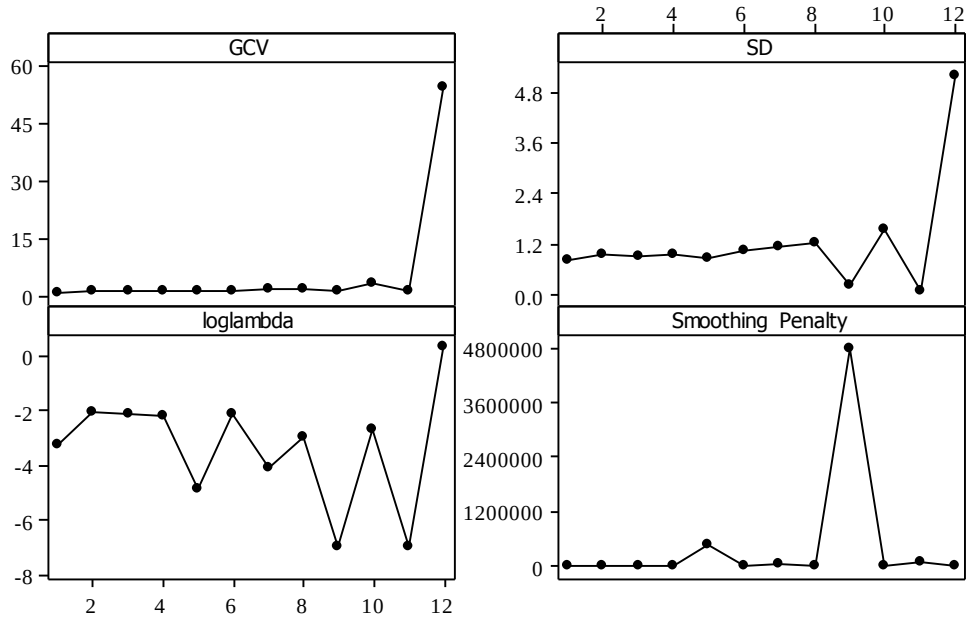

Fit Linear line for 40 Obs Contaminated with high-leverage points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 40 Obs Contaminated with high-leverage points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 40 Obs Contaminated with high-leverage points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 30 Obs Contaminated with high-leverage points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 30 Obs Contaminated with high-leverage points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 30 Obs Contaminated with high-leverage points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 20 Obs Contaminated with high-leverage points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 20 Obs Contaminated with high-leverage points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 20 Obs Contaminated with high-leverage points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)


Fit Linear line for 15 Obs Contaminated with high-leverage points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 15 Obs Contaminated with high-leverage points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 15 Obs Contaminated with high-leverage points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 10 Obs Contaminated with high-leverage points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 10 Obs Contaminated with high-leverage points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 10 Obs Contaminated with high-leverage points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Linear line for 5 Obs Contaminated with high-leverage points
90% Classic CI of Predicted Y
predicted values of Y
Fit Linear line for 5 Obs Contaminated with high-leverage points
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Linear line for 5 Obs Contaminated with high-leverage points
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)

