استخدام توزيع بواسون ذى الأصفار الزائدة وتوزيع هاردل بواسون فى نمذجة تكرار المطالبات فى تأمين السيارات

د. أمانى محمد عجوة
مدرس التأمين
بالجامعة العمالية – فرع القاهرة

الملخص

يهدف هذا البحث إلى إلقاء الضوء على استخدام توزيع بواسون ذى الأصفار الزائدة وتوزيع هاردل بواسون لتحسين جودة توفيق البيانات التى تحتوى على أصفار زائدة. وقد تم استخدام بيانات تكرار المطالبات لقاعدة بيانات تأمين السيارات بسنغافورة والمتاحة على الشبكة الدولية للمعلومات. وتم اختبار البيانات لاكتشاف الأصفار الزائدة وتقدير معلمات توزيع بواسون ذى الأصفار الزائدة بطريقة العزوم وطريقة الامكان الأعظم، وتم اشتقاق صيغ الوسط الحسابى والتباين لتوزيع هاردل بواسون واشتقاق مقدرات طريقة العزوم وطريقة الامكان الأعظم للمعلمات المجهولة، وتقدير المعلمات المجهولة لتوزيع هاردل بواسون بطريقة العزوم وطريقة الامكان الأعظم. وتم نمذجة البيانات باستخدام توزيع بواسون، وتوزيع بواسون ذى الأصفار الزائدة وتوزيع هاردل بواسون، واختبارجودة التوفيق باستخدام اختبار كاى تربيع، وتم الاختيار بين التوزيعات المختلفة المستخدمة فى الدراسة باستخدام معيارى (AIC)، (BIC)، وقد وجد أن توزيع هاردل بواسون هو الأفضل لتمثيل البيانات.

Zero-inflated Poisson Distribution and Hurdle Poisson Distribution for modeling Claim Frequency in Motor Insurance

Abstract

The aim of this paper is to highlight the use of Zero inflated Poisson distribution and Hurdle Poisson distribution to improve goodness of fit of data with excess zeros. Claim frequency data were used for the Singapore motor insurance database available on the internet. The data are tested for the detection of excess zeros. Parameters of Zero inflated Poisson distribution are estimated using method of moments and method of maximum likelihood. Mean and variance formulas are derived for the distribution of Hurdle Poisson and estimators of method of moment and maximum likelihood of unknown parameters are derived. Parameters of Hurdle Poisson are estimated. The data are modeled using Poisson distribution, Zero inflated Poisson distribution and Hurdle Poisson distribution. Goodness of fit are tested using Chi Square test. The best distribution in the study is selected using AIC & BIC criteria. Hurdle Poisson is the best distribution for modeling the data.

١. مقدمة

تعد الخطوة الأولى فى تسعير التأمينات العامة هى اختيار التوزيعات التى تلائم بيانات المطالبات. فيجب توخى الدقة عند اختيار توزيع تكرار المطالبات وتوزيع قيم المطالبات، نظراً لأنه عند اختيار توزيع تكرار المطالبات بشكل خاطئ فإن هذا يؤدى إلى حدوث خطأ فى التقدير مما يؤدى فى النهاية إلى سياسة تسعير خاطئة. لذلك يجب اختيار التوزيع الملائم لتكرار المطالبات حتى يتم التنبؤ الصحيح بعدد المطالبات والتى تمثل منتصف الطريق للحصول على سعر عادل. كذلك يجب اختيار التوزيع الملائم لقيم المطالبات.

توزيع بواسون هو من التوزيعات الشائعة الاستخدام فى نمذجة مطالبات تأمين السيارات، ومع ذلك فى بعض الأحيان نلاحظ عدم تساوى المتوسط المحسوب من بيانات تكرار المطالبات مع التباين وهو الشرط الأساسى لاستخدام توزيع بواسون، وقد يكون توزيع بواسون هو التوزيع الملائم لها إلا أنه بسبب زيادة عدد الأصفار الموجودة فى بيانات تكرار المطالبات عن تلك المتوقعة فى توزيع بواسون فإن هذا يؤدى إلى زيادة التباين عن المتوسط، بذلك تصبح البيانات بها تشتت أكثر من اللازم (Over-dispersion). لذلك يصعب استخدام توزيع بواسون المعتاد وفى هذه الحالة يجب البحث عن توزيعات تستوعب وجود الأصفار الزائدة مثل توزيع بواسون ذى الأصفار الزائدة (Zero inflated Poisson distribution)، وتوزيع هاردل بواسون (Hurdle Poisson distribution)

(١-١) مشكلة البحث

فى بعض الأحيان لا يتم الابلاغ عن بعض المطالبات فى تأمين السيارات بسبب وجود حد تحمل، أو رغبة حامل الوثيقة فى أن يحصل على خصم عدم المطالبة، أو رغبته فى ألا يزيد القسط فى العام القادم، كل هذا يؤدى فى النهاية إلى أن يصبح عدد الأصفار الموجودة فى التوزيع التجريبى أكبر من تلك المتوقع وجودها فى التوزيع الاحتمالى النظرى. وفى بعض الأحيان توجد أصفار زائدة بسبب تحسن خبرة المؤمن لهم فى هذه المحفظة فلا يوجد بالفعل حوادث لعدد كبير من حملة الوثائق، فتظهر أيضا مشكلة الأصفار الزائدة. وعند استخدام توزيع بواسون العادى فيجب أن يتساوى المتوسط والتباين فى توزيع بواسون، ومع ذلك قد نلاحظ وجود تشتت أكثر من اللازم فى البيانات (Over-dispersion) من خلال زيادة التباين عن المتوسط؛ بذلك فإن أهم شرط لتطبيق توزيع بواسون أصبح غير موجود. لذلك يجب البحث عن توزيع آخر يأخذ فى اعتباره ظاهرة الأصفار الزائدة فى التوزيع التجريبى.

(١-٢) هدف البحث

يهدف هذا البحث إلى إلقاء الضوء على توزيع بواسون ذى الأصفار الزائدة، وتوزيع هاردل بواسون لتحسين جودة التوفيق للبيانات التى بها أصفار زائدة، واستخدام توزيع بواسون ذى الأصفار الزائدة وتوزيع هاردل بواسون فى نمذجة تكرار المطالبات فى تأمين السيارات. ويتحقق هذا الهدف من خلال القيام بالآتى:

- اختبار البيانات لاكتشاف الأصفار الزائدة.
- تقدير معلمات توزيع بواسون ذى الأصفار الزائدة باستخدام طريقة العزوم وطريقة الامكان الأعظم.
- تقدير معلمات توزيع هاردل بواسون باستخدام طريقة العزوم وطريقة الامكان الأعظم
- اختبار جودة توفيق البيانات باستخدام اختبار Chi Square.
- اختيار أفضل نموذج للبيانات من خلال استخدام معيارى (AIC), (BIC) .

(١-٣) أهمية البحث

ترجع أهمية البحث إلى كون نمذجة تكرار المطالبات فى التأمينات العامة هو الخطوة الأولى فى تحديد المتوسط العام للخسارة الذى يترتب عليه اتخاذ قرارات أخرى مثل التسعير وتكوين الاحتياطيات فى التأمينات العامة. فعندما تتم نمذجة تكرار المطالبات بشكل خاطئ فإن هذا يؤدى إلى استخدام نموذج لا يمثل البيانات وبالتالى يعطى نتائج مضللة.

بذلك يجب أن تتم عملية نمذجة تكرار المطالبات بشكل صحيح حتى نصل فى النهاية إلى تحديد السعر العادل والكافى والقادر على المنافسة حتى يمكن لشركة التأمين الاستمرار فى سوق التأمين.

(١-٤) الدراسات السابقة

دراسة (Yip et al., 2005) قدمت هذه الدراسة نمذجة لبيانات تكرار المطالبات فى التأمينات العامة التى تحتوى على أصفار زائدة وقدم العديد من التوزيعات ذات الأصفار الزائدة فى اطار استخدام نماذج الانحدار التى تفترض تبعية المتغير التابع لأحد هذه التوزيعات.

دراسة (Moualassim et al., 2012) قامت هذه الدراسة بعمل Review لنموذج بواسون العادى ونموذج بواسون ذى الأصفار الزائدة، وتم المقارنة بين النموذجين من خلال نمذجة بيانات تكرار المطالبات فى التأمين الصحى الخاص.

دراسة (Wolny-Dominiak, 2013) قدمت هذه الدراسة نمذجة لتكرار المطالبات باستخدام النماذج ذات الأصفار الزائدة مع استخدام اختبار Score Test لاختبار الأصفار الزائدة للمتغير التابع.

