



تطبيق النماذج المركبة في توفيق بيانات مطالبات التأمينات العامة

إعداد

د. محمد مسعد المعداوي

مدرس بقسم الاحصاء والتأمين - كلية التجارة - جامعة الزقازيق

moh_elmadawye@yahoo.com

المجلة العلمية للدراسات والبحوث المالية والتجارية

كلية التجارة - جامعة دمياط

المجلد السادس - العدد الثاني - الجزء الرابع - يوليو ٢٠٢٥

التوثيق المقترح وفقاً لنظام APA:

المعداوي، محمد مسعد. (٢٠٢٥). تطبيق النماذج المركبة في توفيق بيانات مطالبات التأمينات العامة،
المجلة العلمية للدراسات والبحوث المالية والتجارية، كلية التجارة، جامعة دمياط، ٦(٢)٤، ١٠٧-
١٣١.

رابط المجلة: <https://cfdj.journals.ekb.eg/>

تطبيق النماذج المركبة في توفيق بيانات مطالبات

التأمينات العامة

د. محمد مسعد المعداوي

ملخص البحث:

يهدف هذا البحث إلى استخدام النماذج المركبة للتوزيع اللوغاريتمي الطبيعي-باريتو (Composite Lognormal-Pareto) والنماذج المركبة لتوزيع وايبل - باريتو (Composite Weibull-Pareto) في نمذجة بيانات المطالبات في شركات التأمين، ثم المقارنة بين هذه النماذج لاختيار أفضلهم من خلال إيجاد المقاييس الإحصائية (NLL, AIC, BIC)، وبالتالي الوصول إلى أفضل نموذج، ثم محاولة تحسين هذا النموذج للوصول إلى قيمة (threshold point) المثلى التي تحقق أقل قيمة (AIC, BIC) في النموذج الأفضل. وكانت أهم النتائج التي توصلت إليها الدراسة أن النموذج Composite Lognormal-Pareto الثالث هو الأفضل لملاءمة للبيانات لأنه يحقق أقل قيم لـ (NLL, AIC, BIC)، نماذج Composite Weibull-Pareto تؤدي أداءً جيداً ولكنها ليست قوية مثل نموذج Composite Lognormal-Pareto، وعند تحسين النموذج الثالث Composite Lognormal-Pareto، كانت القيمة المثلى $(\theta) = 19000$ ، وكانت $AIC = 3093.282$ ، أما عند تحسين النموذج الثالث Composite Weibull-Pareto، كانت القيمة المثلى $(\theta) = 20000$ وكانت $AIC = 3255.723$ ، $BIC = 3267.162$ ، وهذا يشير إلى جودة ملاءمة النموذج، وإتفاق معيار (AIC, BIC) على نفس القيمة المثلى يعكس استقرار النموذج وقوته الإحصائية. وقد أوصت الدراسة باستخدام نموذج Composite Lognormal-Pareto الثالث لأنه النموذج الأفضل لنمذجة بيانات مطالبات التأمين، وأهمية استخدام توزيع باريتو المعمم في نمذجة مطالبات التأمين ذات الذيل الثقيل بدقة. وأيضاً استخدام نماذج مركبة أخرى في نمذجة مطالبات التأمين، كما توصى الدراسة بالجمع بين النماذج المركبة وتقنيات التنقيب في البيانات (Data Mining) لتحليل بيانات المطالبات بشكل أعمق واستخلاص نتائج أكثر دقة.

الكلمات المفتاحية: النماذج المركبة - النموذج المركب (اللوغاريتمي الطبيعي-باريتو) - توزيع باريتو المعمم - النموذج المركب (وايبل-باريتو) - AIC - BIC

مقدمة:

في الدراسات الاكتوارية، يتم استخدام التوزيع اللوغاريتمي الطبيعي وتوزيع باريتو بشكل متكرر لنمذجة بيانات مطالبات التأمين، لأن هذه البيانات غالباً ما تكون ذات إلتواء موجب بدرجة عالية، ويتم استخدام توزيع باريتو غالباً لنمذجة قيم البيانات الكبيرة، حيث يتميز بذيله الطويل والسُميك ناحية اليمين. ومن ناحية أخرى، نجد أن التوزيع اللوغاريتمي الطبيعي غالباً ما يكون مناسباً للمطالبات الكبيرة والأقل تكراراً وللمطالبات الصغيرة والأعلى تكراراً. وعلى الرغم من أن النموذج اللوغاريتمي الطبيعي يغطي المطالبات الكبيرة ذات التكرارات الأقل، إلا أنه يتلاشى إلى الصفر بشكل أسرع من نموذج باريتو، مما يجعله أقل كفاءة في التعامل مع المطالبات ذات القيم الكبيرة. علاوة على ذلك، فإن نموذج باريتو لا يكون مناسباً للمطالبات ذات القيم الصغيرة نظراً لشكل دالة كثافته الاحتمالية

المتناقضة. ولذلك، مع الأخذ في الاعتبار سلوك الذيل للمطالبات الصغيرة والكبيرة، تم استخدام النماذج المركبة لنمذجة بيانات مطالبات التأمين، حيث يجمع النموذج بين خصائص التوزيعين للإستفادة من دقة التوزيع المناسب في الجزء من المطالبات ذات القيم الصغيرة وقوة التوزيع الأخر المناسب في الجزء من المطالبات ذات القيم الكبيرة. وتتمثل النماذج المركبة في النموذج المركب اللوغاريتمي الطبيعي- باريتو (Composite Lognormal-Pareto)، والنموذج المركب واييل- باريتو (Composite Weibull-Pareto)، حيث أن هذه النماذج تعتمد على دالة كثافة الاحتمال للتوزيع اللوغاريتمي الطبيعي أو واييل، لتغطي الخسائر الأقل من قيمة غير معلومة (threshold point) ثم بعد ذلك دالة كثافة الاحتمال لتوزيع باريتو لتغطي قيم الخسائر الأكبر من هذه القيمة (Preda and Ciumara, 2006 & Cooray and Ananda, 2005).

وقد تناولت دراسة (Cooray and Ananda, 2005) النموذج المركب الأول من التوزيع اللوغاريتمي الطبيعي - باريتو، حيث يعتمد هذا النموذج على دالة الكثافة الاحتمالية للتوزيع اللوغاريتمي الطبيعي لتصل إلى قيمة غير معلومة، ثم دالة الكثافة الاحتمالية لتوزيع باريتو ثنائي المعلمت بعد ذلك. واقترحوا أن هذا النموذج قد يكون مناسباً للاستخدام من قبل الخبراء الاكتواريين عندما يواجهون بيانات قيمتها صغيرة وذات تكرار كبير بالإضافة إلى بيانات قيمتها كبيرة وذات تكرار صغير. إن الميزة الرئيسية لهذا النموذج هو أنه يسمح باختيار نقطة القطع (threshold point) الأكثر ملاءمة للبيانات. وفرض (Cooray and Ananda, 2005) شروطاً على معلمت النموذج لضمان الإتصال وقابلية الإشتقاق عند نقطة القطع. وتناولت دراسة (Preda and Ciumara, 2006) مقارنة بين النموذجين المركبين اللوغاريتمي الطبيعي - باريتو الذي تم إنشائه بواسطة (Cooray and Ananda, 2005) وواييل- باريتو الذي تم إنشاؤه من قبل (Ciumara, 2006) بنفس طريقة النموذج اللوغاريتمي الطبيعي - باريتو. وثبت أن هذه النماذج أظهرت سلوكاً مشابهاً ويمكن استخدامها في مجال التأمين لنمذجة بيانات المطالبات، وخاصة في الحالات ذات مبالغ الخسارة الكبيرة، وأظهرت النتائج أن شكل دالة الكثافة الاحتمالية للتوزيع اللوغاريتمي الطبيعي- باريتو، وتوزيع واييل- باريتو المركب لها ذيل أكبر مقارنة بالتوزيع اللوغاريتمي الطبيعي أو واييل بشكل منفصل، ولكن تتسم بذيل أصغر من باريتو، ومع ذلك، كان شكل هذه الكثافات المركبة متشابهاً. وتناولت دراسة (Scollnik, 2007) ثلاثة نماذج مختلفة للنموذج المركب اللوغاريتمي الطبيعي- باريتو الذي يتم استخدامه مع البيانات ذات طبيعة الالتواء الموجب والتي تنشأ غالباً في الدراسات الاكتوارية لصناعة التأمين. واتضح في هذه الدراسة أن النموذج الأول المقترح (اللوغاريتمي الطبيعي- باريتو) من (Cooray and Ananda, 2005) له شكل مقيد للغاية بسبب أوزان الخطل الثابتة الخاصة به. وتم تطوير النموذج المركب الأول اللوغاريتمي الطبيعي- باريتو، ليميز النموذج المركب الثاني بأوزان خطل غير ثابتة على عكس النموذج الأول. وجاء النموذج الثالث مشابهاً للنموذج الثاني، ولكن تم إنشاؤه باستخدام توزيع باريتو من النوع الثاني (باريتو المعمم) بدلاً من توزيع باريتو المعياري في ضوء ملاحظات (McNeil, 1997) وكانت نتائج الدراسة أن النموذج الثالث أفضل لملاءمة للبيانات. وقدمت دراسة (Scollnik, and Sun, 2012) نماذج واييل- باريتو المركبة، واقترحت استخدامها في نمذجة بيانات مطالبات التأمين. وتناولت دراسة (Walena and Sandile, 2024) تحليلاً شاملاً للنماذج المركبة والنماذج المختلطة لبيانات مطالبات التأمين، لأن هذه النماذج قد نالت اهتماماً في الأدبيات الاكتوارية لأنها توفر طرقاً مرنة لتوفيق المنحنيات، حيث أخذت في الاعتبار ٢٥٦ نموذجاً مركباً و٢٥٦ نموذجاً مختلطاً مشتقاً من ١٦ توزيعاً معيارياً شائعاً، وتناولت الدراسة معايير اختيار النموذج ومقاييس الخطر لأفضل ٢٠ نموذجاً في كل فئة (مركب / مزيج)، وتوصلت إلى أن النماذج المركبة

