

Research paper

Modélisation du Risque d'un Portefeuille d'Assurance-Crédit via l'Application de la Simulation de Monte Carlo

Youssef Diouri^{1*}, Jamal Zahi¹

¹ Laboratoire de Modélisation Mathématique et de Calcul Economique, Faculté d'Economie et de Gestion, Université Hassan 1^{er}, Settat, Maroc

PAPER INFO

Paper History

Received 17 October
2023

Accepted 11 August
2024

Keywords

Monte Carlo

Lof Normale

Loi de Poisson

Engagement

Portefeuille

Mutualisation

Assurance Crédit

ABSTRACT

L'assurance constitue un pilier majeur de l'économie en fournissant une protection contre les événements imprévus, permettant ainsi aux individus et aux entreprises de se prémunir contre des pertes financières potentiellement dévastatrices. Cette étude se concentre sur le produit d'assurance-crédit immobilier, qui garantit les banques contre les défauts de paiement de leurs clients à la suite du risque de décès. Ce type d'assurance garantit des montants importants, exposant ainsi l'assureur à des risques considérables, notamment en cas de variation du ratio de sinistralité. Par conséquent, l'objectif principal de cette recherche est d'évaluer avec précision le risque actuariel associé au portefeuille d'assurance-crédit immobilier afin de préserver sa viabilité financière. Dans le cadre de cette étude, nous nous focalisons sur la modélisation du nombre de sinistres annuels et de la charge financière qui en découle, en utilisant diverses lois de modélisation statistique pour quantifier le risque global auquel l'assureur pourrait être confronté, et anticiper ainsi ses engagements envers les assurés grâce à l'application de la simulation de Monte Carlo, reconnue comme l'une des méthodes les plus efficaces pour résoudre des problèmes numériques complexes. Par conséquent, nous proposons un modèle de quantification des risques, qui représente un outil de prise de décision essentiel pour les compagnies d'assurance.

1. Introduction

L'assurance joue un rôle essentiel dans l'économie, offrant une protection contre les éventuels aléas et permettant aux individus et aux entreprises de se prémunir contre des pertes financières potentiellement dévastatrices. Les produits d'assurance couvrent une gamme étendue de risques, allant des accidents de la route aux catastrophes naturelles, en passant par les problèmes de santé et la responsabilité civile.

L'engagement de l'assureur envers l'assuré consiste à le couvrir financièrement lors de la survenance d'un événement spécifié dans le contrat, moyennant le paiement d'une prime d'assurance. Cette prime est calculée en fonction de l'espérance de survenance du sinistre, avec pour objectif de rester proche de la moyenne statistique. Ainsi, le principe de mutualisation est appliqué, permettant à l'assureur de souscrire un grand nombre de contrats pour indemniser les risques sinistrés.

Cependant, les événements aléatoires couverts par une assurance peuvent parfois entraîner une sous-estimation de la charge probable à indemniser, ce qui peut avoir des conséquences financières importantes et mettre en péril la solvabilité de l'assureur. Il est donc essentiel que l'assureur dispose de réserves suffisantes pour faire face à ses engagements financiers envers les assurés, non seulement pour se conformer aux exigences réglementaires, mais aussi pour éviter tout risque de faillite en cas de sinistres exceptionnels.

Ce travail de recherche se focalise sur le produit d'assurance-crédit, où l'assureur garantit les banques contre le défaut de paiement de leurs clients suite au décès. Étant donné que ce type d'assurance garantit des montants importants, l'exposition au risque de l'assureur peut être très élevée, surtout en cas d'augmentation inattendue des sinistres.

L'objectif principal de cette étude est d'évaluer précisément le risque actuariel associé au portefeuille d'assurance-crédit afin de garantir sa stabilité financière à long terme. Pour ce faire, nous allons modéliser le nombre de sinistres annuels et la charge associée, en utilisant différentes lois statistiques pour recourir par la suite à l'application de la simulation de Monte Carlo qui est l'une des méthodes les plus robustes pour résoudre des problèmes numériques complexes.

2. Méthodologie de recherche

2.1 L'assurance crédit-immobilier

Le produit concerné par notre étude est le produit d'Assurance-Crédit immobilier. Il s'agit d'un produit très courant dans les compagnies d'assurance, surtout chez les bancassureurs qui gèrent à la fois la souscription de la police d'assurance et la souscription de l'emprunt. Dans ce qui suit, nous allons décrire l'assurance-crédit, ses garanties ainsi que ses caractéristiques.