دراسة (Sarul et al., 2015) قدمت هذه الدراسة نمذجة لمطالبات تأمين السيارات فى السوق التركى باستخدام نموذج انحدار بواسون ذى الأصفار الزائدة، ونموذج انحدار ذى الحدين السالب ذى الأصفار الزائدة ، ونموذج انحدار هاردل بواسون، ونموذج انحدار هاردل ذى الحدين السالب.

دراسة (Guillen et al., 2018) اهتمت هذه الدراسة بنمذجة بيانات تكرار المطالبات التى تحتوى على عدد كبير من الأصفار فى تأمين السيارات، مع التركيز على أسباب وجود الأصفار، باستخدام نموذج انحدار بواسون ذى الأصفار الزائدة.

دراسة (Kusuma et al., 2019) اهتمت هذه الدراسة بنمذجة تكرار المطالبات فى التأمين الصحى باستخدام نموذج انحدار بواسون ذى الأصفار الزائدة، وتم المقارنة بين نموذج انحدار بواسون العادى ونموذج انحدار بواسون ذى الأصفار الزائدة.

وقدمت بعض الدراسات تقديرات لمعلمات توزيع بواسون ذى الأصفار الزائدة :

دراسة (Backett et al., 2014) اهتمت هذه الدراسة بتقدير معلمات توزيع بواسون ذى الأصفار الزائدة بطريقة العزوم وطريقة الامكان الأعظم، وتم استخدامه فى نمذجة بيانات الكوارث الطبيعية.

دراسة (Wagh et al., 2018) قدمت هذه الدراسة العديد من النماذج ذات الأصفار الزائدة، وتم الاهتمام بتقدير معلمات توزيع بواسون ذى الأصفار الزائدة.

دراسة (Sakthivel et al., 2018) اهتمت هذه الدراسة بتقديم تقديرات العزوم والامكان الأعظم لمعلمات توزيع بواسون ذى الأصفار الزائدة، واستخدام مؤشر الأصفار الزائدة لاختبار ما إذا كان يوجد بالبيانات أصفار زائدة أم لا.

ملاحظات على الدراسات السابقة

- اهتمت بعض الدراسات السابقة مثل (Sarul et al., 2015) بنمذجة بيانات تكرار المطالبات فى تأمين السيارات باستخدام نماذج الانحدار ذات الأصفار الزائدة، ونماذج انحدار هاردل والمقارنة بينهم. واهتم كل من (Guillen et al., 2018)، (Wolny-Dominiak, 2013) بنمذجة تكرار المطالبات فى تأمين السيارات باستخدام نماذج الانحدار ذات الأصفار الزائدة. وركز كلا من (Kusuma et al., 2019)، (Moualassim et al., 2012) على نمذجة تكرار المطالبات فى التأمين الصحى باستخدام نموذج انحدار بواسون ذى الأصفار الزائدة ومقارنته مع نموذج انحدار بواسون العادى. أما دراسة (Yip et al., 2005) فقد ركزت على نمذجة تكرار المطالبات فى التأمينات العامة باستخدام نماذج الانحدار ذات الأصفار الزائدة. وبذلك لم تستخدم أى دراسة من هذه الدراسات التوزيعات الأصلية واستخدمت نماذج الانحدار فقط.
- واهتم كل من (Backett et al., 2014) ، (Wagh et al., 2018) ، (Sakthivel et al., 2018) بتقدير معلمات توزيع بواسون ذى الأصفار الزائدة، ولم تتطرق أى دراسة لتقدير معلمات توزيع هاردل بواسون.

وبناء على ما تقدم، يهتم البحث الحالى باستخدام توزيع بواسون ذى الأصفار الزائدة، وتوزيع هاردل بواسون فى نمذجة بيانات مطالبات تأمين السيارات. ويهتم هذا البحث أيضا باشتقاق المتوسط والتباين لتوزيع هاردل بواسون واشتقاق مقدرات العزوم والامكان الأعظم لمعلمات توزيع هاردل بواسون المجهولة.

(١-٥) خطة البحث

الجزء الأول: مقدمة البحث
الجزء الثانى: الاختبارات المستخدمة لاكتشاف الأصفار الزائدة
الجزء الثالث: التوزيعات الاحتمالية المستخدمة فى النمذجة
الجزء الرابع: الدراسة التطبيقية

٢. الاختبارات المستخدمة لاكتشاف الأصفار الزائدة

(٢-١) مؤشر الأصفار الزائدة Zero-inflation index

قدم (Puig& Valero, 2006) مؤشر الأصفار الزائدة كمقياس لاكتشاف الأصفار الزائدة Zero inflation والتشتت الزائد (Over-dispersion) فى توزيع بواسون. فإذا كان Y هو رقم صحيح غير سالب بمتوسط θ، P_0 هو نسبة الأصفار فى البيانات

وتعرف بالاحتمال التجريبى للأصفار فى البيانات التى حجمها n فإن هذا المؤشر يعرف بأنه:

$$Z = 1 + \frac{\ln(P_0)}{\theta} \qquad (1)$$

فإذا كانت قيمة Z تساوى الصفر فإن البيانات تتبع التوزيع البواسونى، وإذا كانت قيمة Z أكبر من الصفر فإن البيانات بها أصفار زائدة.

(٢-٢) استخدام اختبار (Score test) للأصفار الزائدة فى توزيع بواسون

قدم (Van den Broek, 1995) اختبارا لتحديد ما إذا كان يوجد بالبيانات أصفارا زائدة أم لا. فعندما يكون لدينا n من المشاهدات من بينها n_0 أصفار بدون وجود متغيرات مستقلة، فإن احصاء الاختبار (Score test) التى تستخدم لاختبار ما إذا كان توزيع بواسون يلائم البيانات أم لا هى:

$$S(\tilde{B}_1) = \frac{(n_0 - n\tilde{p}_0)^2}{n\tilde{p}_0(1-\tilde{p}_0) - n\tilde{P}_0{}^2\bar{y}} \qquad (2)$$

حيث:

$\tilde{p}_0$ هى قيمة احتمال حدوث مطالبة صفرية فى توزيع بواسون

n هى عدد المشاهدات

n_0 هى عدد المطالبات الصفرية فى البيانات.

$\bar{y}$ متوسط البيانات

الفرض العدمى فى هذا الاختبار يقضى بعدم وجود تشتت أكثر من اللازم فى البيانات ناتج عن زيادة فى الأصفار (Zero inflation) . وتحت صحة الفرض العدمى فإن هذه الاحصاءة لها توزيع كاى تربيع بدرجة حرية واحدة. وفى حالة رفض الفرض العدمى، فإن هذا يعنى وجود أصفار زائدة فى البيانات نتج عنها تشتت أكثر من اللازم (Over-dispersion) .

٣. التوزيعات الاحتمالية المستخدمة فى النمذجة

فى هذا الجزء يتم عرض ثلاثة توزيعات هم توزيع بواسون، توزيع بواسون ذى الأصفار الزائدة، وتوزيع هاردل بواسون.

(٣-١) توزيع بواسون (Poisson distribution)

توزيع بواسون من التوزيعات الهامة لأنه يمثل الأحداث النادرة، ويعد أيضا تقريبا جيدا لتوزيعى ذى الحدين وذى الحدين السالب عند توافر شروط معينة (Ashour, et al., 2009) ويمكن عرض دالة الاحتمال لتوزيع بواسون كالتالى:

$$f(y) = \frac{e^{-\lambda}\lambda^{y}}{y!} \qquad y = 0,1,2,3,\ldots\ldots \qquad (3)$$

وفى توزيع بواسون يتساوى المتوسط مع التباين ويكونا هما قيمة المعلمة المقدرة λ .

وعند استخدام طريقة الامكان الأعظم (Maximum likelihood method) لتقدير المعلمة المجهولة λ، تكون دالة الامكان الأعظم كالتالى:

$$L(y;\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y!} \qquad (4)$$

$$\ln\big(L(y;\lambda)\big) = -n\lambda + \left(\sum_{i=1}^{n} y_i\right)\ln(\lambda) - \sum_{i=1}^{n} ln(y_i!) \qquad (5)$$

وباجراء التفاضل بالنسبة للمعلمة λ والمساواة بالصفر حتى نصل للقيمة العظمى للمعلمة λ :

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (6)$$

ويتميز توزيع بواسون أن تقدير الامكان الأعظم يتساوى مع تقدير طريقة العزوم للمعلمة المجهولة λ.

(٣-٢) توزيع بواسون ذى الأصفار الزائدة Zero-inflated Poisson distribution

يمكن عرض الصورة العامة للتوزيعات ذات الأصفار الزائدة (Zero-inflated distributions) كالتالى (Wolny-Dominiak, 2013) :

$$P(Y=y)=\begin{cases}\pi+(1-\pi)f(y) & y=0\\ (1-\pi)f(y) & y>0\end{cases} \qquad (7)$$

وبذلك يمكن عمل توزيع ذى أصفار زائدة من أى توزيع متقطع مثل توزيع بواسون، توزيع ذى الحدين السالب وغيرها.