حققت تقديرات أفضل للخطر. وقامت دراسة (Abu Bakar *et al.*, 2015) بتطوير العديد من النماذج المركبة التي تعتمد على توزيع Weibull لبيانات مطالبات التأمين ذات الذيل الثقيل، ويفترض النموذج المركب توزيعات مختلفة لرأس وذيل التوزيع، وتم اختبار مدى ملاءمتها، ثم تناولت إحدى طرق قياس الخطر. وفي دراسة (جيهان المعداوي، ٢٠١٦) تم استخدام بعض التوزيعات الاحتمالية ذات طبيعة الالتواء الموجب التي تستخدم على نطاق واسع في توفيق بيانات مطالبات التأمين، وتم تقدير معالم هذه التوزيعات باستخدام طريقة الإمكان الأعظم، واختيار أفضل توزيع من بين هذه التوزيعات من خلال عدة معايير وهي: (NLL, AIC, BIC). وقدمت دراسة (أمانى عوجة، ٢٠١٧) النمذجة الاكتوارية لبيانات مطالبات التأمين الهندسي باستخدام بعض التوزيعات الإحصائية، وتم تقدير معالم التوزيع المجهولة، ثم إختبار جودة التوفيق باستخدام إختبار كولموجروف سيمرنوف. وفي دراسة (أمانى عوجة، ٢٠١٩) تم استخدام توزيع بواسون ذى الأصفار الزائدة وتوزيع هاردل بواسون في نمذجة تكرار المطالبات في تأمين السيارات. وفي دراسة (هاشم، ٢٠٢٢) تم استخدام أربع خوارزميات في التنقيب في البيانات للوصول إلى النمذجة الاكتوارية للمطالبات في فرع تأمين البترول، وقد توصلت إلى إمكانية تطبيق الخوارزميات الأربعة المقترحة في نمذجة مطالبات تأمين البترول، وأن هناك اختلاف بين تطبيق النماذج المختلفة للخوارزميات المستخدمة في النمذجة الاكتوارية للمطالبات. وفي دراسة (هاشم وآخرون، ٢٠٢٣) تم استخدام النماذج الخطية المعممة وتطبيق خوارزمية الأعشاب الضارة لنمذجة المطالبات في فرع تأمين السيارات الإيجارى. وفي دراسة (العربي، ٢٠٢٤) قدمت هذه الدراسة تسع توزيعات احتمالية مختلفة لتوفيق بيانات مطالبات تأمينات الممتلكات في ثلاثة فروع (تأمين اجسام السفن، التأمين الهندسي، والتأمين الطبى).

باستعراض الدراسات السابقة، وجد أن بعض الدراسات استخدمت التوزيعات الاحتمالية التقليدية، وبعض الدراسات قد استخدمت خوارزميات في التنقيب في البيانات والنماذج الخطية المعممة وتطبيق خوارزمية الأعشاب الضارة لنمذجة بيانات المطالبات. ودراسة واحدة قد استخدمت توزيعات احتمالية مختلطة في نمذجة بيانات المطالبات. وفي هذه الدراسة سوف يتم استخدام النماذج المركبة للتوزيع اللوغاريتمي الطبيعي-باريتو (Composite Lognormal-Pareto) ولتوزيع واييل - باريتو (Composite Weibull-Pareto) لنمذجة بيانات المطالبات في شركة التأمين.

وتتمثل مشكلة البحث في أن التوزيعات الاحتمالية التقليدية المستخدمة في نمذجة بيانات المطالبات التي تتسم بالالتواء الشديد والذيل الثقيل (مثل التوزيع اللوغاريتمي الطبيعي، واييل، جاما، وباريتو) أصبحت أقل استخداماً في الدراسات الاكتوارية بشكل متزايد. ومن ناحية أخرى، فقد اكتسبت النماذج المركبة المزيد من الاهتمام في الأونة الأخيرة بسبب مرونتها. ولذلك سوف يتم استخدام نماذج التوزيعات المركبة (composite models) لنمذجة بيانات المطالبات في شركات التأمين.

ويهدف هذا البحث إلى استخدام النماذج المركبة للتوزيع اللوغاريتمي الطبيعي - باريتو (Composite Lognormal-Pareto) ولتوزيع واييل- باريتو (Composite Weibull-Pareto) في نمذجة بيانات المطالبات في شركات التأمين، ثم المقارنة بين هذه النماذج لاختيار أفضلهم من خلال إيجاد المقاييس الإحصائية (NLL, AIC, BIC)، وبالتالي الوصول إلى أفضل نموذج، ثم محاولة تحسين هذا النموذج للوصول إلى قيمة threshold point المثلى التي تحقق أقل قيمة (AIC, BIC) في النموذج الأفضل.

وترجع أهمية البحث في أنه يتناول موضوع من الموضوعات الهامة لشركات التأمين وهو نمذجة المطالبات، بالإضافة إلى حداثة استخدام نماذج التوزيعات الاحتمالية المركبة من التوزيع اللوغاريتمي الطبيعي - باريتو، وتوزيع واييل- باريتو في نمذجة المطالبات. وأيضاً مساعدة شركات تأمينات الممتلكات والمسئولية المصرية على تحسين أدائها مما يمكنها من إتمام عملية تسعير وثائقها بدقة، وبالتالي تقييم أفساط التأمين، وتقييم وضعها المالي، وتقييم احتمالات التعثر، وما إلى ذلك.

وستعتمد الدراسة التطبيقية في هذا البحث على بيانات مطالبات التأمين في فرع الحريق لإحدى شركات التأمين المصرية.

وسيتم مناقشة نماذج التوزيعات المركبة على النحو التالي:

النماذج المركبة (نماذج التوزيعات الاحتمالية المركبة) (composite models) يقصد بها تقسيم البيانات إلى مجموعتين بناءً على قيمة (θ) (Threshold point)، ثم نستخدم توزيعين مختلفين للتعامل مع كل مجموعة من البيانات بحيث يتم استخدام توزيع للبيانات التي تقع تحت هذه القيمة (θ) واستخدام توزيع آخر للبيانات التي تقع فوق هذه القيمة (θ) . وفي هذه الدراسة سوف يتم استخدام التوزيع اللوغاريتمي الطبيعي أو واييل للبيانات الأقل من (θ) واستخدام توزيع باريتو أو باريتو المعمم للبيانات الأكبر من (θ) .

أولاً: النمذجة الاكتوارية باستخدام نماذج التوزيع اللوغاريتمي الطبيعي- باريتو المركب (Composite Lognormal-Pareto Models):

(Preda and Ciumara, 2006; Scollnik, 2007; Teodorescu and Vernic, 2009; Sun, 2010; Nadarajah, and Abu Bakar, 2014; Stefanut Ciochina, 2019 & Walena and Sandile, 2024)

تم إنشاء النموذج اللوغاريتمي الطبيعي- باريتو المركب بواسطة (Cooray and Ananda, 2005)، بفرض أن X متغير عشوائي، فإن دالة كثافة الاحتمال للنموذج المركب تأخذ الصورة التالية:

$$f(x) = \begin{cases} cf_1(x) & \text{if } 0 < x \leq \theta \\ cf_2(x) & \text{if } \theta \leq x < \infty \end{cases}$$

حيث: C : ثابت، $f_1(x)$: ترمز لدالة كثافة الاحتمال للتوزيع اللوغاريتمي الطبيعي بمعلمتين، وتأخذ الشكل التالي:

$$f_1(x) = \frac{(2\pi)^{-1/2}}{x\sigma} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right], x > 0$$

$f_2(x)$: ترمز لدالة كثافة الاحتمال لتوزيع باريتو بمعلمتين، وتأخذ الشكل التالي:

$$f_2(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}} \quad x > \theta$$

$\theta, \mu, \sigma, \alpha$ معالم غير معلومة، حيث أن:

$$\theta > 0, \quad -\infty < \mu < \infty, \quad \sigma > 0, \quad \alpha > 0$$

ويفترض النموذج الإستمرارية وقابلية الإشتقاق عند θ (Preda and Ciumara, 2006) أي:

$$\int_0^{\infty} f(x)dx = 1, \quad f_1(\theta) = f_2(\theta), \quad f_1'(\theta) = f_2'(\theta)$$

وإستنتاج (Cooray and Ananda, 2005) دالة كثافة الاحتمال للنموذج اللوغاريتمي الطبيعي - باريتو المركب على النحو التالي:

$$f(x) = \begin{cases} \frac{\alpha\theta^\alpha}{(1 + \Phi(k))x^{\alpha+1}} \exp\left\{-\frac{\alpha^2}{2k^2} \ln^2\left(\frac{x}{\theta}\right)\right\} & \text{if } 0 < x \leq \theta \\ \frac{\alpha\theta^\alpha}{(1 + \Phi(k))x^{\alpha+1}} & \text{if } \theta \leq x < \infty \end{cases}$$

حيث أن: $\Phi(\cdot)$ تمثل دالة التوزيع التراكمية للتوزيع الطبيعي المعياري ودالة التوزيع التراكمية للنموذج اللوغاريتمي الطبيعي - باريتو المركب كالتالي:

$$F(x) = \begin{cases} \frac{1}{(1 + \Phi(K))} \Phi\left\{\left(\frac{\alpha}{k}\right) \ln\left(\frac{x}{\theta}\right) + k\right\} & \text{if } 0 < x \leq \theta \\ 1 - \frac{1}{(1 + \Phi(K))} \left(\frac{\theta}{k}\right)^\alpha & \text{if } \theta \leq x < \infty \end{cases}$$

حيث أن: k : ثابت معلوم يتم الحصول على قيمته بالحل الموجب للمعادلة التالية:

$$\exp(-k^2) = 2\pi k^2$$

حيث أن: $k \approx 0.372238898$

(Preda and Ciumara, 2006; Scollnik, 2007 and Sun, 2010)

ويتطلب إنشاء هذا النموذج اللوغاريتمي الطبيعي - باريتو المركب أيضاً أن:

$$\frac{\ln(\theta) - \mu}{\sigma} = \alpha\sigma = k$$

$$c = \frac{1}{(1 + \Phi(K))}$$

وبالتالي، يكون للنموذج في الواقع معلمتين مجهولين فقط وهما: $\theta > 0$, $\alpha > 0$

واقترح (Cooray and Ananda, 2005) أن نموذجهما يكون مناسباً للاستخدام من قبل الإكتواريين عندما يواجهون بيانات ذات قيم صغيرة وتكرراتها عالية بالإضافة إلى بيانات ذات قيم كبيرة وتكرراتها أقل.