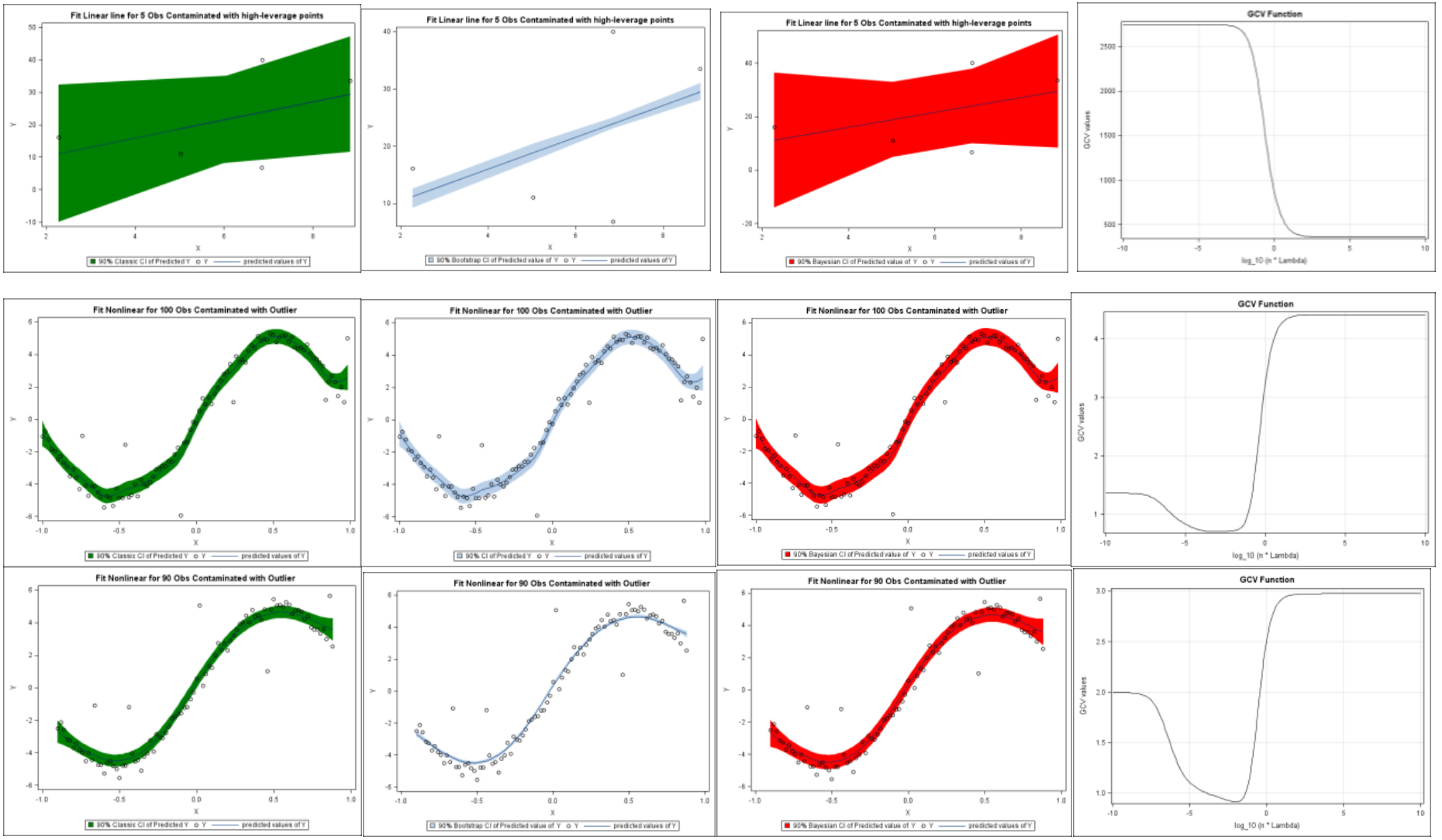
الشكل (4): يوضح الرسوم البيانية لكل من حدود الثقة بوتستراب والبيزية والكلاسيكية، فضلاً عن الرسم البياني لدالة الصلاحية التقاطع المعممة للانموذج الخطي في ظل وجود المشاهدات ذات القوة الرافعة ولحجوم عينات مختلفة.