ويجب ملاحظة أن المعلمة π عندما تتروا‌ح بين الصفر والواحد الصحيح فإن التوزيع يصبح توزيع ذى أصفار زائدة (Zero-inflated distribution) ، أما إذا كانت قيمة المعلمة π سالبة فإن هذا يشير إلى وجود أصفار قليلة وذلك يعطى توزيع ذى أصفار قليلة (Zero-deflated distribution) (Yip et al., 2005).

نماذج الأصفار الزائدة عبارة عن نماذج مختلطة (Mixture models) حيث يتم تقديم التوزيع الكامل للنتائج عن طريق مكونين منفصلين، الجزء الأول هو نمذجة الأصفار الزائدة، والجزء الثانى لنمذجة الأصفار غير الزائدة وقيم المتغير غير الصفرية(Loeys et al., 2012) .

تم تقديم نموذج بواسون للأصفار الزائدة لأول مرة عن طريق (Lambert, 1992) للتعامل مع الزيادة فى الأصفار الموجودة فى البيانات المتقطعة (Count data) بالتطبيق على عيوب الصناعة (Defects in manufacturing) وقد أشار إلى أن الأصفار لها مصدرين فى هذا النموذج هما الأصفار الحقيقية (Real zeros) وهى الأصفار التى تتبع توزيع بواسون، والأصفار الزائدة (Excess zeros) وهى الأصفار الزائدة التى لم يستطيع توزيع بواسون العادى نمذجتها.

يمكن عرض دالة التوزيع البواسونى ذى الأصفار الزائدة (Beckett et al., 2014) :

$$P(Y=y)=\begin{cases}\pi+(1+\pi)\exp(-\theta) & y=0\\ (1-\pi)\dfrac{\exp(-\theta).\theta^{y}}{y!} & y>0\end{cases} \qquad (8)$$

ويجب ملاحظة أن $0<\pi<1$ ،$\theta>0$

وكما نلاحظ من صيغة توزيع بواسون ذى الأصفار الزائدة، أن π هى معلمة اضافية مسئولة عن حساب احتمال الأصفار الزائدة فى التوزيع، ويتم التعويض فى دالة توزيع بواسون عن $y=0$ للحصول على احتمال الأصفار الحقيقية مرجحا بالقيمة $(1-\pi)$ ويتم جمع القيمتين للحصول على احتمال أن يأخذ المتغير العشوائى القيمة صفر. ويتم التعويض فى دالة توزيع بواسون للحصول على احتمال أن يأخذ المتغير العشوائى أى قيمة أخرى غير الصفر ويتم ترجيح الاحتمال بالقيمة $(1-\pi)$.

ويمكن الحصول على متوسط وتباين توزيع بواسون ذى الأصفار الزائدة كالتالى : (Beckett et al., 2014)

$$E(Y) = \theta(1-\pi) \qquad (9)$$

$$Var(Y) = \theta(1-\pi)(1+\theta\pi) \qquad (10)$$

وكما نلاحظ من معادلة المتوسط والتباين أن التباين فى توزيع بواسون ذى الأصفار الزائدة يزيد عن المتوسط.

(٣-٢-١) تقدير معلمات توزيع بواسون ذى الأصفار الزائدة

عرض كلا من (Beckett et al., 2014)، (Sakthivel et al., 2018)، (Wagh, et al., 2018) طريقة العزوم ، وطريقة الامكان الأعظم لتقدير معلمات توزيع بواسون ذى الأصفار الزائدة كالتالى:

(٣-٢-١-١) طريقة العزوم (Method of moments)

نحصل على مقدرات طريقة العزوم بمساواة المتوسط والتباين المحسوب من البيانات محل الدراسة بالصيغة النظرية للمتوسط والتباين لتوزيع بواسون ذى الأصفار الزائدة.

وبذلك نحصل على تقديرات للمعلمات المجهولة لتوزيع بواسون ذى الأصفار الزائدة كالتالى:

$$\tilde{\theta} = \bar{Y} + \left(\frac{S^2}{\bar{Y}}\right) - 1 \qquad (11)$$

$$\tilde{\pi} = \frac{S^2 - \bar{Y}}{\bar{Y}^2 + (S^2 - \bar{Y})} \qquad (12)$$

حيث $\tilde{\theta}$ هو تقدير المعلمة θ باستخدام طريقة العزوم، $\tilde{\pi}$ هو تقدير المعلمة π باستخدام طريقة العزوم.

$\bar{Y}$ هو متوسط البيانات، S^2 هو تباين البيانات

$$\bar{Y} = \sum_{i=1}^{n} \frac{Y_i}{n} \qquad (13)$$

$$S^2 = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{n-1} \qquad (14)$$

(٣-٢-١-٢) طريقة الامكان الأعظم (Maximum likelihood estimator)

بفرض أن $y_1, y_2, \dots \dots \dots . y_n$ هى عينة عشوائية مأخوذة من توزيع بواسون ذى الأصفار الزائدة، بذلك فإن دالة الامكان الأعظم للتوزيع يمكن كتابتها كالتالى:

$$L = \left[\pi + (1-\pi)e^{-\theta}\right]^{n_0} \left[(1-\pi)e^{-\theta}\right]^{(n-n_0)} \left(\prod_{i=1}^{n} \frac{\theta^{y_i}}{y_i!}\right)$$

$$\begin{aligned} \ln L = n_0 \ln\left(\pi + (1-\pi)e^{-\theta}\right) + (n-n_0)\ln(1-\pi) \\ - (n-n_0)\theta + A[ln(\theta)] \\ - \sum_{i=1}^{n} \ln(y_i!) \end{aligned} \qquad (15)$$

حيث:

$$n_0 = \sum_{i=1}^{n} I\{y = 0\}$$

وهذه المعادلة تعنى أن n_0 هى مجموع الأصفار فى البيانات.

$$A = \sum_{i=1}^{n} y_i \qquad (16)$$

هذه المعادلة تعنى أن A هى مجموع قيم المتغير العشوائى y

وبعد اجراء التفاضلات اللازمة يمكن الحصول على المعادلات التالية:

$$\bar{Y}\left(1 - e^{-\hat{\theta}}\right) = \hat{\theta}\left(1 - \frac{n_0}{n}\right) \qquad (17)$$

ويمكن تبسيط المعادلة والحصول على تقدير الامكان الأعظم للمعلمة θ كالتالى:

من المعادلة (16)

$$\bar{Y} = \frac{A}{n} \qquad\qquad A = n\bar{Y}$$

وبالتعويض فى المعادلة (17) نحصل على الصيغة التالية:

$$\hat{\theta} = \frac{A(1 - e^{-\hat{\theta}})}{n - n_0} \qquad (18)$$

ويمكن حل هذه المعادلة باستخدام طريقة نيوتن رافسون (Newton Raphson)

والمعادلة الخاصة بالمعلمة π يمكن كتابتها كالتالى:

$$\hat{\pi} = 1 - \frac{\bar{Y}}{\hat{\theta}} \qquad (19)$$

وبالتعويض عن قيمة $\hat{\theta}$ ، وقيمة $\bar{Y}$ فى المعادلة (19) يمكن الحصول على شكل آخر لتقدير المعلمة π كالتالى:

$$\hat{\pi} = \frac{P_0 - e^{-\hat{\theta}}}{1 - e^{-\hat{\theta}}} \qquad (20)$$

حيث : $P_0 = \frac{n_0}{n}$ وهى نسبة الأصفار الموجودة فى البيانات.

(٣-٣) توزيع هاردل بواسون Hurdle Poisson distribution

تم تقديم نموذج هاردل (Hurdle model) لأول مرة عن طريق (Mullahy, 1986) للتعامل مع التكرارات التى تحتوى على أصفار زائدة أكبر من تلك المسموح بها بموجب افتراضات التوزيع الاحتمالى سواء كان توزيع بواسون أو توزيع ذى الحدين السالب، ويعد نموذج هاردل بديلا عن نموذج الأصفار الزائدة.

يمكن عرض التوزيع الاحتمالى لنموذج هاردل كالتالى (Hofstetter, et al., 2016):

$$P(Y = y) = \begin{cases} \pi_0 & y = 0 \\ (1 - \pi_0)\dfrac{f(y, \lambda)}{1 - f(0, \lambda)} & y > 0 \end{cases} \qquad (21)$$

وكما يتضح من الصيغة السابقة، عندما تكون $y = 0$ فإن احتمال الأصفار كلها يتم تحديده عن طريق تقدير المعلمة π_0 ، وعندما تكون $y > 0$ يمكن حساب الاحتمال باستخدام صيغة التوزيع المبتور (Truncated distribution) مرجحا بالقيمة $(1 - \pi_0)$.