(١) النموذج المركب اللوغاريتمي الطبيعي- باريتو (Lognormal-Pareto Composite) الأول:

تأخذ دالة كثافة الاحتمال لهذا النموذج الشكل التالي:

$$f(x) = \begin{cases} \psi \frac{1}{\Phi(k)} f_1(x) & \text{if } 0 < x \leq \theta \\ (1 - \psi) f_2(x) & \text{if } \theta < x \leq \infty \end{cases}$$

حيث أن: $\theta > 0$, $\alpha > 0$

$f_1(x)$: دالة كثافة الاحتمال للتوزيع اللوغاريتمي الطبيعي

$f_2(x)$: دالة كثافة الاحتمال لتوزيع باريتو

$$k \approx 0.37224$$

$$\psi = \frac{\Phi(k)}{1 + \Phi(k)} \approx 0.39215$$

$$1 - \psi = \frac{1}{1 + \Phi(k)} \approx 0.60785$$

$$\Phi(k) = \Phi\left(\frac{\ln(\theta) - \mu}{\sigma}\right)$$

النموذج المركب بوزن ثابت (Ψ) يفسر على أنه نموذج خليط مكون من مكونين بأوزان خلط ثابتة ومعروفة مسبقاً، المكون الأول هو نموذج لوغاريتمي طبيعي يتم قطعه عند القيمة (θ) والمكون الثاني هو نموذج باريتو ثنائي المعلمات يبدأ من القيمة (θ).

$$\theta = \exp(\mu + k\sigma)$$

ويفترض النموذج أن نقطة القطع أو النقطة الفاصلة بين التوزيعين (θ) (Threshold point) هي (٦٤,٥١٤%) للنموذج اللوغاريتمي الطبيعي بدالة كثافة احتمال $f_1(x)$ ، ونلاحظ من شروط الإستمرارية وقابلية الإشتقاق المفروضة على هذا النموذج، أن ($\Psi = 0.39215$)، وهذا يعنى أن ٣٩,٢١٥% من البيانات تتبع النموذج اللوغاريتمي الطبيعي حتى القيمة (θ) والباقي ٦٠,٧٨٥% من البيانات تتبع نموذج باريتو بعد القيمة (θ).

(٢) النموذج المركب اللوغاريتمي الطبيعي- باريتو (Lognormal-Pareto Composite) الثاني:

النموذج التالي هو النموذج المركب اللوغاريتمي الطبيعي - باريتو بنقطة قطع (θ)، ولكن بأوزان خلط غير محددة مسبقاً. وبفرض أن X متغير عشوائى له دالة كثافة الاحتمال التالية:

$$f(x) = \begin{cases} r \frac{1}{\Phi(k)} f_1(x) & \text{if } 0 < x \leq \theta \\ (1-r)f_2(x) & \text{if } \theta < x \leq \infty \end{cases}$$

حيث أن:

$$\Phi(k) = \Phi\left(\frac{\ln(\theta) - \mu}{\sigma}\right), \quad 0 \leq r \leq 1$$

$$\theta > 0, \quad -\infty < \mu < \infty, \quad \sigma > 0, \quad \alpha > 0, \quad \theta, \mu, \sigma, \alpha \text{ معالم مجهولة،}$$

وبفرض إتصال الدالة عند θ أى: $f_1(\theta) = f_2(\theta)$ ، فإن:

$$r = \sqrt{2\pi}\alpha\sigma\Phi(k) \left[\frac{e^{1/2\left(\frac{\ln(\theta) - \mu}{\sigma}\right)^2}}{e^{1/2\left(\frac{\ln(\theta) - \mu}{\sigma}\right)^2} + 1} \right]$$

نلاحظ أن وزن الخلط (r) ليس قيمة ثابتة ومعلومة كما كان وزن الخلط في النموذج المركب الأول، حيث أن قيمة (r) تتراوح بين الصفر والواحد الصحيح، وتعتمد قيمتها على قيم $(\theta, \mu, \sigma, \alpha)$ ويفرض قابلية الإشتقاق عند θ أي: $f_1'(\theta) = f_2'(\theta)$

وبالتالي، فإن: $\frac{\ln(\theta) - \mu}{\sigma} = \alpha\sigma = k$ وبذلك يمكن كتابة (r) على الصورة التالية:

$$r = \sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma) \left[\frac{e^{1/2(\alpha\sigma)^2}}{e^{1/2(\alpha\sigma)^2} + 1} \right]$$

وهذا النموذج له ثلاث معالم مجهولة:

$$(\theta > 0, \quad \sigma > 0, \quad \alpha > 0)$$

ولتخفيض عدد المعالم يفترض تساوى المشتقة الثانية للدالتين كالتالي:

$$\theta = 50, \alpha = 0.5 \text{، وبوضع } \frac{\ln(\theta) - \mu - 1}{\sigma} = \alpha\sigma \text{، فإن: } f_1''(\theta) = f_2''(\theta)$$

$$\text{فإن: } \mu = \ln(\theta) - k\sigma \approx 3.63$$

$$\sigma = \frac{k}{\alpha} = 0.74$$

(معالم النموذج اللوغاريتمي الطبيعي)

(٣) النموذج المركب اللوغاريتمي الطبيعي- باريتو (Lognormal-Pareto Composite): الثالث:

تم تطوير النموذج المركب الثالث للتوزيع اللوغاريتمي الطبيعي- باريتو حيث يتركب من نموذج التوزيع اللوغاريتمي الطبيعي حتى القيمة (θ) ثم توزيع باريتو المعمم (Generalized Pareto Distribution, GPD) بعد القيمة (θ) بدلاً من توزيع باريتو ذو المعلمتين المعياري، ويطلق عليه أيضاً توزيع باريتو من النوع الثاني أو توزيع لوماكس. وتم استخدام باريتو المعمم في هذا النموذج من أجل تطوير نموذج بيانات المطالبات ذات الذيل الثقيل، حيث تأخذ دالة التوزيع لباريتو المعمم الشكل التالي:

$$G(x) = 1 - \left(1 + \frac{x - \theta}{\alpha\beta}\right)^{-\alpha}$$

$$\text{حيث: } \theta > 0, \quad x > \theta, \quad \beta > 0, \quad \alpha > 0, \quad -\infty < \theta < \infty$$

وتكون دالة كثافة الاحتمال لتوزيع باريتو المعمم $g(x)$ على الصورة التالية:

$$g(x) = \frac{\alpha(\alpha\beta)^\alpha}{(\alpha\beta - \theta + x)^{\alpha+1}}, \quad x > \theta$$

$$\text{حيث: } \theta > 0, \quad \alpha > 0, \quad \beta > 0$$

وبفرض أن: $(\lambda = \alpha\beta - \theta)$ ، وبالتالي يمكن كتابة دالة كثافة الاحتمال لتوزيع باريتو

المعمم على الصورة التالية:

$$g(x) = \frac{\alpha(\lambda + \theta)^\alpha}{(\lambda + x)^{\alpha+1}}, \quad x > \theta$$

$$(\theta > 0, \quad \alpha > 0, \quad \lambda > -\theta)$$

وتعتبر الصيغة السابقة لدالة كثافة الاحتمال لتوزيع باريتو المعمم هي الصيغة الشائعة في الدراسات الاكتوارية.

وتكون دالة كثافة الاحتمال لنموذج Composite Lognormal-Pareto الثالث على الصورة التالية:

$$f(x) = \begin{cases} r \frac{1}{\Phi\left(\frac{\ln(\theta) - \mu}{\sigma}\right)} f_1(x) & \text{if } 0 < x \leq \theta \\ (1 - r) g(x) & \text{if } \theta < x \leq \infty \end{cases}$$

نلاحظ أن:

(r) : وزن الخلط وقيمه تتراوح بين الصفر والواحد الصحيح ($0 \leq r \leq 1$)

$f_1(x)$: دالة كثافة الاحتمال للتوزيع اللوغاريتمي الطبيعي

$g(x)$: دالة كثافة الاحتمال لتوزيع باريتو المعمم

المعلمت في $f_1(x)$ ، $g(x)$ هي $(\theta, \mu, \sigma, \alpha, \lambda)$ ، حيث أن:

$$(\theta > 0, \quad -\infty < \mu < \infty, \quad \sigma > 0, \quad \alpha > 0, \quad \lambda > -\theta)$$

وبفرض إتصال الدالة عند θ ، فإن:

$$r = \sqrt{2\pi}\alpha\theta\sigma\Phi(k) \left[\frac{e^{1/2\left(\frac{\ln(\theta)-\mu}{\sigma}\right)^2}}{e^{1/2\left(\frac{\ln(\theta)-\mu}{\sigma}\right)^2} + \lambda + \theta} \right]$$

وبفرض قابلية الإشتقاق عند θ ، أي: $f_1'(\theta) = f_2'(\theta)$

وبالتالي، فإن: $\sigma = \frac{\ln(\theta) - \mu}{\lambda + \theta}$ ، وإذا تم وضع $\lambda = 0$ فإننا نكون بصدد النموذج

المركب الثاني، ولتخفيض عدد المعالم يفترض تساوى المشتقة الثانية للدالتين كالتالي:

$$\frac{\ln(\theta) + \sigma^2 - \mu - 1}{\theta^2 \sigma^2} = \frac{\alpha + 1}{(\lambda + \theta)^2} \quad \text{وبالتالي فإن:} \quad f_1''(\theta) = f_2''(\theta)$$

ونلاحظ في هذا النموذج أنه يستخدم توزيع باريتو المعمم (GPD) بدلاً من توزيع باريتو ذو المعلمتين المعياري، حيث يضيف GPD معلمة إضافية (λ)، مما يسمح له بنمذجة مجموعة أوسع من سلوكيات الذيل. وهذه المرونة الإضافية تجعل النموذج أكثر تكيفاً مع أنواع مختلفة من البيانات ذات الذيل الثقيل. على عكس النموذج الأول، حيث تكون الأوزان (ψ) ثابتة، والنموذج الثاني، حيث يعتمد الوزن r على المعلمت، ولكنه قد لا يعكس كل التباين، وبالتالي يقوم النموذج الثالث بنمذجة سلوك الذيل بدقة باستخدام GPD، مما يجعله أكثر ملائمة للبيانات ذات الذيل الثقيل.

ثانياً: النمذجة الاكتوارية باستخدام النماذج المركبة وايبل-باريتو (Weibull-Pareto Composite):

(Preda and Ciumara, 2006; Teodorescu, and VERNIC, 2009; Sun, 2010; Scollnik, and Sun, 2012; Stefanut Ciochina, 2019 & Walena and Sandile, 2024).