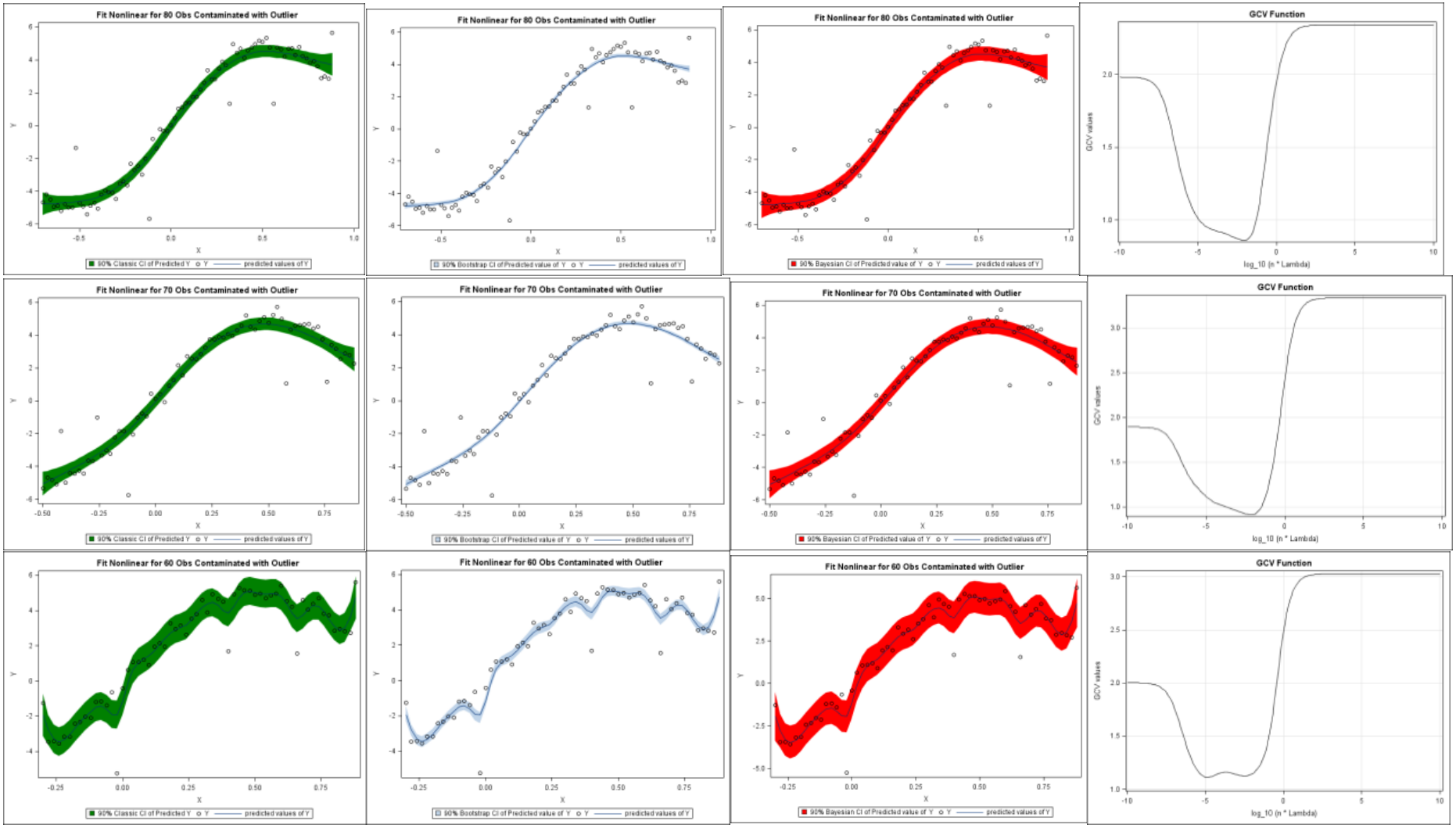
الجدول (3): يوضح الاحصاءات للتقديرات النهائية للانموذج اللاخطي في ظل وجود المشاهدات الشاردة