- **L'assurance-Crédit :**

L'assurance-crédit est un contrat, nécessaire lors de la souscription d'un prêt. Il est temporaire, c'est-à-dire limité à la durée du crédit et permet de protéger le créancier contre différents risques de défaillance de l'emprunteur pouvant l'empêcher de rembourser ses échéances.

Ce type d'assurance prend généralement la forme d'un contrat collectif, dans la mesure où sa souscription se fait au niveau des établissements de crédit auprès d'une compagnie d'assurance en faveur des emprunteurs ayant la qualité d'adhérents.

- **Garanties : IAD/IPT**

Les garanties offertes en assurance décès emprunteur, concernent principalement le risque de décès d'une part et le risque d'invalidité d'une autre part.

En effet, en cas de décès de l'assuré, l'assureur prendra à sa charge le versement du capital restant dû à la contractante.

2.2 Le modèle collectif du risque

Nous définissons la variable aléatoire S comme étant le montant total des réclamations découlant d'un risque sur une année.

Soit la variable aléatoire N qui représente le nombre de réclamations du risque au cours de cette année.

Et soit la variable aléatoire X_i qui représente le montant de la i ème réclamation. Le montant total des réclamations agrégées est simplement la somme des montants de réclamation individuels. Par conséquent, nous pouvons écrire [2] :

$$S = \sum_{i=1}^N X_i$$

Nous faisons deux hypothèses importantes :

Tout d'abord, nous supposons que $\{X_i\}_{i=1, \dots, \infty}$ est une séquence de variables aléatoires indépendantes et identiquement distribuées.

Deuxièmement, nous supposons que la variable aléatoire N est indépendante de $\{X_i\}_{i=1, \dots, \infty}$.

Dans ce cas, nous avons [3] :

$$E(S) = E(X) * E(N)$$

2.3 La distribution du modèle

L'objectif de cette partie est de construire un modèle probabiliste pour représenter S , la charge globale.

Le modèle nécessiterait un composant qui modélise le nombre de sinistres et un autre qui modélise les montants de ces sinistres. [2]

Pour ce qui est de la fréquence : les deux modèles les plus connus étant : la loi de poisson et la loi binomial négative. On cherchera le modèle le plus adapté dans notre cas via des tests de validation.

Dans la suite dans la partie pratique on va opter pour la loi de poisson qui a comme probabilité :

$$p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

L'espérance étant $\mu = \lambda$.

Le choix de la distribution du montant d'un sinistre est crucial dans la modélisation du risque supporté par l'assureur. En actuariat, pour modéliser le comportement aléatoire de la sévérité, on a généralement recours à une loi continue avec un support appartenant à \mathbb{R}^+ .

Pour ce qui est de la sévérité : les lois continues les plus utilisés sont : log normale, exponentielle, pareto, gamma et weibull. On va démontrer dans la partie pratique que la loi de log normale est la plus adaptée.

Dans ce cas, si X suit une distribution lognormale $LN(\mu, \sigma)$, alors $\ln(X)$ suit une distribution normale $N(\mu, \sigma^2)$, avec $\mu = E(X)$. Pour ce qui est de la fonction de densité, on a [5]:

$$f(x|\mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$

- $X > 0$ est la variable aléatoire,
- μ la moyenne de la distribution de la variable logarithmique,
- σ l'écart type de la distribution de la variable logarithmique.

2.4 Estimation des paramètres

Pour l'estimation paramétrique, nous allons nous intéresser à la méthode la plus utilisée à savoir la méthode du maximum de vraisemblance. [6]

Soient X_1, \dots, X_n n variable aléatoire indépendantes et identiquement distribuées de fonction de densité $f(x, \theta)$. On définit la fonction de vraisemblance comme suit :

$$L(\theta) = \prod_{i=1}^n f(x_i, \theta)$$

L'estimateur du maximum de vraisemblance est défini comme la statistique $\hat{\theta} = T(X_1 \dots X_n)$ telle que :

$$L(\theta^{MV}) = \text{Max}_{\theta} L(\theta)$$

Il s'agit donc de chercher la valeur de θ fonction des observations $(x_1 \dots x_n)$ qui assure la plus grande probabilité d'avoir ces observations.