طور (Ciumara, 2006) نموذج Composite Weibull-Pareto للاستخدام مع البيانات الاكتوارية باستخدام نفس أسلوب (Cooray and Ananda, 2005) الذي قدم Composite Lognormal-Pareto، وتم تعريف نموذج Composite Weibull-Pareto وبنائه كالتالي: بفرض أن X متغير عشوائي، فإن دالة كثافة الاحتمال لنموذج Composite Weibull-Pareto تأخذ الصورة التالية:

$$f(x) = \begin{cases} cf_1(x) & \text{if } 0 < x \leq \theta \\ cf_2(x) & \text{if } \theta \leq x < \infty \end{cases}$$

حيث: C : ثابت، $f_1(x)$: ترمز لدالة كثافة الاحتمال لتوزيع وايبل، وتأخذ الشكل التالي:

$$f_1(x) = \left(\frac{\tau}{x}\right) \left(\frac{x}{\phi}\right)^\tau \exp\left[-\left(\frac{x}{\phi}\right)^\tau\right] \quad x > 0$$

$f_2(x)$: ترمز لدالة كثافة الاحتمال لتوزيع باريتو، وتأخذ الشكل التالي:

$$f_2(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}} \quad x > \theta$$

$\theta > 0$, $\alpha > 0$, $\tau > 0$, $\phi > 0$ ، θ , α , τ , ϕ معالم مجهولة،

ولإستكمال نموذج Composite Weibull-Pareto، إقترض (Ciumara, 2006)

الإستمرارية وقابلية الإشتقاق عند (θ) ، بمعنى أن:

$$f_1(\theta) = f_2(\theta) \quad f_1'(\theta) = f_2'(\theta)$$

وقد أوضح كلاً من (Ciumara 2006 and Cooray, 2009) أن دالة كثافة الاحتمال لنموذج (Composite Weibull-Pareto)، تأخذ الشكل التالي:

$$f(x) = \begin{cases} \left(\frac{(k+1)^2\alpha}{(2k+1)x}\right) \left(\frac{x}{\theta}\right)^{ak} \exp\left\{-\left(\frac{k+1}{k}\right) \left(\frac{x}{\theta}\right)^{ak}\right\} & \text{if } 0 < x \leq \theta \\ \left(\frac{k+1}{2k+1}\right) \left(\frac{\alpha}{x}\right) \left(\frac{\theta}{x}\right)^\alpha & \text{if } \theta \leq x < \infty \end{cases}$$

حيث: k : ثابت معلوم يتم الحصول على قيمته بالحل الموجب للمعادلة التالية:

$$\exp\left(1 + \frac{1}{k}\right) = k + 1$$

حيث نجد أن : $k \approx 2.8573348$ ، ويتطلب أيضاً نموذج Composite Weibull-Pareto أن:

$$1 + \frac{\tau}{\alpha} = \exp\left(1 + \frac{\alpha}{\tau}\right) = \exp\left\{\left(\frac{\theta}{\phi}\right)^\tau\right\}$$

حيث أن: $k = \frac{\tau}{\alpha}$

$$c = \frac{k+1}{2k+1} \approx 0.5744638$$

وبالتالي يكون في النموذج معلمتين مجهولتين وهما: (θ, α) ، حيث: $(\theta, \alpha > 0)$

(١) النموذج المركب وايبل-باريتو (Composite Weibull-Pareto) الأول:

تأخذ دالة كثافة الاحتمال لهذا النموذج الشكل التالي:

$$f(x) = \begin{cases} \eta \left(\frac{1+k}{k}\right) f_1(x) & \text{if } 0 < x \leq \theta \\ (1-\eta) f_2(x) & \text{if } \theta \leq x < \infty \end{cases}$$

حيث: $(\theta, \alpha > 0)$ ، $f_2(x), f_1(x)$ دوال كثافة الاحتمال لتوزيع وايبل وتوزيع باريتو.

$$k \approx 2.8573348$$

$$\eta = \frac{k}{2k+1} \approx 0.4255362$$

$$1 - \eta = c = \frac{k+1}{2k+1} \approx 0.5744638$$

وتكون دالة التوزيع لنموذج وايبل طبقاً لدالة كثافة الاحتمال كالتالي:

$$F_1(x) = 1 - \exp\left\{-\left(\frac{x}{\phi}\right)^\tau\right\}, x > 0$$

$$F_1(\theta) = 1 - \exp\left\{-\left(\frac{\theta}{\phi}\right)^\tau\right\} = \frac{k}{k+1}$$

مما سبق يمكن كتابة دالة كثافة الاحتمال للنموذج المركب Weibull-Pareto الأول على

الصورة التالية:

$$f(x) = \begin{cases} \eta \frac{1}{F_1(\theta)} f_1(x) & \text{if } 0 < x \leq \theta \\ (1-\eta) f_2(x) & \text{if } \theta \leq x < \infty \end{cases}$$

ويقترض أن (Threshold point (θ)) هي (٠,٧٤,٠٧٥٪) لنموذج وايبل بدالة كثافة احتمال

$f_1(x)$ ، ونلاحظ أنه يستخدم أوزان ثابتة ($\eta = 0.4255362$) ، وهذا يعني أن ٤٢,٥٥٪ من

البيانات تتبع توزيع وايبل حتى القيمة (θ) ، و ٥٧,٤٥٪ من البيانات تتبع توزيع باريتو بعد القيمة

(θ) .

(٢) النموذج المركب وايبل - باريتو (Composite Weibull-Pareto) الثاني:

وتأخذ دالة كثافة الاحتمال لهذا النموذج الشكل التالي:

$$f(x) = \begin{cases} r \frac{1}{F_1(\theta)} f_1(x) & \text{if } 0 < x \leq \theta \\ (1-r) f_2(x) & \text{if } \theta \leq x < \infty \end{cases}$$

حيث أن: $(0 \leq r \leq 1)$ ، $f_2(x), f_1(x)$ دوال كثافة الاحتمال لتوزيع وايبل وتوزيع باريتو، $\theta > 0$ ، $\alpha > 0$ ، $\tau > 0$ ، $\phi > 0$

وبفرض الإستمرارية عند (θ) ، فإن: $f_1(\theta) = f_2(\theta)$ وبالتالي:

$$r = \frac{\frac{\alpha}{\tau}}{\frac{\left(\frac{\theta}{\phi}\right)^\tau}{\exp\left\{\left(\frac{\theta}{\phi}\right)^\tau\right\} - 1} + \frac{\alpha}{\tau}}$$

وبفرض قابلية الإشتقاق عند (θ) ، $f_1'(\theta) = f_2'(\theta)$

وبالتالي، ينتج عن ذلك:

$$\left(\frac{\theta}{\phi}\right)^\tau = \frac{\alpha}{\tau} + 1$$

ويمكن تبسيط صيغة r السابقة على النحو التالي:

$$r = \frac{\alpha \exp\left(\frac{\alpha}{\tau} + 1\right) - \alpha}{\alpha \exp\left(\frac{\alpha}{\tau} + 1\right) + \tau}$$

$$0 \leq r \leq 1$$

(Preda, and Ciumara, 2006 & Ciumara, 2006)

$$\frac{\tau}{\alpha} = k = 2.857334$$

وبالتالي يكون نموذج Composite Weibull-Pareto الثاني له ثلاثة معالم مجهولة:

$\theta > 0$ ، $\alpha > 0$ ، $\tau > 0$ ، وبايجاد المشتقة الثانية عند (θ) :

وبالتالي، ينتج عن ذلك: $f_1''(\theta) = f_2''(\theta)$

$$\left(\frac{\theta}{\phi}\right)^\tau = \left(\frac{\alpha}{\tau} + 1\right) \left(\frac{1}{1-\tau}\right)$$

(٣) النموذج المركب واييل - باريتو (Composite Weibull-Pareto) الثالث:

وفي هذا النموذج يتم استخدام توزيع باريتو المعمم (GPD) حيث تأخذ دالة كثافة الاحتمال الشكل التالي:

$$g(x) = \frac{\alpha(\alpha\beta)^\alpha}{(\alpha\beta - \theta + x)^{\alpha+1}}, \quad x > \theta$$

حيث: $\theta > 0$ ، $\alpha > 0$ ، $\beta > 0$

ويمكن كتابة الدالة السابقة على الصورة التالية، وبوضع: $(\lambda = \alpha\beta - \theta)$

$$g(x) = \frac{\alpha(\lambda + \theta)^\alpha}{(\lambda + x)^{\alpha+1}}, \quad x > \theta$$

حيث: $\theta > 0$ ، $\alpha > 0$ ، $\lambda > -\theta$

وتأخذ دالة كثافة الاحتمال لهذا النموذج الشكل التالي:

$$f(x) = \begin{cases} r \frac{1}{F_1(\theta)} f_1(x) & \text{if } 0 < x \leq \theta \\ (1-r)g(x) & \text{if } \theta \leq x < \infty \end{cases}$$

حيث أن: $(0 \leq r \leq 1)$ ، $f_1(x)$: دالة كثافة الاحتمال لتوزيع واييل، $g(x)$: دالة كثافة الاحتمال لتوزيع باريتو المعمم. والمعالم التي ظهرت في $f_1(x)$ ، $g(x)$ هي: $\theta, \alpha, \tau, \phi, \lambda$ ، حيث:

$$\theta > 0, \quad \alpha > 0, \quad \tau > 0, \quad \phi > 0, \quad \lambda > -\theta$$

وبفرض إتصال الدالة عند θ أي: $f_1(\theta) = f_2(\theta)$ ، فإن:

$$r = \frac{\frac{\alpha}{\tau}}{\frac{\lambda + \theta}{\theta} \frac{\left(\frac{\theta}{\phi}\right)^\tau}{\exp\left\{\left(\frac{\theta}{\phi}\right)^\tau\right\} - 1} + \frac{\alpha}{\tau}}$$

وبفرض قابلية الإشتقاق عند (θ) ، $f_1'(\theta) = f_2'(\theta)$ ، وبالتالي، ينتج عن ذلك:

$$\left(\frac{\theta}{\phi}\right)^\tau = \frac{\alpha\theta - \lambda}{(\lambda + \theta)^\tau} + 1$$

له أربعة معالم مجهولة: $\lambda > -\theta$ ، $\tau > 0$ ، $\alpha > 0$ ، $\theta > 0$ ، ويمكن تخفيض هذا النموذج

إلى النموذج الثاني بوضع $\lambda = 0$ ، وبإيجاد المشتقة الثانية عند (θ) ، $f_1''(\theta) = f_2''(\theta)$ ، فإن:

$$\left(\frac{\theta}{\phi}\right)^\tau = \frac{\theta^2(\alpha + 1)}{(\lambda + \theta)^2(1 - \tau)\tau} - \frac{1}{\tau}$$

$$\alpha = \frac{\theta^2 - (\tau(\lambda + \theta) + \theta)(1 - \tau)(\lambda + \theta)}{\theta(1 - \tau)(\lambda + \theta) - \theta^2}$$