وبذلك يتكون نموذج هاردل (Hurdle model) من جزئين؛ الأول لنمذجة الأصفار فقط، والجزء الثانى هو توزيع مبتور الصفر حيث يتم نمذجة بقية القيم بدون وجود الصفر (Hofstetter et al., 2016)

يمكن عرض دالة التوزيع الاحتمالى لتوزيع هاردل بواسون كالتالى (Sarul et al., 2015):

$$P(Y = y) = \begin{cases} \pi_0 & y = 0 \\ (1 - \pi_0)\dfrac{e^{-\alpha}\alpha^y}{(1 - e^{-\alpha})y!} & y > 0 \end{cases} \qquad (22)$$

وتمثل π_0، α معلمات التوزيع المجهولة.

$\frac{e^{-\alpha}\alpha^y}{(1-e^{-\alpha})y!}$ ويمثل هذا المقدار توزيع بواسون المبتور الصفر (Zero truncated Poisson)

(٣-٣-١) اشتقاق متوسط وتباين توزيع هاردل بواسون

يمكن الوصول إلى صيغة المتوسط والتباين لتوزيع هاردل بواسون من المبادئ الأولية للاحصاء كالتالى:

(٣-٣-١-١) متوسط توزيع هاردل بواسون

يمكن حساب متوسط توزيع هاردل بواسون كالتالى:

$$E(Y) = (0 * \pi_0) + 1 * \frac{(1 - \pi_0)e^{-\alpha}\alpha^1}{1!\,(1 - e^{-\alpha})} + 2 * \frac{(1 - \pi_0)e^{-\alpha}\alpha^2}{2!\,(1 - e^{-\alpha})} + \cdots$$

$$E(Y) = (1 - \pi_0)\sum_{y=1}^{\infty} y\,\frac{e^{-\alpha}\alpha^y}{y!\,(1 - e^{-\alpha})}$$

$$E(Y)=\frac{(1-\pi_0)\alpha e^{-\alpha}}{(1-e^{-\alpha})}\sum_{y=1}^{\infty}\frac{\alpha^{(y-1)}}{(y-1)!}$$

Let X=(y-1)

$$E(Y)=\frac{(1-\pi_0)\alpha e^{-\alpha}}{(1-e^{-\alpha})}\sum_{X=0}^{\infty}\frac{\alpha^{X}}{X!}$$

Since

$$\sum_{X=0}^{\infty}\frac{\alpha^{X}}{X!}=e^{\alpha}$$

$$E(Y)=\frac{\alpha(1-\pi_0)}{(1-e^{-\alpha})} \qquad (23)$$

(٣-١-٣-٢) حساب التباين لتوزيع هاردل بواسون

يمكن حساب التباين عن طريق حساب العزوم العاملى

$$E\big(Y(Y-1)\big)=\frac{(1-\pi_0)e^{-\alpha}}{(1-e^{-\alpha})}\sum_{y=1}^{\infty}\frac{y(y-1)\alpha^{y}}{y!}$$

$$E\big(Y(Y-1)\big)=\frac{(1-\pi_0)\alpha^2 e^{-\alpha}}{(1-e^{-\alpha})}\sum_{y=1}^{\infty}\frac{\alpha^{(y-2)}}{(y-2)!}$$

Let X= y-1

$$E\big(Y(Y-1)\big)=\frac{(1-\pi_0)\alpha^2 e^{-\alpha}}{(1-e^{-\alpha})}\sum_{X=0}^{\infty}\frac{\alpha^{(X-1)}}{(X-1)!}$$

$$E\big(Y(Y-1)\big)=\frac{(1-\pi_0)\alpha^2}{(1-e^{-\alpha})}$$

$$E(Y^2) = E\big(Y(Y-1)\big) + E(Y)$$

$$E(Y^2) = \frac{(1-\pi_0)\alpha^2}{(1-e^{-\alpha})} + \frac{(1-\pi_0)\alpha}{(1-e^{-\alpha})}$$

$$E(Y^2) = \frac{(1-\pi_0)(\alpha^2+\alpha)}{(1-e^{-\alpha})} \qquad (24)$$

$$Var(Y) = E(Y^2) - (E(Y))^2$$

$$Var(Y) = \frac{(1-\pi_0)(\alpha^2+\alpha)}{(1-e^{-\alpha})} - \left(\frac{\alpha(1-\pi_0)}{(1-e^{-\alpha})}\right)^2$$

$$Var(Y) = \frac{\alpha(1-\pi_0)}{(1-e^{-\alpha})}\left[(\alpha+1) - \frac{\alpha(1-\pi_0)}{(1-e^{-\alpha})}\right] \qquad (25)$$

وبذلك يمكن كتابة متوسط توزيع هاردل بواسون بدلالة المتوسط والمعلمة α كالتالى:

$$Var(Y) = E(Y)[(\alpha+1) - E(Y)] \qquad (26)$$

(٣-٣-٢) تقدير معلمات توزيع هاردل بواسون باستخدام طريقة العزوم

يتم مساواة العزم الأول والثانى حول الصفر للتوزيع بالعزم الأول والثانى حول الصفر المحسوب من البيانات، ويمكن توضيح ذلك كالتالى:

$$\acute{m}_1 = \frac{\alpha(1-\pi_0)}{(1-e^{-\alpha})} \qquad (27)$$

$$\acute{m}_2 = \frac{(1-\pi_0)(\alpha^2+\alpha)}{(1-e^{-\alpha})} \qquad (28)$$

بقسمة المعادلة (27) على المعادلة (28)

$$\frac{\acute{m}_2}{\acute{m}_1} = \frac{\alpha^2+\alpha}{\alpha}$$

$$\tilde{\alpha} = \frac{\acute{m}_2}{\acute{m}_1} - 1$$

$$\tilde{\alpha} = \frac{\acute{m}_2 - \acute{m}_1}{\acute{m}_1} \qquad (29)$$

ومن المعادلة (27) يمكن إيجاد تقدير المعلمة π_0

$$\tilde{\pi}_0 = 1 - \left(\frac{\acute{m}_1(1 - e^{-\alpha})}{\alpha}\right) \qquad (30)$$

(٣-٣-٣) تقدير معلمات توزيع هاردل بواسون باستخدام طريقة الامكان الأعظم

باستخدام المعادلة (22) يمكن ايجاد دالة الامكان الأعظم لتوزيع هاردل بواسون كالتالى:

$$L = (\pi_0)^{n_0} \prod_{\substack{i=1 \\ y_i \neq 0}}^{n-n_0} \frac{(1-\pi_0)e^{-\alpha}\alpha^{y_i}}{(1-e^{-\alpha})y_i!} \qquad (31)$$

نلاحظ فى دالة الامكان الأعظم فصل كامل بين المشاهدات الصفرية والمشاهدات غير الصفرية.

وبأخذ لوغاريتم الطرفين للأساس الطبيعى e للمعادلة (31) نحصل على الآتى:

$$\begin{aligned} \ln(L) = n_0 \ln(\pi_0) + (n-n_0)\ln(1-\pi_0) - (n-n_0)\alpha \\ - (n-n_0)\ln(1-e^{-\alpha}) \\ + \sum_{\substack{i=1 \\ y_i \neq 0}}^{n-n_0} y_i(\ln(\alpha)) \\ - \sum_{\substack{i=1 \\ y_i \neq 0}}^{n-n_0} \ln(y_i!) \qquad (32) \end{aligned}$$

Let

$$B = \sum_{\substack{i=1 \\ y_i \neq 0}}^{n-n_0} y_i \qquad (33)$$

بالتعويض من المعادلة (33) فى المعادلة (32)

$$\ln(L) = n_0 \ln(\pi_0) + (n - n_0)\ln(1 - \pi_0) - (n - n_0)\alpha - (n - n_0)\ln(1 - e^{-\alpha}) + B(\ln(\alpha)) - \sum_{\substack{i=1 \\ y_i \neq 0}}^{n-n_0} \ln(y_i!) \qquad (34)$$

باجراء التفاضل الأول للمعادلة (34) بالنسبة للمعلمة π_0

$$\frac{\partial \ln(L)}{\partial \pi_0} = \frac{n_0}{\pi_0} - \frac{n - n_0}{1 - \pi_0} \qquad (35)$$

وبمساواة المعادلة (35) بالصفر لنحصل على تقديرللقيمة العظمى للمعلمة π_0 وهو:

$$\hat{\pi}_0 = \frac{n_0}{n} \qquad (36)$$

باجراء التفاضل الأول للمعادلة (34) بالنسبة للمعلمة α

$$\frac{\partial \ln(L)}{\partial \alpha} = -\frac{(n - n_0)}{(1 - e^{-\alpha})} - (n - n_0) + \frac{B}{\alpha} \qquad (37)$$

وبمساواة المعادلة (37) بالصفر نحصل على تقدير للقيمة العظمى للمعلمة α وهو:

$$\widehat{\alpha} = \frac{B(1 - e^{-\alpha})}{n - n_0} \qquad (38)$$

مع ملاحظة أن الرمز B تم تعريفه فى المعادلة (33).