Dans le cas où la fonction de vraisemblance est continue et deux fois dérivable par rapport au paramètre θ , l'estimateur θ^{MV} n'est autre que la solution du système:

$$\begin{aligned} - & \left(\frac{\partial L}{\partial \theta} \right)_{\theta^{MV}} = 0 \\ - & \left(\frac{\partial^2 L}{\partial \theta^2} \right)_{\theta^{MV}} < 0 \end{aligned}$$

2.5 Test d'ajustement

Afin de choisir la loi qui ajuste au mieux les données, nous utilisons les tests d'hypothèses permettant de prendre une décision au détriment d'une autre.

2.5.1 Détermination de la loi de fréquence

Pour choisir entre une loi Poisson ou Binomiale négative, il est naturel de penser à comparer entre la fréquence et variance empirique des données puisqu'en cas d'égalité entre ces deux moments, nous pensons directement à l'équi-dispersion qui constitue la caractéristique principale d'une loi Poisson.

Dans ce sens, deux tests d'hypothèses seront analysés: le test de sur-dispersion basé sur la statistique liée à l'indice de dispersion de Fisher et le test du rapport de vraisemblance.

- **Test de sur dispersion**

L'hypothèse testée est $H_0 : E(N) = V(N)$ Contre $V(N) > E(N)$.

En d'autres termes, l'existence de sur-dispersion ou son absence.

La statistique T pour estimer l'indice de dispersion qui est égale à $\frac{V(N)}{E(N)}$ est :

$$T = \frac{(n-1)\bar{S}^2}{\bar{X}}$$

Avec \bar{X} et \bar{S}^2 respectivement la moyenne et la variance empirique de la loi de fréquence N .

Sous l'hypothèse H_0 , Hoel (1943) a montré que la statistique nT , connue sous le nom d'indice de dispersion de Fisher, est asymptotiquement distribuée comme une loi de khi 2 à $n - 1$ degrés de liberté avec n le nombre d'observations de N .

- **Test du rapport de vraisemblance**

Le test de rapport de vraisemblance permet de tester si l'échantillon de N suit une distribution donnée contre une distribution alternative. Dans cette étude, on s'intéressera à tester l'hypothèse suivante :

$$H_0 : N \text{ suit une loi poisson} \quad \text{vs} \quad H_1 : N \text{ suit une loi binomiale négative}$$

La statistique du rapport de vraisemblance est donnée par : $\mathbf{T} = 2(\ln(\mathbf{L}_1) - \ln(\mathbf{L}_0))$.

Avec :

- L_0 : La vraisemblance de la loi de poisson
- L_1 : La vraisemblance de la binomiale négative

La statistique T tend asymptotiquement en loi vers une loi du Khi-deux à un degré de liberté.

2.5.2 Détermination de la loi du coût de sinistre

Puisqu' il s'agit des lois continues, nous testons l'ajustement par le biais des tests suivants : Kolmogorov – Smirnov et Anderson – Darling.

- **Kolmogorov – Smirnov Test**

Le test d'ajustement de Kolmogorov – Smirnov est un test non paramétrique basé sur la distance maximale entre une fonction de répartition théorique et la fonction de répartition empirique de l'échantillon. [7]

Il permet de tester l'hypothèse H_0 selon laquelle les données observées sont engendrées par une loi de probabilité théorique considérée comme étant un modèle convenable.

Nous considérons ainsi une variable aléatoire X de fonction de répartition F_x , que nous voulons comparer à une fonction de répartition théorique F_0 continue.

Nous testons alors :

$$H_0: F_x(x) = F_0(x) \forall x \quad \text{Contre} \quad \exists x F_x(x) \neq F_0(x)$$

Soit $\{X_1, X_2 \dots X_n\}$ un n – échantillon de X , la fonction de répartition empirique F_n associée à cet échantillon est :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{]-\infty, x]}(X_i)$$

$F_n(x)$ la proportion des observations dont la valeur est inférieure ou égale à x .

L'écart entre les valeurs observées et les valeurs théoriques du modèle déduites de la fonction de répartition F_0 peut donc être mesuré par la variable aléatoire :

$$D_n = \sup_{x \in x} |F_n(x) - F_0(x)|$$

La statistique D_n tend asymptotiquement vers une loi du Khi-deux à $n-1$ degré de liberté.