الدراسة التطبيقية:

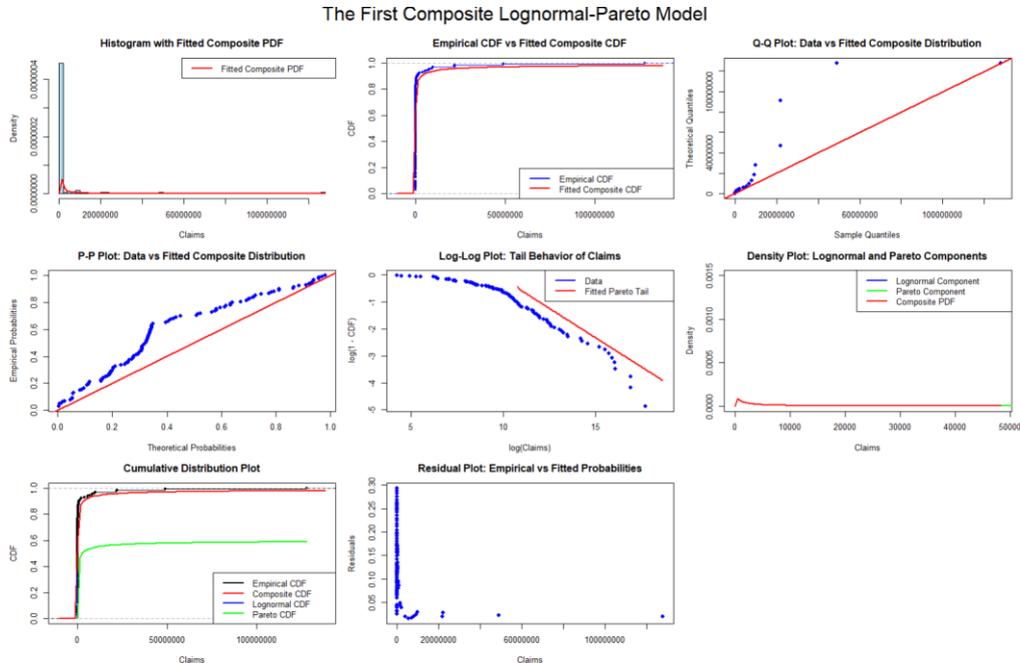
ستعتمد الدراسة التطبيقية في هذا البحث على بيانات مطالبات الحريق لإحدى شركات التأمين المصرية. وتم استخدام برنامج (R) في إجراء الدراسة التطبيقية على النحو التالي: (Core Team, 2023; Davison, 2023; Marie and Christophe, 2015; Scollnik, and Sun, 2012; Dutang *et al.*, 2008 and Scollnik, 2007).

أولاً: الإحصاءات الوصفية لبيانات مطالبات الحريق:

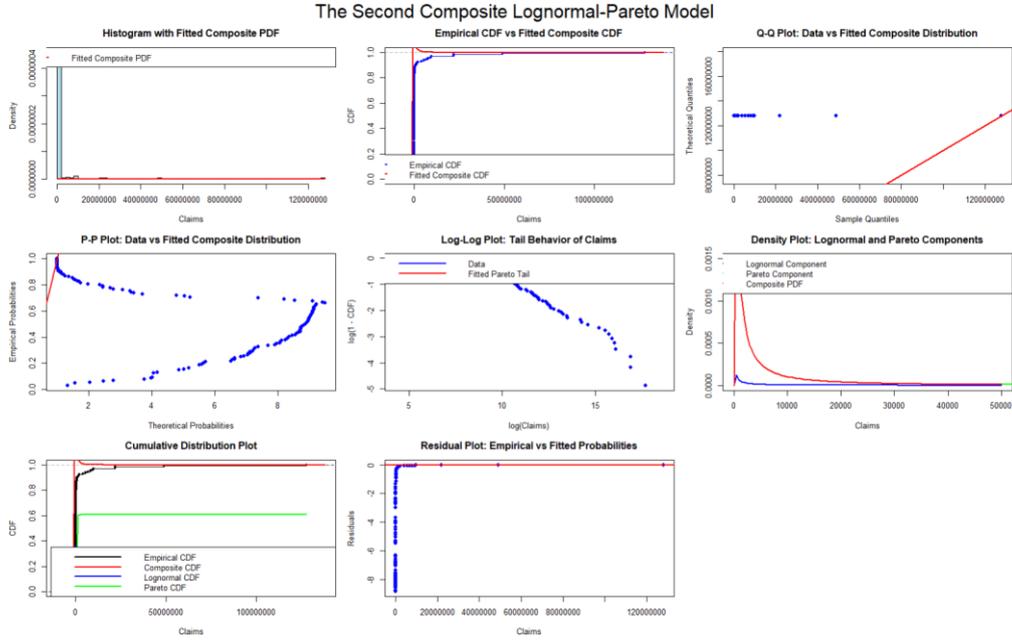
جدول (١): يوضح الإحصاءات الوصفية لبيانات مطالبات الحريق:

Statistic	Value	Percentile	Value
Sample Size	129	Min	70
Range	127.926651×10^6	5%	164
Mean	21.5531×10^5	10%	600
Variance	152.26547×10^{12}	25% (Q1)	3912
Std. Deviation	12.33959×10^6	50% (Median)	26679
Coef. of Variation	5.7252	75% (Q3)	1.15×10^5
Std. Error	1.0864×10^6	90%	8.84831×10^5
Skewness	8.8005	95%	7.600465×10^6
Excess Kurtosis	83.72651	Max	1.27926721×10^8

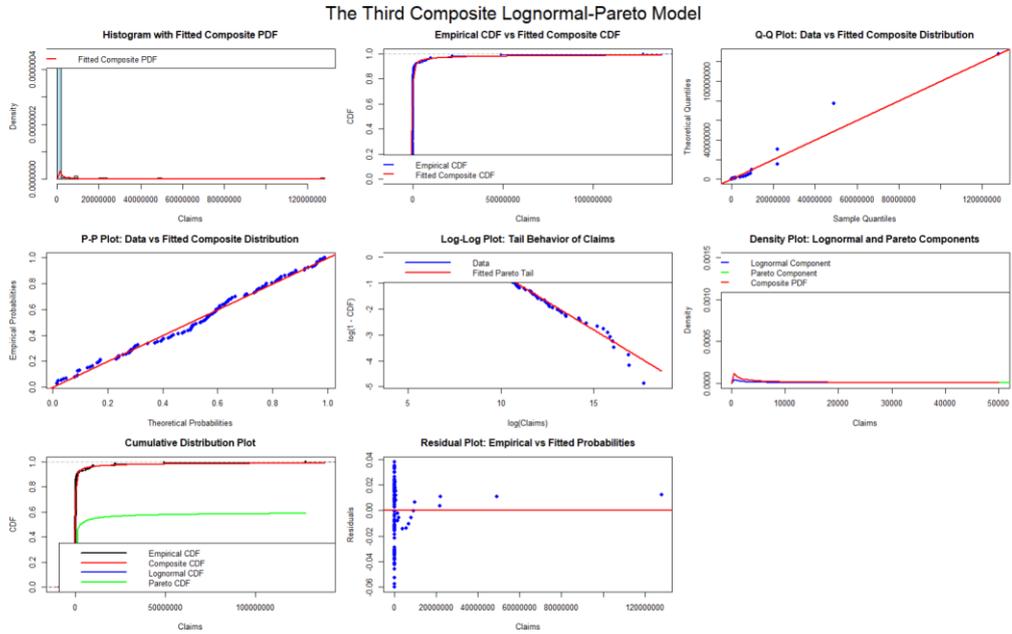
ثانياً: توفيق بيانات المطالبات وفقاً للنماذج المختلفة:



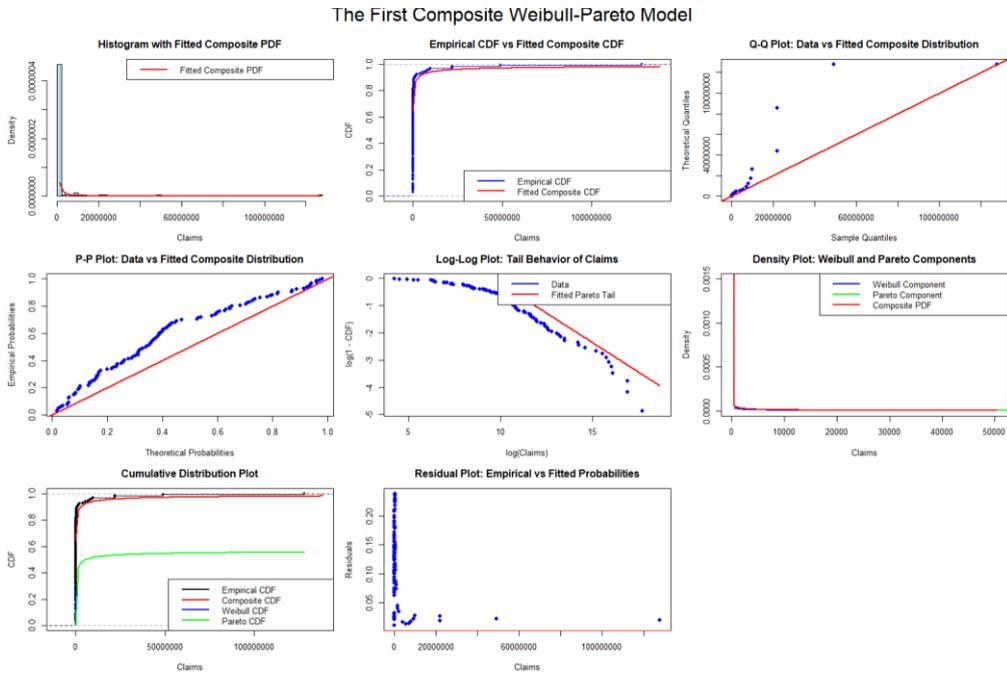
شكل (١) النموذج المركب اللوغاريتمي الطبيعي-باريتو (Composite Lognormal-Pareto) الأول