(٣-٤) الاختيار بين توزيع بواسون ذى الأصفار الزائدة، وتوزيع هاردل بواسون

يرجع الاختلاف الرئيسى بين توزيع هاردل بواسون وتوزيع بواسون ذى الأصفار الأصفار الزائدة هو أن توزيع هاردل بواسون يقوم بنمذجة البيانات الصفرية بشكل مستقل تماما عن نمذجة البيانات غير الصفرية؛ حيث تمثل المعلمة π_0 توزيع الأصفار كلها، والمعلمة α معلمة توزيع بواسون المبتور الصفر الذى يمثل البيانات غير الصفرية. وبالنسبة لتوزيع بواسون ذى الأصفار الزائدة، يفترض هذا التوزيع وجود نوعين من الأصفار الأصفار الحقيقية، والأصفار الزائدة حيث تمثل المعلمة θ معلمة توزيع بواسون والمعلمة π هى معلمة التشتت الزائد (Over-dispersion) (Sarul et al., 2015).

(٣-٤-١) الاختيار على أساس الاعتبارات النظرية

إن الاختيار بين توزيع هاردل بواسون وتوزيع بواسون ذى الأصفار الزائدة يعتمد على الاعتبارات النظرية (Theoretical consideration) حيث يفترض توزيع بواسون ذى الأصفار الزائدة وجود نوعين من الأصفار أصفار حقيقية، وأصفار زائدة. وعلى العكس من ذلك يفترض توزيع هاردل بواسون أن الأصفار كلها أصفار حقيقية (Hofstetter et al., 2015).

(٣-٤-٢) الاختيار على أساس المعايير الاحصائية

يمكن الاختيار بين التوزيعات المختلفة التى أثبتت اختبارات جودة التوفيق صلاحيتها لتمثيل البيانات باستخدام معيار (Akaike's Information Criterion) (AIC) ومعيار (Schwarz Bayesian Information Criterion) (BIC) ، ويمكن عرض شكل إحصاء الاختبار كالتالى (Maydeu-Olivares et al., 2010) :

معيار AIC

$$AIC = -2\ln(L) + 2k \qquad (39)$$

معيار BIC

$$BIC = -2\ln(L) + k.ln(n) \qquad (40)$$

حيث:

$\ln(L)$: هو لوغاريتم دالة الامكان الأعظم بعد التعويض عن المعلمات المجهولة والقيم الأخرى من البيانات.

k : هو عدد المعلمات المقدرة

$\ln(n)$: هو لوغاريتم عدد المشاهدات

ويتم اختيار التوزيع الذى له أقل قيمة لمعيار AIC أو معيار BIC .

وبناء على ما تقدم يتم الاختيار بين توزيع بواسون ذى الأصفار الزائدة وتوزيع هاردل بواسون على أساس منشأ الأصفار، فإذا كان القائم بعملية النمذجة يعلم أن بعض الأصفار حقيقية وتمثل حدث عدم وجود مطالبات بالفعل، وبعض الأصفار الأخرى غير حقيقية (زائدة) وتمثل حوادث وقعت بالفعل ولم يتم الابلاغ عنها لأسباب تتعلق بحد التحمل الموجود فى الوثائق أو عدم رغبة حامل الوثيقة فى أن يخسر خصم عدم المطالبة للعام التالى، بذلك يكون الافتراض الواقعى هو اختيار توزيع ذى أصفار زائدة. أما إذا كانت جميع الأصفار الموجودة فى بيانات تكرار المطالبات هى أصفار حقيقية بمعنى أنها تمثل عدم وقوع حادث بالفعل يكون الاختيار الأمثل هو اختيار توزيع هاردل للتوزيع المتقطع الذى تتبعه البيانات. أما إذا كان من الصعب تحديد منشأ الأصفار فى البيانات التى نرغب فى نمذجتها تتم عملية النمذجة باستخدام توزيع ذى أصفار زائدة ، وتوزيع هاردل لنفس التوزيع والاختيار بينهما باستخدام المعايير الاحصائية مثل معيار (AIC)، معيار (BIC).

٤. الدراسة التطبيقية

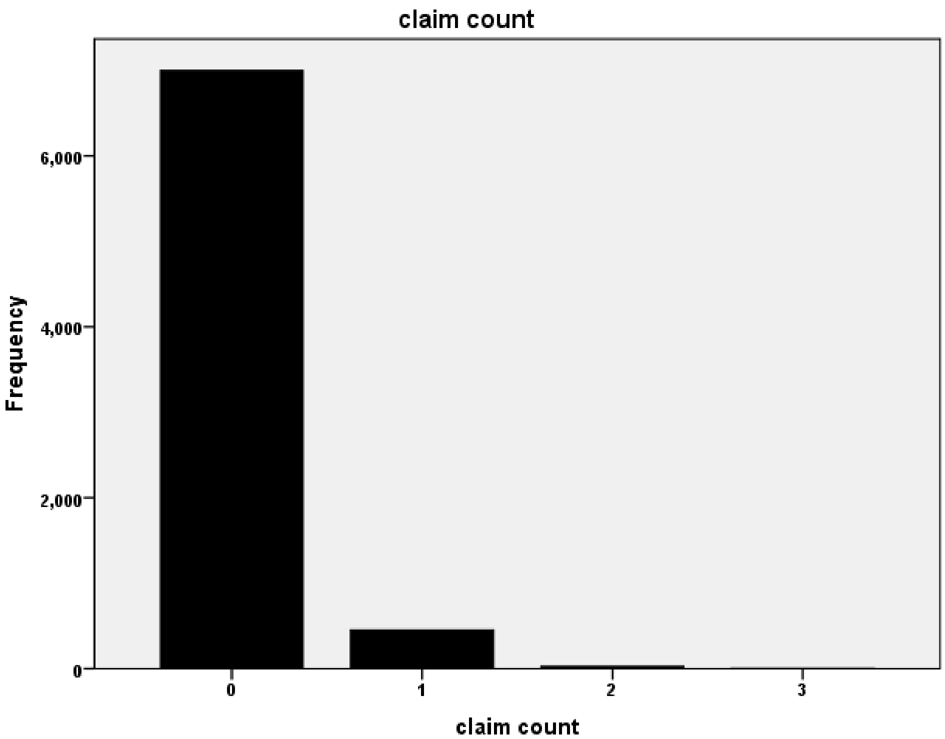
لقد تم الاعتماد على قاعدة بيانات تأمين السيارات لسوق سنغافورة والمتاحة على الشبكة الدولية للمعلومات (قائمة المراجع، المرجع رقم 18) ، ومصدر البيانات الرئيسى هو جمعية التأمينات العامة بسنغافورة General Insurance Association of Singapore، حيث تم استخدام بيانات تكرار المطالبات.

(٤-١) وصف بيانات البحث

جدول (١) التوزيع التكرارى لعدد المطالبات

قيم المطالبات	التكرار المشاهد	التكرار النسبى المشاهد	التكرار التراكمى النسبى المشاهد
0	6996	93.5	93.5
1	455	6.1	99.6
2	28	.4	99.9
3	4	.1	100.0
المجموع	7483	100.0	

وكما نلاحظ من جدول التوزيع التكرارى وجود 93.5% من المشاهدات لها قيمة صفرية. والعدد الكلى للمشاهدات هو 7483، والتكرارت تقل كلما زاد عدد المطالبات، فى البداية عندما نرغب فى نمذجة هذ النوع من البيانات فإن أول توزيع يأتى فى المقدمة هو توزيع بواسون، لأنه هو التوزيع الذى يمثل الأحداث النادرة، ولكن يجب التأكد أولا من توافر شرط تطبيق توزيع بواسون وهو تساوى المتوسط الحسابى للتوزيع مع تباينه.



شكل (١) المدرج التكرارى الذى يمثل تكرار المطالبات

نلاحظ من الشكل السابق أن بيانات تكرار المطالبات تحتوى على عدد كبير من الأصفار، والبيانات ملتوية ناحية اليمين، ويقل تكرار المطالبة كلما زاد عدد المطالبات.