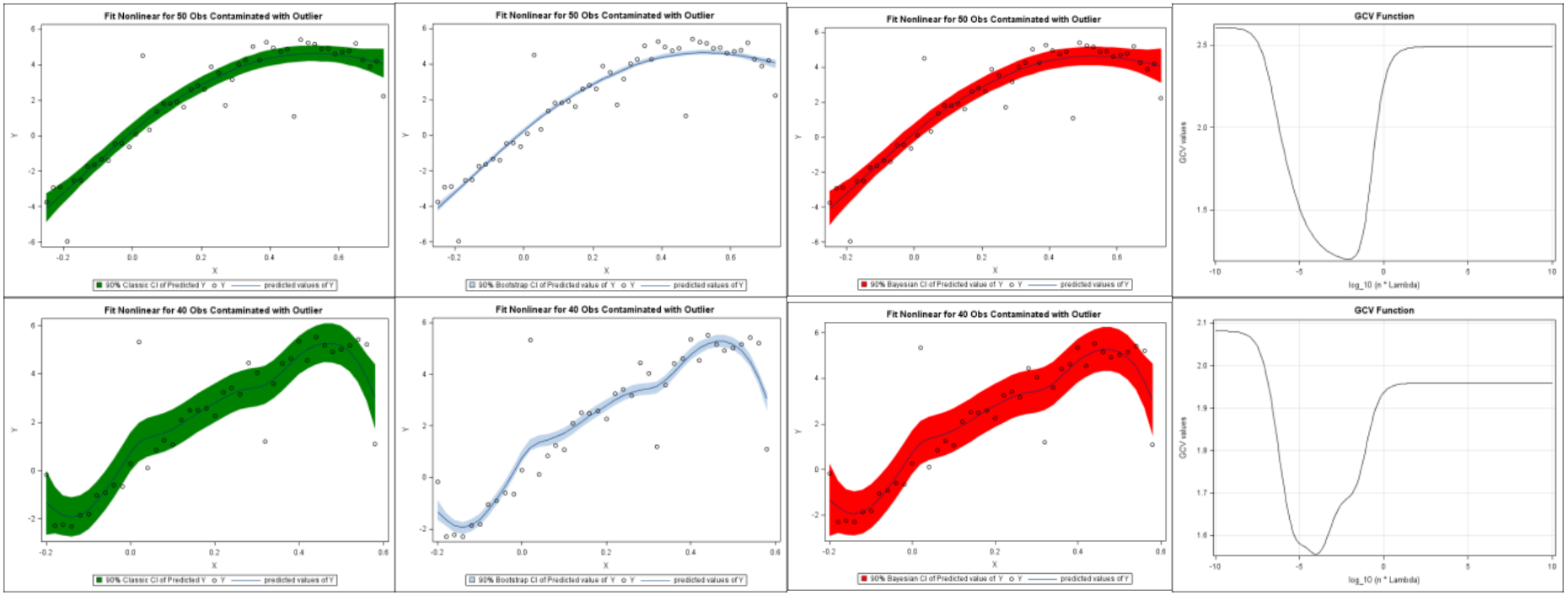
NO of observations	loglambda	SMOOTHING PENALTY	Residual	DF	SD	GCV
100	-3.2311	7055.6430	53.5584	13.0682	0.7849	0.708726
90	-2.0069	1215.1608	70.9038	6.3704	0.9208	0.912419
80	-2.0935	1006.5407	58.8419	6.0172	0.8918	0.860036
70	-2.1718	1103.3531	54.1068	5.5919	0.9165	0.913078
60	-4.8731	486287.3070	30.3667	19.5572	0.8665	1.114034
50	-2.1157	647.0535	50.4123	4.1750	1.0489	1.200351
40	-4.0394	50432.7360	38.1933	8.6672	1.1041	1.556276
30	-2.9831	2655.8819	38.9258	4.1340	1.2267	1.745452
20	-6.9773	4796272.1017	0.0352	19.1236	0.2005	0.905041
10	-2.6935	580.6901	17.3717	2.4862	1.5205	3.077003
5	-6.9773	92441.7743	0.0001	4.9823	0.0568	0.907823
4	0.3291	0.0013	54.2786	2.0001	5.2097	54.284200