- **Test d'Anderson-Darling**

Le test d'Anderson-Darling est basé sur le calcul d'une distance entre la fonction de distribution empirique et la fonction ajustée définie comme suit :

$$A_n^2 = n * \int_{-\infty}^{+\infty} [F_x(x) - F_0(x)]^2 * \frac{1}{F_0(x) * [1 - F_0(x)]} * dF(x)$$

Bien que le test d'Anderson Darling et le test de Kolmogorov ont la même application, la différence entre eux réside dans le fait que pour le deuxième test seul l'écart maximum entre la distribution empirique et la distribution d'ajustement compte, alors que l'indicateur d'écart du premier test capte mieux l'ensemble des données dans la mesure où toute la somme des écarts intervient. On peut donc déduire que le test de Kolmogorov est beaucoup plus sensible à l'existence de points aberrants dans un échantillon que le test d'Anderson Darling.

2.6 Simulation Monte Carlo

Les techniques de simulation sont fort utiles lorsque la complexité du modèle envisagé rend extrêmement difficile toute approche analytique [8].

Une fois avoir générer les variables aléatoires par une loi donnée, la méthode de Monte Carlo consiste à approcher $E(X)$ par la moyenne empirique des simulations de la variable aléatoire X : (La loi des grands nombres).

$$\bar{X} = \frac{1}{\text{Nombre - simulations}} \sum_{i=1}^{\text{Nombre-simulations}} X_i \xrightarrow{\text{Nombre-simulations} \rightarrow \infty} E(X)$$

3. Résultats et discussion

L'objectif de ce travail est d'évaluer la charge globale probable que l'assureur doit anticiper relative au portefeuille d'assurance-crédit immobilier ce qui coïncide avec les fonds que doit détenir l'assureur pour honorer ses engagements futurs sans menacer sa stabilité financière. Ce portefeuille regroupe 2365 assurés.

Pour quantifier le risque, il est nécessaire d'analyser la sinistralité passée afin de modéliser celle future. Dans ce cadre on se basera sur une base de données sinistre qui est dotée de ses informations suivantes historique de 8 ans de 2015 à 2022 qui intègre les données suivantes : date de survenance, l'identifiant du sinistre, montant du sinistre

Cependant Pour que les données historiques soient comparables entre elles quel que soit leur année de survenance, nous procéderons à une opération qui s'appelle mise en situation « As If » et consiste à :

- ❖ Actualiser le montant du sinistre en se basant sur le taux d'inflation entre l'année de survenance et l'année de cotation soit 2023. On retient les taux annuels suivants enregistrés au Maroc :

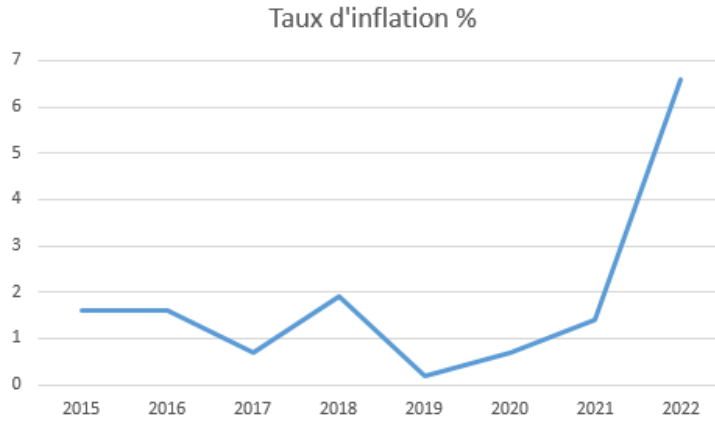


Fig. 1. Taux d'inflation annuels

En se basant sur les taux d'inflations représentés ci-dessus, nous revalorisons chaque sinistre X de la manière suivante :

$$X^{As\ if} = X * \prod_{K=J}^n (1 + r_k)$$

Avec :

- n : L'année de cotation-1 (2022 pour notre cas)
- r_k : L'indice d'inflation de l'année k
- J : L'année de survenance

- ❖ Actualiser le nombre de sinistres annuels en supposant l'existence d'une relation de proportionnalité entre le nombre de sinistres survenu annuellement et le nombre d'adhérents par exercice, d'où l'expression suivante:

$$n_k^{As\ if} = n_{adh2023} * \frac{n_k}{n_{adhe_k}}$$

Avec :

- n_k : Nombre de sinistres relatif à l'année k
- $n_k^{As\ if}$: Nombre de sinistres As if
- n_{adhe_k} : Nombre d'adhérents de l'année k

Le graphe ci-après représente les fréquences et les charges totales revalorisées entre 2015 et 2022 :