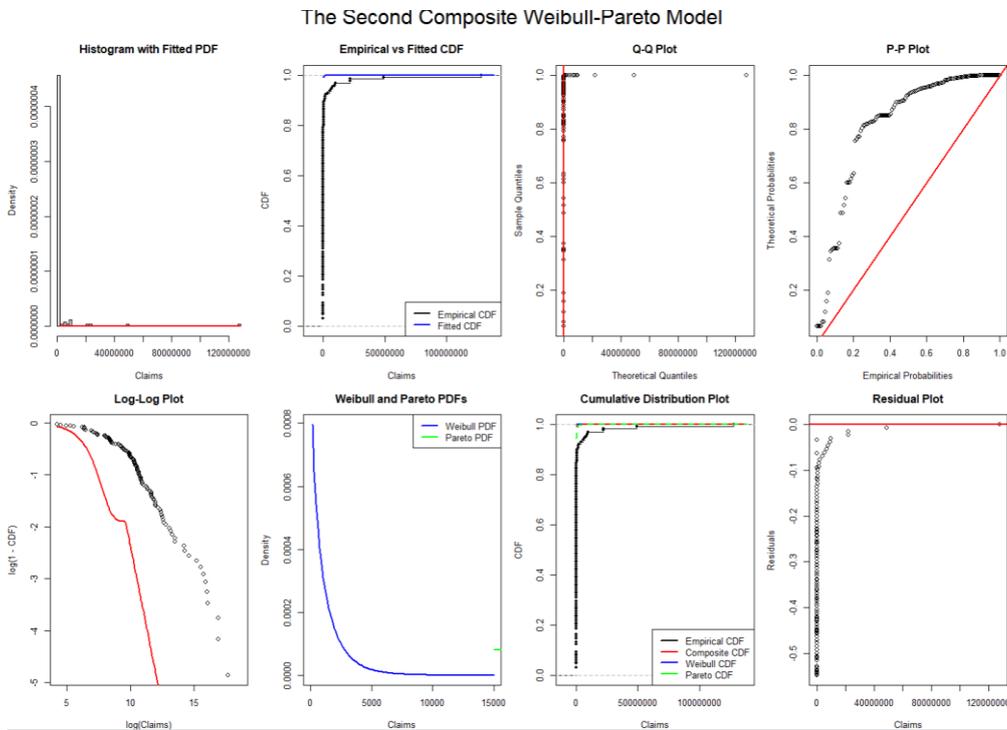
شكل (٢) النموذج المركب اللوغاريتمي الطبيعي- باريتو (Composite Lognormal-Pareto) الثاني



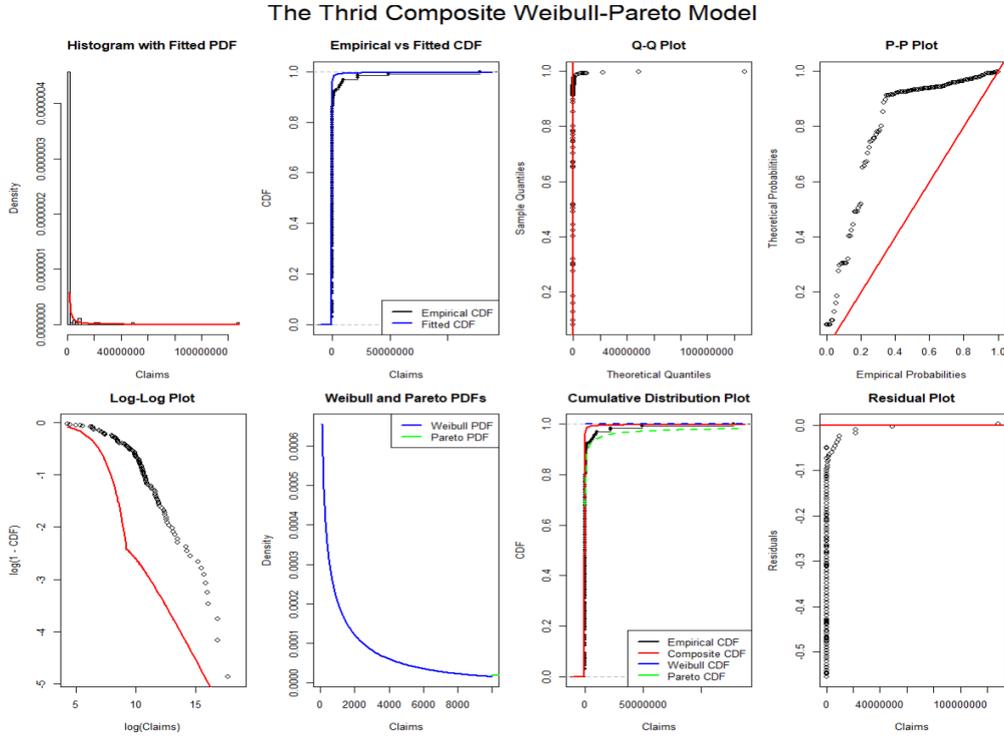
شكل (٣) النموذج المركب اللوغاريتمي الطبيعي- باريتو (Composite Lognormal-Pareto) الثالث



شكل (٤) النموذج المركب وايبل - باريتو (Composite Weibull -Pareto) الأول



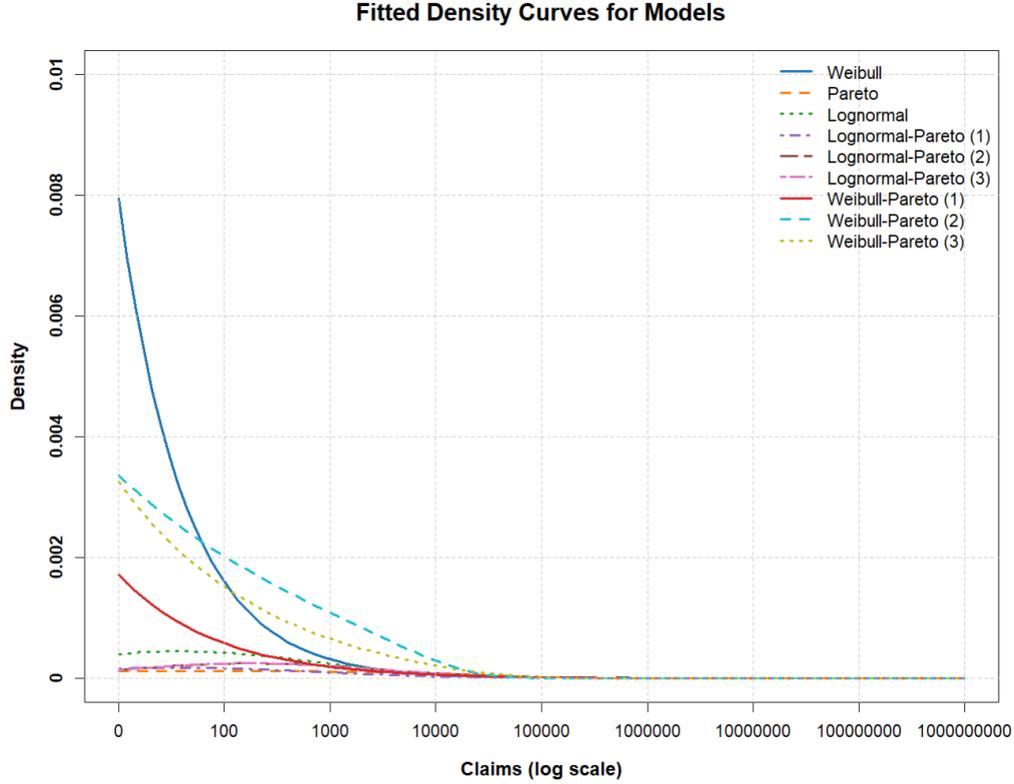
شكل (٥) النموذج المركب وايبل - باريتو (Composite Weibull -Pareto) الثاني



شكل (٦) النموذج المركب وايبل - باريتو (Composite Weibull -Pareto) الثالث

وبصفة عامة، توضح الرسوم البيانية في الأشكال (١، ٢، ٤، ٥، ٦) أن النموذج المركب يمثل الجزء الأكبر من البيانات بشكل جيد ولكنه يعاني من تمثيل الذيل (القيم الكبيرة). ويؤكد ذلك رسم Log-Log أن توزيع باريتو يقلل من تقدير ثقل الذيل، بينما توضح الرسوم البيانية في شكل (٣) أن النموذج المركب الثالث اللوغاريتمي الطبيعي - باريتو يمثل الجزء الأكبر من البيانات والذيل بشكل فعال. ويوضح رسم Log-Log في شكل (٣) أن توزيع باريتو المعمم يمثل سلوك الذيل بدقة، وهذا يؤكد على أهمية استخدامه لنمذجة المطالبات ذات القيم العالية.

يوضح شكل (٧) منحنيات الكثافة لتوزيعات وايبل، باريتو، اللوغاريتمي الطبيعي، والنموذج المركب اللوغاريتمي الطبيعي- باريتو، ووايبل- باريتو أن المنحنيات سلسلة ومحددة جيداً، مما يشير إلى ملاءمة جيدة للنماذج. بينما نماذج باريتو والمركب اللوغاريتمي الطبيعي- باريتو كانت أكثر ملاءمة لنمذجة البيانات ذات الذيل الثقيل.



شكل (٧): يوضح منحنيات دالة الكثافة الاحتمالية لكل من (Weibull, Pareto, Lognormal) ومنحنيات دالة الكثافة الاحتمالية المركبة لكل من (Composite Lognormal-pareto and Composite Weibull-Pareto)

ثالثاً: إختيار النموذج الأفضل:

تم استخدام عدة معايير لاختبار جودة التوفيق لقياس مدى ملائمة النماذج (التوزيعات) المختارة، وهذه المعايير تقيم صلاحية النماذج المختارة، حيث أن أقل قيمة للمعيار تعطي مؤشر لأفضل نموذج، وهذه المعايير (Abu Bakar *et al.*, 2015) تتمثل في الآتي:

١- Negative Log- Likelihood (NLL) :

هذا المعيار يهدف إلى تصغير قيمة دالة الإمكان الأعظم السالبة، ويعرف على النحو التالي:

$$NLL = - \ln(\theta)$$

حيث أن: $\ln(\theta)$: لو غار يتم دالة الإمكان الأعظم للنموذج

ويستخدم معيار NLL لمقارنة التوزيعات، وتشير أقل قيمة لهذا المعيار إلى أنسب توزيع لتوفيق البيانات.