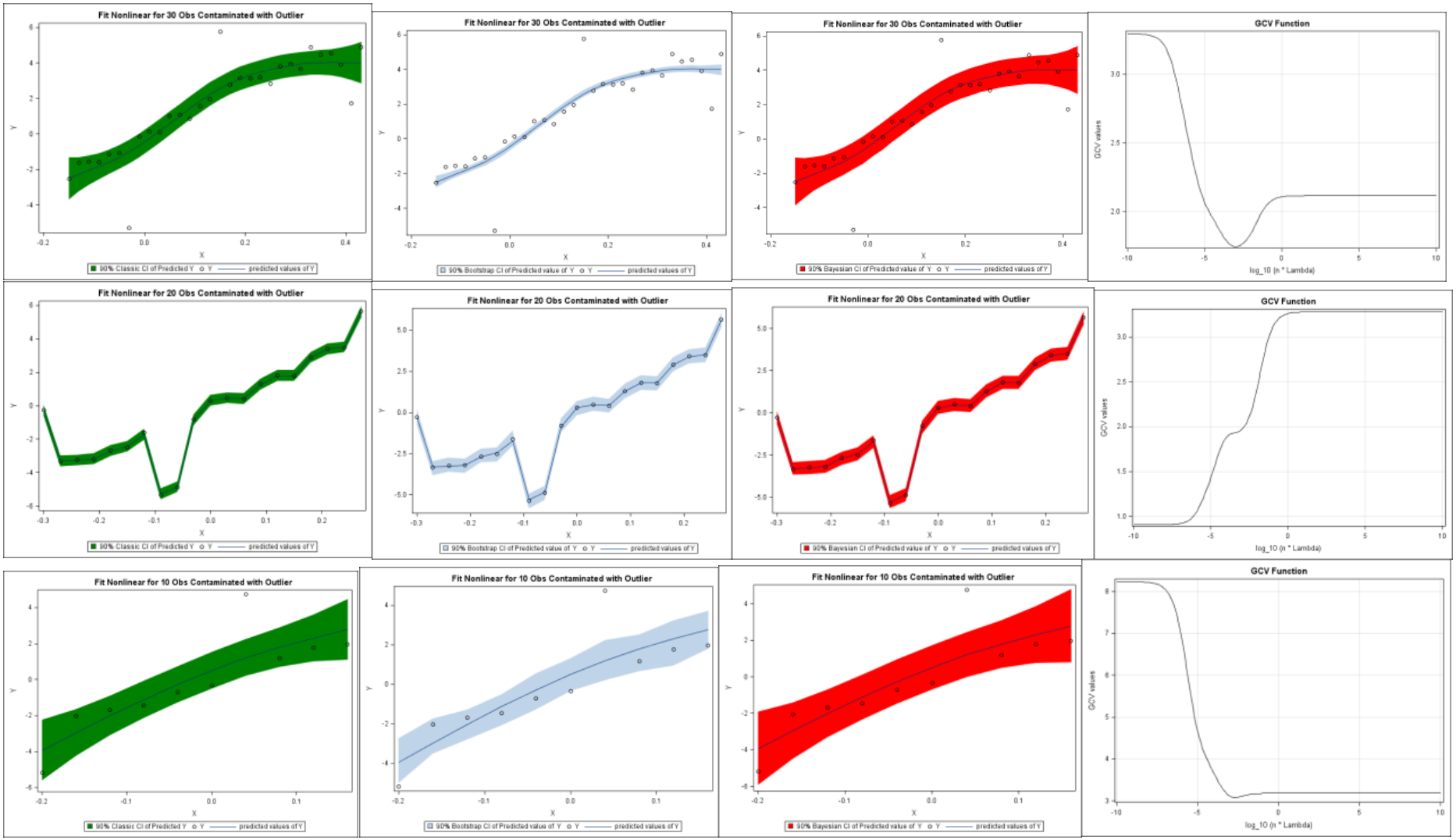


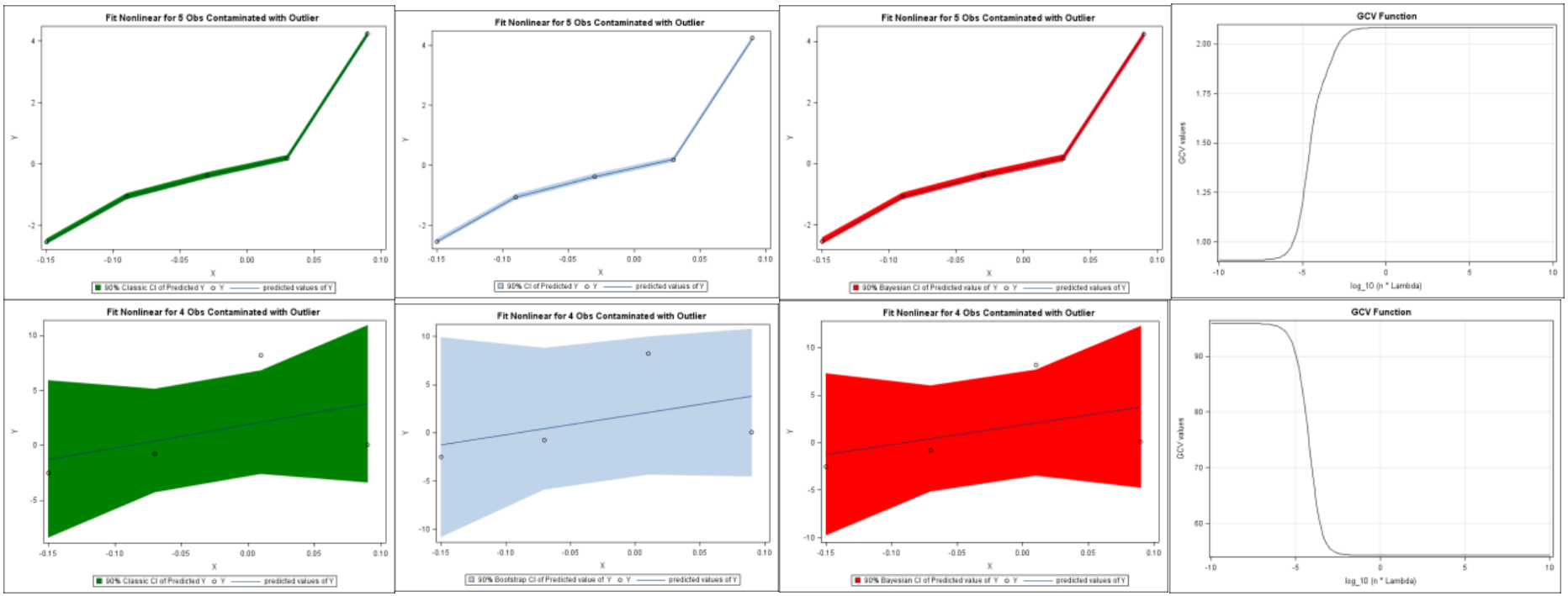
الشكل (5): يوضح الرسوم البيانية لإحصاءات التقديرات النهائية الواردة في الجدول (3) للانموذج اللاخطي في ظل وجود المشاهدات الشاردة، ولحجوم عينات مختلفة