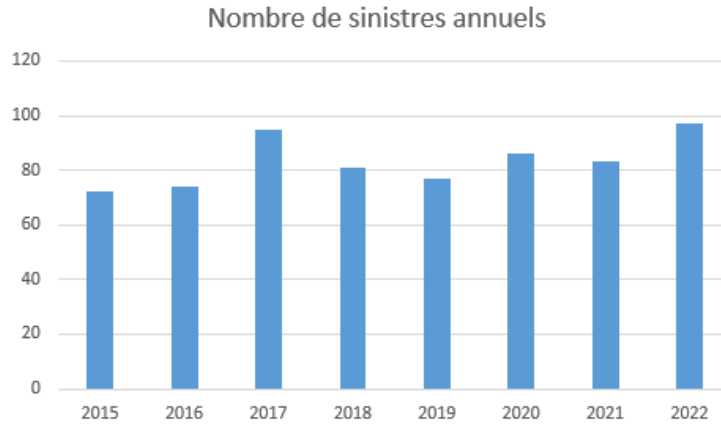


Fig. 2. Nombre de sinistres annuels

Maintenant que nous avons construit notre vecteur de fréquence As If et des charges As If, l'étape suivante consiste à chercher un modèle qui ajuste bien les données.

3.1 Estimation du modèle

Dans ce qui suit, on se basera sur le modèle collectif du risque :

$$S = \sum_{i=1}^N X_i$$

Avec : N le nombre de sinistres annuels

X_i : Le coût As if de l'i-ème sinistre

S : La charge totale annuelle

En se basant sur les notions théoriques précédemment expliquées, nous allons modéliser la charge annuelle avant de procéder à la simulation de Monte Carlo. Pour cela, on commencera par modéliser la charge du sinistre X_i et par la suite le nombre N.

3.2 Détermination de la loi de fréquence

Les lois les plus utilisées pour modéliser la distribution de la variable aléatoire de la fréquence N sont la loi de Poisson et la loi Binomiale Négative. Pour choisir, on opte pour le test de dispersion.

TABLE I. TEST DU RAPPORT DE DISPERSION

Surdispersion	\bar{X}	\bar{S}^2	IF
Indice Fisher (IF)	83,13	84,41	7,11

L'indice de Fisher est à comparer avec le quantile 95% d'une loi Khi2 à 7 degrés de liberté qui est égal à 14,07. On constate que IF est 7,11 donc inférieur à 14,07 et par suite on rejette l'hypothèse de l'existence d'une surdispersion des données et par suite la loi de poisson est la plus adaptée.

Estimation des paramètres :

Pour estimer les paramètres, nous utilisons la méthode de maximum de vraisemblance. On obtient :

$$\lambda = 83,13$$

Pour ce qui est du nombre de sinistres on confirme le choix du model par le test du rapport de vraisemblance étudié ci-après :

TABLE II. TEST DU RAPPORT DE VRAISEMBLANCE

Modèle	$\ln(L_0)$	$\ln(L_1)$	RV
	- 33,78	- 20,42	28,1

Avec le test du rapport de vraisemblance, la loi de poisson est acceptée puisque RV est supérieur au quantile 95% d'une loi χ^2 à 1ddl qui est égal à 3,84 ce qui approuve la qualité d'ajustement.

3.3 Détermination de la loi du coût de sinistre

Pour la modélisation de la charge du sinistre on va utiliser la loi lognormal qui est largement utilisée dans la modélisation des variables continues en actuariat. Dans la suite, nous allons vérifier si l'ajustement par la loi log normal est effectivement approuvé par les tests statistiques.

Pour ce qui est des paramètres de la loi Log Normale estimés par la méthode de maximum de vraisemblance. On obtient $\mu = 12.39$ et $\sigma = 0.92$

Pour valider la qualité de l'ajustement, on aura recours aux tests de Kolmogorov et de celui d'Anderson Darling pour le coût de sinistre.

On obtient les résultats suivants :

- **Test de Kolmogorov :**

TABLE III. TEST DU RAPPORT DE KOLMOGOROV

Modèle	Dn	C	P-value
Log Normal	0,053	0,086	0,063

Au seuil 5%, nous acceptons l'hypothèse sous laquelle :

- Le coût par sinistre suit une LN (12.39,0.92)

En effet, la p-value obtenue est supérieure à 0,05.

- **Test d'Anderson Darling**

TABLE IV. TEST D'ANDERSON DARLING

	A	P-Value
Test Anderson	0,23	0,079

Le test d'Anderson Darling confirme aussi la qualité d'ajustement puisque la p-value associée à chaque test est supérieure à 0,05.