٢- Akaike Information Criterion (AIC) :

هذا المعيار يقيس الجودة النسبية للنموذج (Akaike, 1974) ويعرف كالتالي:

$$AIC = 2k - 2 \ln(\theta)$$

حيث أن (k): تمثل عدد معالم التوزيع، ويعتبر أفضل توزيع هو ذات أقل قيمة (AIC)

٣- Bayesian information criterion (BIC):

ويطلق على هذا المعيار أيضًا Schwarz information criterion (Schwarz, 1978) ويحسب هذا المعيار من المعادلة التالية:

$$BIC = k \ln(n) - 2 \ln(\theta)$$

حيث أن (n) : تمثل عدد المشاهدات (حجم العينة)، وهذا المعيار يشابه المعيارين السابقين، حيث أن أقل قيمة (BIC) تشير لأفضل توزيع من بين التوزيعات المقارن بينها.

وتم تقدير المعالم وحساب المقاييس الإحصائية (NLL, AIC, BIC) في جدول (٢).

جدول (٢): يوضح القيم المقدرة للنماذج المستخدمة لنمذجة بيانات مطالبات الحريق في شركة التأمين والمقاييس الإحصائية (NLL, AIC, BIC).

Distribution	Parameter MLEs	NLL	AIC	BIC
Weibull	$\hat{\tau}=0.4081122$ $\hat{\phi}=90560.18$	3044.181	6092.362	6099.34
Pareto	$\hat{\alpha} = 0.5551576$ $\hat{\theta}=9747.191$	3182.153	6090.141	6097.119
Lognormal	$\hat{\sigma}=2.534954$ $\hat{\mu}=10.1778$	3031.514	6067.029	6074.007
نموذج Composite Lognormal-Pareto الأول	$\hat{\theta}=48280.19$ $\hat{\alpha} = 0.4337139$ $\hat{\mu}=8.436369$ $\hat{\sigma}=1.897962$	1657.86	3323.72	3335.159
نموذج Composite Lognormal-Pareto الثاني	$\hat{\theta}=50000$ $\hat{\alpha} = 1.59643$ $\hat{\sigma}=2.441346$	1749.932	3505.864	3514.444
نموذج Composite Lognormal-Pareto الثالث	$\hat{\theta}=50000$ $\hat{\alpha} = 0.4333273$ $\hat{\mu}=9.919873$ $\hat{\sigma}=2.647628$ $\hat{\lambda}=1000$	1623.519	3257.038	3271.337
نموذج Composite Weibull -Pareto الأول	$\hat{\theta}=50470.07$ $\hat{\alpha} = 0.4324715$ $\hat{\tau}=0.5284934$ $\hat{\phi}=50472.23$	1637.058	3282.116	3293.556
نموذج Composite Weibull -Pareto الثاني	$\hat{\theta}= 10000$ $\hat{\tau} = 0.3081401$ $\hat{\alpha} = 0.3081401$ $\hat{\phi}=1014.443$	1640.582	3289.164	3300.603
نموذج Composite Weibull -Pareto الثالث	$\hat{\phi}=2694.899$ $\hat{\alpha} = 0.4156669$ $\hat{\tau} = 0.6743488$ $\hat{\lambda} = 20879.84$	1627.275	3262.551	3273.99

وبناءً على المقاييس الإحصائية (NLL, AIC, BIC) في جدول (٢) يتضح مايلي:

- تفوق النماذج المركبة على النماذج التقليدية (Lognormal, Pareto, Weibull) من حيث (NLL, AIC, BIC).
- النموذج Composite Lognormal-Pareto الثالث يناسب البيانات بشكل أفضل من Composite Lognormal-Pareto الأول والثاني لأنه يحقق أقل قيم (NLL, AIC, BIC).
- النموذج Composite Weibull-Pareto الثالث يناسب البيانات بشكل أفضل من Composite Weibull-Pareto الأول والثاني لأنه يحقق أقل قيم (NLL, AIC, BIC).
- النموذج Composite Lognormal-Pareto الثالث هو الأفضل ملائمة للبيانات لأنه يحقق أقل قيم (NLL, AIC, BIC).
- نماذج Composite Weibull-Pareto تؤدي أداءً جيداً ولكنها ليست قوية مثل نموذج Composite Lognormal-Pareto.

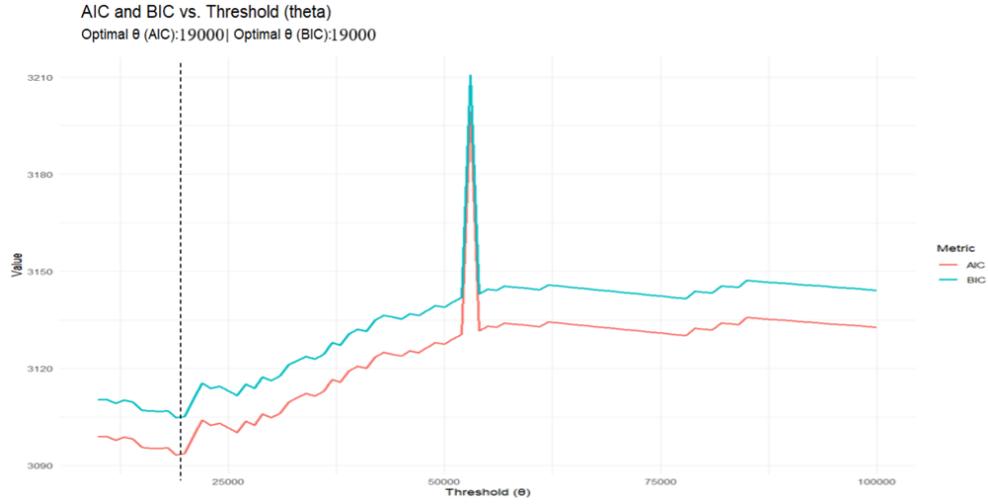
رابعاً: تحسين النموذج المركب Composite Lognormal-Pareto:

لا يمكن تحسين (θ) في نموذج Composite Lognormal-Pareto الأول حيث يتم اشتقاقها من معلمات التوزيع اللوغاريتمي الطبيعي (σ, μ)، ولكن يمكن تحسين (θ) في نموذج Composite Lognormal-Pareto الثاني والثالث حيث أنها يتم تقديرها من البيانات، ولكن نموذج Composite Lognormal-Pareto الثالث هو النموذج الأفضل، لأنه يستخدم توزيع باريتو المعمم (GPD)، حيث يكون أكثر ملائمة لبيانات المطالبات الكبيرة، وهذا يتوافق مع طبيعة بيانات التأمين التي تحتوى على مطالبات كبيرة جداً، لذلك تم تحسين (θ) في نموذج Composite Lognormal-Pareto الثالث، حيث كانت:

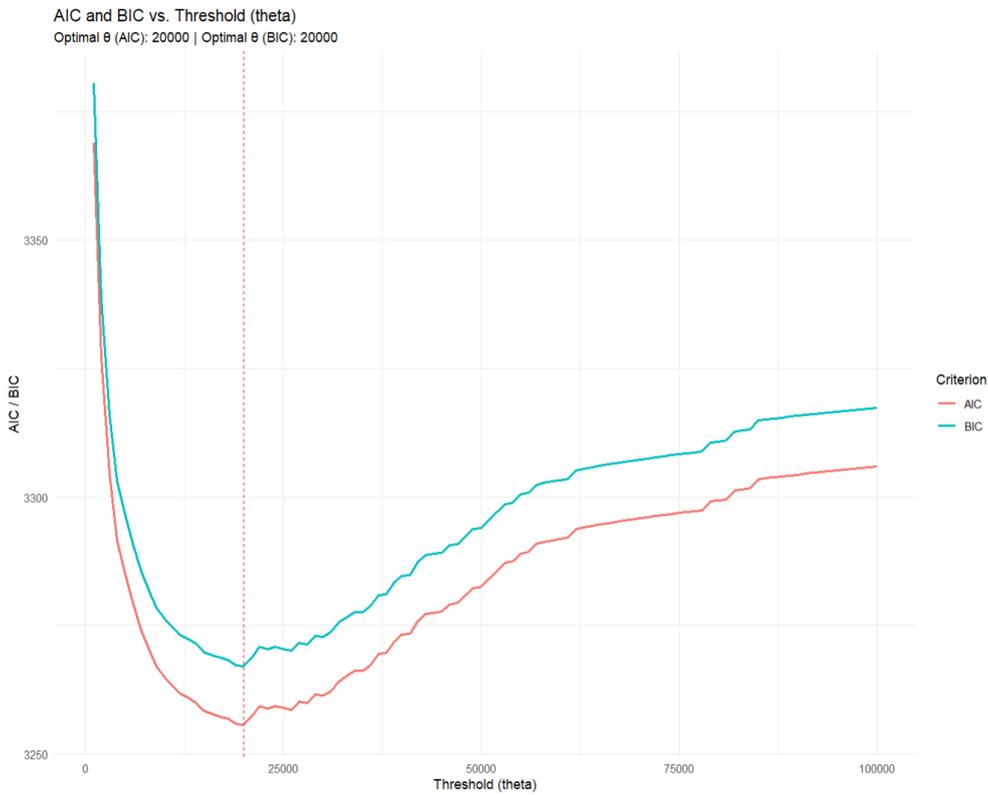
- القيمة المثلى $\theta = 19000$ ، وهذا يعني أن النموذج يحقق أفضل أداء عندما يتم تحديد Threshold point $(\theta) = 19000$ ، بمعنى أن المطالبات الأقل من 19000 سيتم نمذجتها باستخدام التوزيع اللوغاريتمي الطبيعي، المطالبات الأكبر من 19000 سيتم نمذجتها باستخدام توزيع باريتو المعمم.
- عند $(\theta) = 19000$ ، كانت قيم AIC, BIC عند أدنى مستوى (AIC = 3093.282, BIC = 3104.721)، وهذا يشير إلى جودة ملائمة النموذج، وإتفاق معيار AIC, BIC على نفس القيمة المثلى يعكس إستقرار النموذج وقوته الإحصائية، وذلك موضح في شكل (٨).

خامساً: تحسين النموذج المركب Composite Weibull-Pareto:

- يمكن عمل التحسين للنماذج الثلاثة ولكن تم تحسين (θ) في نموذج Composite Weibull-Pareto الثالث، لأنه يستخدم توزيع باريتو المعمم (GPD)، حيث يكون أكثر ملائمة لبيانات المطالبات الكبيرة، وهذا يتوافق مع طبيعة بيانات التأمين التي تحتوى على مطالبات كبيرة جداً، حيث كانت:
- القيمة المثلى $\theta = 20000$ ، وهذا يعني أن النموذج يحقق أفضل أداء عندما يتم تحديد Threshold point $(\theta) = 20000$.
- عند $(\theta) = 20000$ ، كانت قيم AIC و BIC عند أدنى مستوى (AIC = 3255.723, BIC = 3267.162)، مما يشير إلى جودة ملائمة النموذج، وإتفاق معيار AIC, BIC على نفس القيمة المثلى يعكس إستقرار النموذج وقوته الإحصائية، وذلك موضح في شكل (٩).