Fit Linear line for 5 Obs Contaminated with high-leverage points
90% Classic CI of Predicted Y
Y
predicted values of Y
Fit Linear line for 5 Obs Contaminated with high-leverage points
90% Bootstrap CI of Predicted value of Y
Y
predicted values of Y
Fit Linear line for 5 Obs Contaminated with high-leverage points
90% Bayesian CI of Predicted value of Y
Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Nonlinear for 100 Obs Contaminated with Outlier
90% Classic CI of Predicted Y
Y
predicted values of Y
Fit Nonlinear for 100 Obs Contaminated with Outlier
90% CI of Predicted Y
Y
predicted values of Y
Fit Nonlinear for 100 Obs Contaminated with Outlier
90% Bayesian CI of Predicted value of Y
Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Nonlinear for 90 Obs Contaminated with Outlier
90% Classic CI of Predicted Y
Y
predicted values of Y
Fit Nonlinear for 90 Obs Contaminated with Outlier
90% Bootstrap CI of Predicted value of Y
Y
predicted values of Y
Fit Nonlinear for 90 Obs Contaminated with Outlier
90% Bayesian CI of Predicted value of Y
Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)


Fit Nonlinear for 80 Obs Contaminated with Outlier
90% Classic CI of Predicted Y
predicted values of Y
Fit Nonlinear for 80 Obs Contaminated with Outlier
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Nonlinear for 80 Obs Contaminated with Outlier
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Nonlinear for 70 Obs Contaminated with Outlier
90% Classic CI of Predicted Y
predicted values of Y
Fit Nonlinear for 70 Obs Contaminated with Outlier
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Nonlinear for 70 Obs Contaminated with Outlier
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Nonlinear for 60 Obs Contaminated with Outlier
90% Classic CI of Predicted Y
predicted values of Y
Fit Nonlinear for 60 Obs Contaminated with Outlier
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Nonlinear for 60 Obs Contaminated with Outlier
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)


Fit Nonlinear for 50 Obs Contaminated with Outlier
90% Classic CI of Predicted Y
Y
predicted values of Y
Fit Nonlinear for 50 Obs Contaminated with Outlier
90% Bootstrap CI of Predicted value of Y
Y
predicted values of Y
Fit Nonlinear for 50 Obs Contaminated with Outlier
90% Bayesian CI of Predicted value of Y
Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Nonlinear for 40 Obs Contaminated with Outlier
90% Classic CI of Predicted Y
Y
predicted values of Y
Fit Nonlinear for 40 Obs Contaminated with Outlier
90% Bootstrap CI of Predicted value of Y
Y
predicted values of Y
Fit Nonlinear for 40 Obs Contaminated with Outlier
90% Bayesian CI of Predicted value of Y
Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)


Fit Nonlinear for 30 Obs Contaminated with Outlier
90% Classic CI of Predicted Y
predicted values of Y
Fit Nonlinear for 30 Obs Contaminated with Outlier
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Nonlinear for 30 Obs Contaminated with Outlier
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Nonlinear for 20 Obs Contaminated with Outlier
90% Classic CI of Predicted Y
predicted values of Y
Fit Nonlinear for 20 Obs Contaminated with Outlier
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Nonlinear for 20 Obs Contaminated with Outlier
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)
Fit Nonlinear for 10 Obs Contaminated with Outlier
90% Classic CI of Predicted Y
predicted values of Y
Fit Nonlinear for 10 Obs Contaminated with Outlier
90% Bootstrap CI of Predicted value of Y
predicted values of Y
Fit Nonlinear for 10 Obs Contaminated with Outlier
90% Bayesian CI of Predicted value of Y
predicted values of Y
GCV Function
GCV values
log_10 (n * Lambda)



الشكل (6): يوضح الرسوم البيانية لكل من حدود الثقة بوتستراب والبيزية والكلاسيكية، فضلاً عن الرسم البياني لدالة الصلاحية التقاطع المعممة للانموذج اللاخطي في ظل وجود المشاهدات الشاردة ولحجوم عينات مختلفة.