جدول (٢) بعض المقاييس الاحصائية للبيانات

المقياس الاحصائى	القيمة
عدد المشاهدات	7483

عدد المشاهدات الصفرية	6996
عدد المشاهدات غير الصفرية	487
العزم الأول حول الصفر =الوسط الحسابى	$0.069891755 \approx 0.07$
العزم الثانى حول الصفر	0.080582654
التباين	$0.075697796 \approx 0.076$
الانحراف المعيارى	0.2751323
معامل الالتواء	4.272
أقل قيمة للمشاهدات	0
أعلى قيمة للمشاهدات	3
مجموع المشاهدات	523
مجموع المشاهدات غير الصفرية	523
مجموع لوغاريتم مضروب المشاهدات $\sum_{i=1}^{7483} ln(y_i!)$	26.57515
مجموع لوغاريتم مضروب المشاهدات غير الصفرية $\sum_{\substack{i=1 \\ y_i \neq 0}}^{487} ln(y_i!)$	26.57515

وكما يتضح من جدول (٢) أن عدد المشاهدات الكلى هو 7483 مطالبة، وعدد المشاهدات الصفرية هو 6996 وعدد المشاهدات غير الصفرية هو 487 ومتوسط البيانات 0.07 وهو قيمة منخفضة نظرا لوجود عدد كبير من المشاهدات الصفرية. والتباين هو 0.076 وبذلك فإن التباين أكبر من الوسط الحسابى وهذا يعنى وجود تشتت أكبر من اللازم (Over-dispersion) وفقا لتوزيع بواسون، والبيانات ملتوية ناحية اليمين نظرا لأن معامل الالتواء قيمته موجبة. وكما نلاحظ أيضا أن أقل قيمة للمشاهدات هى الصفر وهذا يعنى عدم وجود مطالبات خلال عام الوثيقة، وأعلى قيمة للمشاهدات هو 3 وهذا يعنى وجود ثلاث مطالبات خلال عام الوثيقة. ونلاحظ أيضا أن مجموع المشاهدات الكلية يساوى 523 ومجموع المشاهدات غير الصفرية يساوى 523 أيضا لأن قيمة المشاهدة الصفرية لا تؤثر فى المجموع. وقد تم ايجاد مجموع لوغاريتم مضروب المشاهدات كلها، ومجموع لوغاريتم مضروب المشاهدات غير الصفرية، ولا يوجد اختلاف بين القيمتين نظرا لأن مضروب الصفر يساوى الواحد الصحيح ولوغاريتم الواحد الصحيح يساوى الصفر بذلك لا تؤثر القيم الصفرية على

المجاميع. وتم حساب آخر أربع قيم فى جدول (٢) لاستخدامهم فى حساب معيارى (AIC), (BIC).

(٤-٢) اختبار اكتشاف الأصفار الزائدة

يمكن استخدام نوعين من الاختبارات للتأكد من وجود الأصفار الزائدة فى مشاهدات المتغير التابع؛ هما مؤشر الأصفار الزائدة، واختبار (Score test).

(٤-٢-١) تطبيق مؤشر الأصفار الزائدة على تكرارات المطالبات

ومن جدول (٢):

$$n = 7483, \qquad n_0 = 6996$$

$$P_0 = \frac{6996}{7483} = 0.9349191501$$

$$\bar{Y} = \theta = 0.07 \qquad (41)$$

وبذلك يمكن تطبيق مؤشر الأصفار الزائدة

$$z = 1 + \frac{\ln(P_0)}{\theta}$$

$$z = 0.03863965838$$

نلاحظ أن قيمة z لا تساوى صفر، هذا يعنى أن البيانات لا يمكن أن تتبع توزيع بواسون، ونلاحظ أيضا أن قيمة z موجبة فهذا يعنى وجود تشتت أكثر من اللازم (Over-dispersion) ناتج عن زيادة عدد الأصفار فى بيانات تكرار المطالبات.

(٤-٢-٢) تطبيق اختبار Score Test لاكتشاف الأصفار الزائدة فى البيانات

$$S(\beta) = \frac{(6996 - 7483 * e^{-0.07})^2}{7483e^{-0.07}(1 - e^{-0.07}) - 7483 * 0.07 * (e^{-0.07})^2}$$

$$S(\beta) = 21.885$$

وبمقارنة قيمة الاختبار المحسوبة بالقيمة الجدولية لتوزيع كاى تربيع عند درجة حرية واحدة وهى (6.6349) عند مستوى معنوية 1% ، وهذا يعنى رفض الفرض العدمى الذى يقضى بأن المعلمة الاضافية لتوزيع بواسون ذى الأصفار الزائدة

تساوى الصفر، وهذا يعنى وجود تشتت زائد فى البيانات قد يكون مصدره وجود أصفار زائدة فى البيانات.

وللتأكد من نتائج الاختبارات السابقة يتم توفيق البيانات باستخدام توزيع بواسون

(٤-٣) توفيق البيانات باستخدام توزيع بواسون

يتم تقدير معلمة توزيع بواسون المجهولة، ونلاحظ أن استخدام طريقة العزوم أو طريقة الامكان الأعظم يعطى نفس النتيجة.

$$\hat{\lambda} = \bar{Y} = 0.07 \qquad (42)$$

جدول (٣) يوضح حساب الاحتمالات المختلفة باستخدام توزيع بواسون، والتكرار المتوقع، وحساب قيمة احصاء اختبار كاى تربيع Chi square

جدول (٣) توفيق بيانات تكرار المطالبات باستخدام توزيع بواسون

قيم المتغير	التكرار المشاهد	الاحتمالات المحسوبة من توزيع بواسون	التكرار المتوقع	اختبار chi square
0	6996	0.9324	6977.1492	0.050930925
1	455	0.06527	488.41541	2.286147412
2	28	0.00228	17.06124	7.013351335
3	4	0.00005	0.37415	35.13774749
المجموع	7483	1	7483	44.48817716

ومن جدول (٣)، نلاحظ أن القيمة المحسوبة لاختبار كاى تربيع chi square أكبر من القيمة الجدولية عند درجتين حرية (وهى عدد المجموعات مطروحا منها عدد المعلمات المقدرة مطرحا منها واحد صحيح) ومستوى معنوية 1% (9.2104) وبذلك يمكن رفض الفرض العدمى الذى يقضى بتبعية البيانات لتوزيع بواسون.

وبالرجوع إلى نتائج اختبارات الأصفار الزائدة، وما أظهره اختبار جودة التوفيق أن توزيع بواسون العادى لا يلائم البيانات، فهذا يعنى أن توزيع بواسون ذى الأصفار الزائدة قد يعطى توفيقا أفضل للبيانات.

(٤-٤) توفيق البيانات باستخدام توزيع بواسون ذى الأصفار الزائدة

(٤-٤-١) استخدام مقدرات طريقة الامكان الأعظم

تم حساب معلمات التوزيع المجهولة باستخدام طريقة الامكان الأعظم (Maximum likelihood estimator) ، ويتم استخدام الصيغة المبسطة للتقدير حتى يمكن تحويلها إلى معادلة صفرية يسهل حلها على برنامج MathCad

$$\bar{Y}\left(1-e^{-\theta}\right)=\theta\left(1-\frac{n_0}{n}\right)$$

$$0.07-0.07e^{-\theta}-\theta+\frac{6996}{7483}\theta=0$$

$$0.07-0.07e^{-\theta}-0.0650808\theta=0$$

وباستخدام برنامج MathCad يمكن حل هذه المعادلة

$$\hat{\theta}=0.148 \qquad (43)$$

$$\hat{\pi}=1-\frac{\bar{Y}}{\theta}$$

$$\hat{\pi}=0.52703 \qquad (44)$$

وبالتعويض فى دالة توزيع بواسون ذى الأصفار الزائدة يمكن الحصول على الاحتمالات كما هى موضحة فى جدول(٤)

جدول (٤) توفيق بيانات تكرار المطالبات باستخدام توزيع بواسون ذى الأصفار الزائدة (مقدرات طريقة الامكان الأعظم)

قيم المتغير	التكرار المشاهد	قيم الاحتمالات المحسوبة من التوزيع	التكرار المتوقع	اختبار Chi square
0	6996	0.93493	6996.081	9.42E-07
1	455	0.06037	451.7487	0.0234
2	28	0.0045	33.6735	0.955903
3	4	0.0002	1.4966	4.187499
المجموع	7483	1	7483	5.166803

ومن جدول (٤) نلاحظ أن قيمة كاى تربيع Chi square المحسوبة أقل من القيمة الجدولية عند درجة حرية واحدة (عدد المجموعات مطروحا منها عدد المعلمات المقدرة مطروحا منها واحد صحيح) ومستوى معنوية 1% (6.6349) بذلك لا يمكن رفض الفرض العدمى الذى يقضى بتبعية البيانات لتوزيع بواسون ذى الأصفار الزائدة.