شكل (٨): يوضح AIC, BIC عند القيمة المثلى $\theta = 19000$ لنموذج Composite Lognormal- Pareto الثالث



شكل (٩): يوضح AIC, BIC عند القيمة المثلى $\theta = 20000$ لنموذج Composite Weibull- Pareto الثالث.

النتائج:

- (١) تفوق النماذج المركبة على النماذج التقليدية Lognormal, Pareto, Weibull من حيث NLL, AIC, BIC.
- (٢) النموذج المركب Composite Lognormal-Pareto الثالث هو الأفضل لملاءمة للبيانات لأنه يحقق أقل قيم NLL, AIC, BIC.
- (٣) نماذج Composite Weibull-Pareto تؤدي أداءً جيدًا ولكنها ليست قوية مثل نماذج Composite Lognormal-Pareto
- (٤) عند تحسين (θ) في النموذج المركب Composite Lognormal-Pareto الثالث كانت القيمة المثلى $(\theta) = 19000$ ، وعندها كانت قيم AIC, BIC عند أدنى مستوى ($AIC = 3104.721$, $BIC = 3093.282$)، وهذا يشير إلى جودة ملاءمة النموذج، وإتفاق معيار AIC, BIC على نفس القيمة المثلى يعكس استقرار النموذج وقوته الإحصائية.
- (٥) عند تحسين (θ) في النموذج المركب Composite Weibull-Pareto الثالث كانت القيمة المثلى $(\theta) = 20000$ ، وعندها كانت قيم AIC, BIC عند أدنى مستوى ($AIC = 3267.162$, $BIC = 3255.723$)، وهذا يشير إلى جودة ملاءمة النموذج، وإتفاق معيار AIC, BIC على نفس القيمة المثلى يعكس استقرار النموذج وقوته الإحصائية.

التوصيات:

توصى الدراسة باستخدام النموذج المركب Composite Lognormal-Pareto الثالث لأن النموذج الأفضل لنمذجة بيانات مطالبات التأمين، وأهمية استخدام توزيع باريتو المعمم، لنمذجة مطالبات التأمين ذات الذيل الثقيل بدقة. وأيضًا توصى باستخدام نماذج مركبة أخرى في نمذجة مطالبات التأمين، وبالتالي تعزيز فهم شركات التأمين لسلوك المطالبات وتحسين قراراتها المتعلقة بالتسعير وإدارة المخاطر المالية. كما توصى الدراسة بالجمع بين النماذج المركبة وتقنيات التنقيب في البيانات (Data Mining) لتحليل بيانات المطالبات بشكل أعمق واستخلاص نتائج أكثر دقة.

المراجع:

أولاً: المراجع العربية:

- ١- عوجة، أمانى محمد (٢٠١٩): استخدام توزيع بواسون ذى الأصفار الزائدة وتوزيع هاردل بواسون في نمذجة تكرار المطالبات في تأمين السيارات، المجلة المصرية للدراسات التجارية، كلية التجارة، جامعة المنصورة، المجلد ٤٣، العدد ٤.
- ٢- عوجة، أمانى محمد، نها عبد اللطيف عبد الحميد (٢٠١٧): النمذجة الاكتوارية لمطالبات التأمين الهندسى باستخدام بعض التوزيعات الاحتمالية ذات الذيل الثقيل، المجلة المصرية للدراسات التجارية، كلية التجارة، جامعة المنصورة، المجلد ٤١، العدد ١.
- ٣- العربى، عبد الرحمن (٢٠٢٤): تطوير نموذج احتمالي لبيانات المطالبات كأساس للعمليات الاكتوارية في تأمينات الممتلكات (دراسة تطبيقية)، رسالة دكتوراة (غير منشورة)، جامعة المنصورة

- ٤- المعداوي، جيهان مسعد (٢٠١٦): استخدام التوزيعات الاحتمالية في توفيق بيانات المطالبات بالتطبيق على فرع تأمين السيارات التكميلي، المجلة المصرية للدراسات التجارية – كلية التجارة - جامعة المنصورة - المجلد الأربعون – العدد الرابع - الجزء الأول، ص ١١٩-١٤٢.
- ٥- هاشم، محمد محمود (٢٠٢٢): النمذجة الإكتوارية للمطالبات في التأمينات العامة باستخدام تقنيات التنقيب في البيانات (Data Mining) دراسة تطبيقية على فرع تأمين البترول بالسوق المصري، مجلة الدراسات المالية والتجارية، كلية التجارة، جامعة بنى سويف، العدد الثاني، ص ٣٠٧-٣٦٨.
- ٦- هاشم، محمد محمود، محمد السيد حافظ، محمد أحمد عبد النبي (٢٠٢٣): استخدام النماذج الخطية المعممة (خوارزمية أمثلة الأعشاب الضارة (IWO)) في نمذجة المطالبات في التأمينات العامة بالتطبيق على فرع تأمين السيارات الإجباري، مجلة البحوث الإدارية والمالية والكمية، كلية التجارة، جامعة السويس، المجلد الثالث – العدد الرابع، ص ١٧٦-٢١٣.

ثانياً: المراجع الأجنبية:

- 1) Cooray, K., Ananda, M.M.A. (2005): Modeling actuarial data with a composite lognormal-Pareto model, Scandinavian Actuarial Journal, No. 5/2005, 321-334.
- 2) Ciumara, R. (2006): An Actuarial Model Based on the Composite Weibull-Pareto Distribution, Mathematical Report–Bucharest 8(4):401–414.
- 3) Cooray, K. (2009): The Weibull-Pareto Composite Family with Applications to the Analysis of Unimodal Failure Rate Data, Communications in Statistics–Theory and Methods 38: 1901–1915.
- 4) Preda, V., and Ciumara R. (2006): On Composite Models: Weibull-Pareto and Lognormal-Pareto. A Comparative Study, Romanian Journal of Economic Forecasting 7(2): 32–46.
- 5) Scollnik, D.P.M. (2007): On composite lognormal-Pareto models, Scandinavian Actuarial Journal, 1, 20 – 33
- 6) McNeil, A.J. (1997): Estimating the tails of loss severity distributions using extreme value theory, ASTIN Bulletin 27 (1), 117 –137.
- 7) Sun, C. (2010): Modeling with the Composite lognormal-Pareto models and the Composite Weibull-Pareto models, Master dissertation.
- 8) Stefanut Ciochina (2019): Composite models used in actuarial practice, Review of the Air Force Academy No.2 (40).
- 9) Teodorescu, S. and Vernic R. (2009): Some Composite Exponential-Pareto Models for Actuarial Prediction, Romanian Journal of Economic Forecasting – 4/2009.

-
-
- 10) Abu Bakar, S.A., Hamzah, N.A., Maghsoudi, M., Nadarajah, S. (2015): Modeling loss data using composite models, Insurance: Mathematics and Economics 61 (2015) 146–154.
 - a. Nadarajah, S., Abu Bakar, S.A. (2014): New composite models for the Danish fire insurance data, Scandinavian Actuarial Journal, 2014:2, 180-187, DOI:10.1080/03461238.2012.695748. <https://doi.org/10.1080/03461238.2012.695748>.
 - 11) Modeling loss data using composite models, Insurance: Mathematics and Economics 61 (2015) 146–154.
 - 12) Walena Anesu Maram bakuyana and Sandile Charles Shongwe (2024): Composite and Mixture Distributions for Heavy-Tailed Data - An Application to Insurance Claims, Mathematics 2024, 12, 335. <https://doi.org/10.3390/math12020335>
 - 13) Akaike, H. (1974): A new look at the statistical model identification, IEEE Trans. Autom. Control, 19, 716–723.
 - 14) Schwarz, G. (1978): Estimating the Dimension of a Model, Ann. Stat., 6, 461–464.
 - 15) R Core Team (2023): R: A Language and Environment for Statistical Computing, Available online: <https://www.R-project.org>.
 - 16) Davison, A. (2019): SMPracticals: Practical for Use with Davison (2003) Statistical Models. 2019. R Package Version 1.4-3. (accessed on 1 May 2023). Available online: <https://CRAN.R-project.org/package=SMPracticals>.
 - 17) Dutang, C., Goulet, V., Pigeon, M. (2008): actuar: an r package for actuarial science. J. Stat. Softw. 25 (7), 38. URL <http://www.jstatsoft.org/v25/i07>.
 - 18) Marie Laure Delignette-Muller, Christophe Dutang (2015): fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software, 64(4), 1-34. DOI 10.18637/jss.v064.i04.
 - 19) Scollnik, D.P.M., and Sun, C. (2012): Modeling with Weibull-Pareto models, North American Actuarial Journal, Volume 16, Number 2, pp 260–272.

Applying Composite Distributions in Modeling General Insurance Claim's data

Abstract:

This research aims to use the composite lognormal-Pareto and the composite Weibull-Pareto models in modeling claims data in insurance companies, then compare these models to choose the best of them by finding the statistical measures (NLL, AIC, BIC) and thus reach the best model, then try to improve this model to reach the optimal threshold point value that achieves the lowest value (AIC, BIC) in the best model. The most important results of the study were that the third Composite Lognormal-Pareto model is the best fit for the data because it achieves the lowest values for (NLL, AIC, BIC). The Composite Weibull-Pareto models perform well but are not as strong as the Composite Lognormal-Pareto model. When improving the third model (Composite Lognormal-Pareto), the optimal value (θ) = 19000, and it was (AIC = 3093.282, BIC = 3104.721), while when improving the third model (Composite Weibull-Pareto), the optimal value (θ) = 20000, and it was (AIC = 3255.723, BIC = 3267.162), and this indicates the goodness of the model's fit, and the agreement of the AIC, BIC criterion at the same optimal value reflects model stability and statistical strength. The study recommended using the third Composite Lognormal-Pareto model because it is the best model for modeling insurance claims data, and the importance of using the Generalized Pareto distribution in accurately modeling heavy-tailed insurance claims. Furthermore, using other composite models can enhance the accuracy of modeling insurance claims. The study further recommends integrating composite models with advanced data mining techniques, such as machine learning algorithms, to enable a more in-depth analysis of claims data.

Keywords: Composite models - Composite Lognormal-Pareto - Generalized Pareto distribution - Composite Weibull-Pareto - AIC - BIC