3.4 Calcul du risque global à travers la simulation Monte Carlo

Comme nous connaissons les lois de fréquence et de la charge, nous pouvons envisager l'application de la méthode par simulation.

Pour pouvoir appliquer cette méthode, nous avons construit un algorithme de simulation que nous avons implémenté. Le paragraphe suivant présentera les étapes de l'algorithme.

Pour appliquer Monte Carlo, il va falloir simuler un grand nombre de valeurs de la variable aléatoire S . Comme celle-ci est issue d'une loi composée, nous suivrons les étapes suivantes :

- Fixer le nombre de simulations J , pour notre cas $J = 100000$
- Créer un vecteur N contenant J simulations de même loi que N : $N = (n_1 \dots n_j \dots n_j)$ avec n_j la j -ième réalisation de la simulation par la loi N
- Pour $j = 1$ à J :
 - Simuler n_j variables $x_1^{n_j}, x_2^{n_j} \dots \dots x_{n_j}^{n_j}$ issues de la variable aléatoire $LN(12,39 ; 0,92)$.
- Pour $i = 1$ à n_j
 - $s_j = \sum_{i=1}^{n_j} x_i^{n_j}$
- Calculer la moyenne du vecteur $\{s_j\}_{1 \leq j \leq J}$ ce qui correspond à $E(S)$

En appliquant l'algorithme on en déduit que $E(S)$ est égale à 1,97 Millions \$. D'où la force de cette technique pour approcher d'espérances mathématiques complexes.

4. Conclusion

Dans cette étude, nous avons évalué le risque actuariel associé à un portefeuille d'assurance-crédit en modélisant le nombre annuel de sinistres et le coût moyen. Les distributions de Poisson et de Log-Normale se sont révélées être les modèles les plus adaptés pour la fréquence et le coût des sinistres, respectivement. Ces modèles ont été validés à l'aide de divers tests d'adéquation, confirmant leur adéquation avec les données historiques.

De plus, en appliquant la simulation de Monte Carlo, nous avons pu approfondir notre compréhension du risque en simulant divers scénarios et en examinant les résultats pour obtenir une estimation plus précise des charges financières probables. Cette méthode nous a conduit à l'estimation du coût total moyen des sinistres, évalué à environ 1,97 million de dollars.

La capacité de prédire de manière fiable le risque et les charges financières est cruciale pour la gestion du portefeuille d'assurance, permettant à la compagnie d'assurance de maintenir la solvabilité et de se prémunir contre les événements défavorables significatifs. Cette étude démontre l'importance des méthodes statistiques avancées dans l'évaluation du risque actuariel et dans la prise de décisions éclairées dans le domaine de l'assurance.

References

- [1] A. Melnikov, *Risk Analysis in Finance and Insurance*, Taylor & Francis, 2003.
- [2] D. Dickson, *Insurance Risk and Ruin*, Cambridge University Press, 2016.
- [3] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit, *Modern Actuarial Risk Theory*, Springer, 2008.
- [4] M. Denuit and A. Charpentier, *Mathématiques de l'Assurance Non-Vie*, Economica, 2004.
- [5] M. Lefebvre, *Probabilités, Statistiques et Applications*, Presses Internationales Polytechnique, 2011.
- [6] P. Olofsson, *Probability, Statistics, and Stochastic Processes*, John Wiley & Sons, 2005.
- [7] G. K. Kanji, *100 Statistical Tests*, 3rd ed., SAGE Publications, 2006.
- [8] J. Mun, *Modeling Risk: Applying Monte Carlo Risk Simulation, Strategic Real Options, Stochastic Forecasting, and Portfolio Optimization*, 2nd ed., John Wiley & Sons, 2010.
- [9] N. Ellili, H. Nobanee, L. Alsaïari, H. Shanti, B. Hillebrand, N. Hassanain, and L. Elfout, "The Applications of Big Data in the Insurance Industry: A Bibliometric and Systematic Review of Relevant Literature," *The Journal of Finance and Data Science*, vol. 9, 2023.
- [10] V. Peleckienė and K. Peleckis, "Solvency II Assumptions for Increasing the International Competitiveness of EU Insurance Industry," *Procedia - Social and Behavioral Sciences*, vol. 110, pp. 822-831, 2014.
- [11] C.-Y. Chiu, T.-S. Dai, Y.-D. Lyuu, L.-C. Liu, and Y.-T. Chen, "Option Pricing with the Control Variate Technique Beyond Monte Carlo Simulation," *The North American Journal of Economics and Finance*, vol. 62, 2022, Art. no. 101772.