ويمكن حساب متوسط وتباين البيانات باستخدام متوسط وتباين توزيع بواسون ذى الأصفار الزائدة والمعلمات المقدرة بطريقة الامكان الأعظم باستخدام المعادلات أرقام ((9), (10), (42), (43)) كالتالى:

$$E(Y) = 0.06999956 \approx 0.07$$

$$Var(Y) = 0.0754596 \approx 0.075$$

ويتضح من استخدام متوسط وتباين التوزيع أن القيم تقترب بشدة من القيم الفعلية المحسوبة من البيانات والموضحة فى جدول (٢).

(٤-٤-٢) استخدام مقدرات طريقة العزوم

باستخدام المعادلات أرقام ((14) ,(13) ,(12) ,(11)) يمكن الحصول على تقديرات المعلمات المجهولة لتوزيع بواسون ذى الأصفار الزائدة بطريقة العزوم كالتالى:

$$\tilde{\theta} = 0.1557 \qquad (45)$$

$$\tilde{\pi} = 0.5505 \qquad (46)$$

وباستخدام المعادلات أرقام ((46) ,(45) ,(8)) يمكن حساب الاحتمالات المختلفة لتوزيع بواسون ذى الأصفار الزائدة باستخدام مقدرات العزوم كما هو موضح بالجدول التالى:

جدول (٥) توفيق مطالبات تكرار المطالبات باستخدام توزيع بواسون ذى الأصفار الزائدة (مقدرات طريقة العزوم)

قيم المتغير	التكرار المشاهد	قيم الاحتمالات المحسوبة من التوزيع	التكرار المتوقع	Chi square Test
0	6996	0.9352	6998.102	0.0006311
1	455	0.0599	448.2317	0.1022014
2	28	0.0047	35.1701	1.4617625
3	4	0.0002	1.4966	4.1874994
المجموع	7483	1	7483	5.7520944

وكما نلاحظ من جدول (٥) أن القيمة المحسوبة لاختبار كاى تربيع Chi square أقل من القيمة الجدولية عند درجة حرية واحدة وعند مستوى معنوية 1% (6.6349) وهذا يعنى أنه لا يمكن رفض الفرض العدمى الذى يقضى بتبعية البيانات لتوزيع بواسون ذى الأصفار الزائدة.

وباستخدام المعادلتين ((10) ,(9)) يمكن حساب المتوسط والتباين لتوزيع بواسون ذى الأصفار الزائدة:

$$E(Y) = 0.06998715 \approx 0.07$$

$$Var(Y) = 0.0759859 \approx 0.076$$

وكما نلاحظ أن المتوسط والتباين باستخدام مقدرات طريقة العزوم هو تقريبا نفس المتوسط والتباين المحسوب من البيانات.

(٤-٥) نمذجة البيانات باستخدام توزيع هاردل بواسون

(٤-٥-١) استخدام مقدرات طريقة الامكان الأعظم

باستخدام المعادلات أرقام ((38), (36), (33)) وبرنامج MathCad يمكن الحصول على تقديرات الامكان الأعظم للمعلمات المجهولة لتوزيع هاردل بواسون.

$$\hat{\pi}_0 = 0.93491915 \qquad (47)$$

$$\hat{\alpha} = 0.144 \qquad (48)$$

وباستخدام المعادلات أرقام (48) ,(47) ,(22) يمكن حساب الاحتمالات المختلفة لتوزيع هاردل بواسون كما هو موضح فى الجدول التالى:

جدول (٦): توفيق بيانات تكرار المطالبات باستخدام توزيع هاردل بواسون (مقدرات الامكان الأعظم)

قيم المتغير	التكرار المشاهد	الاحتمالات باستخدام توزيع هاردل بواسون	التكرار المتوقع	Chi square Test
0	6996	0.93492	6996.00636	5.78181E-09
1	455	0.06051	452.79633	0.010724825
2	28	0.00436	32.62588	0.655883175
3	4	0.00021	1.57143	3.753238926
المجموع	7483	1	7483	4.419846932

وكما نلاحظ من جدول (٦) أن القيمة المحسوبة لاختبار كاى تربيع Chi square أقل من القيمة الجدولية عند درجة حرية واحدة ومستوى معنوية 1% (6.6349) وهذا يعنى أنه لا يمكن رفض الفرض العدمى الذى يقضى بتبعية البيانات لتوزيع هاردل بواسون.

وباستخدام المعادلتين ((25), (23)) يمكن حساب المتوسط والتباين لتوزيع هاردل بواسون:

$$E(Y) = 0.069879 \approx 0.07$$

$$Var(Y) = 0.07505859 \approx 0.075$$

وهذه القيم تقترب بشدة من القيم المحسوبة من البيانات.

(٤-٥-٢) استخدام مقدرات طريقة العزوم

يمكن إيجاد تقدير المعلمات المجهولة فى توزيع هاردل بواسون باستخدام طريقة العزوم باستخدام المعادلتين ((30) ,(29)) والعزم الأول حول الصفر (الوسط الحسابى) والعزم الثانى حول الصفر المحسوب من البيانات والموضح فى جدول (٢):

$$\acute{m}_1 = 0.069891755$$

$$\acute{m}_2 = 0.080582654$$

$$\tilde{\alpha} = 0.153 \qquad (49)$$

$$\tilde{\pi}_0 = 0.9351924 \qquad (50)$$

وباستخدام المعادلات أرقام ((50) ,(49) ,(22)) يمكن حساب الاحتمالات المختلفة لتوزيع هاردل بواسون باستخدام تقديرات طريقة العزوم كما فى الجدول التالى:

جدول (٧): توفيق بيانات تكرار المطالبات باستخدام توزيع هاردل بواسون (مقدرات طريقة العزوم)

اختبار Chi square	التكرار المتوقع	الاحتمالات باستخدام توزيع هاردل بواسون	التكرار المشاهد	قيم المتغير
0.000587	6998.027	0.93519	6996	0
0.084809	448.8303	0.05998	455	1
1.198064	34.4218	0.0046	28	2
3.017524	1.72109	0.00023	4	3
4.300984	7483	1	7483	المجموع

وكما يتضح من جدول (٧) أن القيمة المحسوبة لاختبار كاى تربيع Chi square أقل من القيمة الجدولية عند درجة حرية واحدة ومستوى معنوية 1% (6.6349) وهذا يعنى أنه لا يمكن رفض الفرض العدمى الذى يقضى بتبعية البيانات لتوزيع هاردل بواسون.

وباستخدام المعادلتين ((23) ,(25)) يمكن حساب المتوسط والتباين لتوزيع هاردل بواسون باستخدام تقديرات طريقة العزوم:

$$E(Y) = 0.069892 \approx 0.07$$

$$Var(Y) = 0.0757 \approx 0.076$$

وهذه القيم هى تقريبا نفس المتوسط والتباين المحسوب من البيانات.

جدول (٨) ملخص المعلمات المقدرة والمتوسط والتباين لكل توزيع بطريقة العزوم والامكان الأعظم

بيان	توزيع بواسون	توزيع بواسون ذى الأصفار الزائدة	توزيع هاردل بواسون
المعلمات المقدرة بطريقة العزوم	$\tilde{\lambda} = 0.07$	$\tilde{\theta} = 0.1557$ $\tilde{\pi} = 0.5505$	$\tilde{\alpha} = 0.153$ $\tilde{\pi}_0 = 0.9351924$
المتوسط باستخدام تقديرات العزوم	0.07	0.07	0.07
التباين باستخدام تقديرات العزوم	0.07	0.076	0.076
المعلمات المقدرة بطريقة الامكان الأعظم	$\hat{\lambda} = 0.07$	$\hat{\theta} = 0.148$ $\hat{\pi} = 0.52703$	$\hat{\alpha} = 0.144$ $\hat{\pi}_0 = 0.93491915$
المتوسط باستخدام تقديرات الامكان الأعظم	0.07	0.07	0.07
التباين باستخدام تقديرات الامكان الأعظم	0.07	0.075	0.075

وكما يتضح من جدول (٨) أن توزيع بواسون لا يعكس التباين الفعلى فى البيانات حيث يفترض تساوى المتوسط مع التباين، ونلاحظ أيضا أن الفرق بين توزيع بواسون ذى الأصفار الزائدة وتوزيع هاردل بواسون ضئيل جدا سواء فى تقديرات العزوم أو فى تقديرات الامكان الأعظم، ونلاحظ التقارب الشديد بين المتوسط والتباين لهذين التوزيعين وبين المتوسط والتباين الفعلى للبيانات الموضح فى جدول (٢).

(٤-٦) الاختيار بين التوزيعات المختلفة المستخدمة فى الدراسة

نلاحظ أنه عند استخدام اختبار جودة التوفيق لكولومجروف سمرنوف كانت نتائج الثلاثة توزيعات المستخدمة هى قبول الفرض العدمى الذى يقضى بتبعية البيانات للتوزيع محل الاختبار. لذلك يجب استخدام معايير أخرى للاختيار بين التوزيعات المختلفة لنمذجة بيانات تكرار المطالبات. ويتم استخدام معيارى ((AIC), (BIC)) للاختيار بين التوزيعات المختلفة.

جدول (٩) حساب AIC ، BIC للتوزيعات المستخدمة فى الدراسة

التوزيع	AIC	BIC
توزيع بواسون (طريقة العزوم وطريقة الامكان الأعظم)	**3884.36**	**3891.28**
توزيع بواسون ذى الأصفار الزائدة (طريقة العزوم)	**3870.56**	**3884.40**
توزيع بواسون ذى الأصفار الزائدة (طريقة الامكان الأعظم)	**3870.36**	**3884.20**
توزيع هاردل بواسون (طريقة العزوم)	**3870.47**	**3884.31**
توزيع هاردل بواسون (طريقة الامكان الأعظم)	**3870.34**	**3884.18**

تم حساب قيم المعايير بهذا الجدول باستخدام صيغ لوغاريتم دالة الامكان الأعظم للثلاثة توزيعات الموضحة فى المعادلات أرقام ((32) ,(15) ,(5)) والمقاييس الاحصائية الموجودة فى جدول (٢)، وتقديرات المعلمات الموجودة فى جدول (٨)، وصيغ المعايير الموضحة فى المعادلتين ((40) ,(39)).

وكما نلاحظ من جدول (٩) أن توزيع بواسون ذى الأصفار الزائدة وتوزيع هاردل بواسون يعطيا نتائج أفضل من توزيع بواسون حيث إن قيم المعايير الخاصة بهما أقل من قيم المعايير الخاصة بتوزيع بواسون، كذلك نلاحظ أن تقديرات الامكان الأعظم تعطى نتائج أفضل قليلا من تقديرات طريقة العزوم. ونلاحظ أيضا أن توزيع هاردل بواسون أفضل قليلا من توزيع بواسون ذى الأصفار الزائدة.

ونظرا لأن الباحث لم يستطع معرفة منشأ الأصفار فى البيانات محل الدراسة لذلك يتم اختيار التوزيع الذى له أقل قيمة وفقا لمعيارى (AIC)، (BIC) ، لذلك يتم اختيار توزيع هاردل بواسون لنمذجة البيانات محل الدراسة، واستخدام مقدرات الامكان الأعظم لأنها تعطى تقديرات أفضل من طريقة العزوم.

النتائج والتوصيات

أولا: النتائج

بعد القيام بهذا البحث توصل الباحث إلى النتائج التالية:

(١) ان استخدام مؤشر الأصفار الزائدة واختبار (Score test) يكشف وجود الأصفار الزائدة فى البيانات، فوجود الأصفار الكثيرة فى البيانات لا يعنى دائما وجود أصفار زائدة ، لذلك يجب اكتشافها بإجراء الاختبارات.

(٢) إن وجود أصفار زائدة فى البيانات يعنى تشتت أكثر للبيانات (Over-dispersion) ويعنى أيضا أن توزيع بواسون لا يكون هو الأفضل لنمذجة البيانات.

(٣) توزيع بواسون ذى الأصفار الزائدة وتوزيع هاردل بواسون أفضل من توزيع بواسون فى نمذجة بيانات البحث، حيث يمكنه استيعاب التشتت الزائد فى البيانات عن طريق المعلمة الاضافية الموجودة فى كل منهما.

(٤) توزيع هاردل بواسون أفضل قليلا من توزيع بواسون ذى الأصفار الزائدة فى نمذجة بيانات البحث.

(٥) تقدير معلمات توزيع بواسون ذى الأصفار الزائدة وتوزيع هاردل بواسون بطريقة الامكان الأعظم أفضل من طريقة العزوم وفقا لمعيارى (AIC), (BIC).

ثانيا: التوصيات

فى نهاية هذا البحث يوصى الباحث بالتوصيات الآتية:

(١) استخدام اختبارات الأصفار الزائدة مثل مؤشر الأصفار الزائدة، اختبار (Score test) لاكتشاف الأصفار الزائدة الموجودة فى تكرار المطالبات عند نمذجة تكرار المطالبات.

(٢) عندما يكون القائم بعملية النمذجة على دراية بمنشأ الأصفار يتم استخدام توزيع ذى أصفار زائدة للتوزيع الملائم للبيانات عندما يوجد بالبيانات أصفار حقيقية وأصفار زائدة مثل تلك الناشئة عن خصم عدم المطالبة وحدود التحمل الموجودة فى الوثيقة. ويتم استخدام توزيع هاردل للتوزيع الملائم للبيانات عندما تكون كل الأصفار حقيقية مثل الأصفار الناشئة عن انتقاء الشركة للعمليات التأمينية التى تقبلها والناتجة أيضا عن تحسن خبرة المطالبات لدى الشركة نتيجة حرص المؤمن لهم على تقليل عدد الحوادث.

(٣) عندما يصعب التوصل لمنشأ الأصفار يتم الاختيار بين التوزيع ذى الأصفار الزائدة وتوزيع هاردل للتوزيع الملائم للبيانات باستخدام المعايير الاحصائية مثل معيارى (AIC)، (BIC).

قائمة المراجع

1. Ashour, S. K. & Salem, S. A. (2009). Probability Distributions. I.S.S.R., Cairo University.
2. Beckett, S., Jee, J., Ncube, T., Pompilus, S., Washington, Q., Singh, A., & Pal, N. (2014). Zero-inflated Poisson (ZIP) distribution: parameter estimation and applications to model data from natural calamities. *Involve, a Journal of Mathematics*, *7*(6), 751-767.
3. Guillén, M., Nielsen, J. P., Ayuso, M., & Pérez Marín, A. M. (2018). Exposure to risk increases the excess of zero accident claims frequency in automobile insurance. *UB Economics–Working Papers, 2018, IR18/10*.
4. Hofstetter, H., Dusseldorp, E., Zeileis, A., & Schuller, A. A. (2016). Modeling caries experience: advantages of the use of the hurdle model. *Caries research*, *50*(6), 517-526.
5. Kusuma, R. D., & Purwono, Y. (2019, March). Zero-inflated Poisson regression analysis on frequency of health insurance claim PT. XYZ. In *12th International Conference on Business and Management Research (ICBMR 2018)*. Atlantis Press.
6. Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*(1), 1-14.
7. Loeys, T., Moerkerke, B., De Smet, O., & Buysse, A. (2012). The analysis of zero-inflated count data: Beyond zero-inflated Poisson regression. *British Journal of Mathematical and Statistical Psychology*, *65*(1), 163-180.
8. Maydeu-Olivares, A., & Garcla-Forero, C. (2010). Goodness of fit testing. *International Encyclopedia of Education*, Vol. 7, 190-196.
9. Mouatassim, Y., & Ezzahid, E. H. (2012). Poisson regression and zero-inflated Poisson regression: application to private health insurance data. *European actuarial journal*, *2*(2), 187-204.

10. Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of econometrics*, *33*(3), 341-365.
11. Puig, P., & Valero, J. (2006). Count data distributions: some characterizations with applications. *Journal of the American Statistical Association*, 101(473), 332-340.
12. Sakthivel, K. M., & Rajitha, C. S. (2018), Estimation of zero-inflation parameter in zero-inflated Poisson model. International Journal of Mathematics Trends and Technology (IJMTT), Vol. 56, No. 2,
13. Sarul, L. S., & Sahin, S. (2015). An application of claim frequency data using zero inflated and hurdle models in general insurance. *Journal of Business Economics and Finance; Vol 4, No 4 (2015)*.
14. Van den Broek, J. (1995). A score test for zero inflation in a Poisson distribution. Biometrics, 738-743.
15. Wagh, Y. S., & Kamalja, K. K. (2018). Zero-inflated models and estimation in zero-inflated Poisson distribution. *Communications in Statistics-Simulation and Computation*, *47*(8), 2248-2265.
16. Wolny-Dominiak, A. (2013). Zero-inflated claim count modeling and testing–a case study. *Econometric*, (39), 144-151.
17. Yip, K. C., & Yau, K. K. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*, *36*(2), 153-163.
18. Data site: Singapore Automobile Claims

https://instruction.bus.wisc.edu/jfrees/jfreesbooks/Regression%20Modeling/BookWebDec2010/data.html