Revue d'Economie et de Statistique Appliquée
Volume 23, number 01, June 2026
ISSN : 1112-234X EISSN : 2600-6642
Pages : 180-196

MODELING THE SEVERITY OF MATERIAL CLAIMS IN MOTOR INSURANCE: AN APPROACH USING A GENERALIZED LINEAR MODEL WITH GAMMA DISTRIBUTION AND LOGARITHMIC LINK

MODÉLISATION DE LA SÉVÉRITÉ DES SINISTRES MATÉRIELS EN ASSURANCE AUTOMOBILE : UNE APPROCHE PAR UN MODÈLE LINÉAIRE GÉNÉRALISÉ DE DISTRIBUTION GAMMA ET LIEN LOGARITHMIQUE

*** Moussa DJEMMALI:**
École Nationale Supérieure de Statistique et Économie Appliquée (ENSSEA) – Koléa, Tipaza - Laboratoire de Modélisation des Phénomènes Stochastiques (LAMOPS)
Email : djemmali.moussa@enssea.edu.dz

Abdelouahab LATRECHE:
École Nationale Supérieure de Statistique et Économie Appliquée (ENSSEA) – Koléa, Tipaza- Laboratoire de Modélisation des Phénomènes Stochastiques (LAMOPS)
Email : latreche.abdelouahab@enssea.edu.dz

Reçu le: 06/05/2026 **Accepté le :** 20/06/2026 **Publication en ligne le**: 28/06/2026

ABSTRACT:
Modeling claim severity represents a major issue in automobile insurance. In Algeria, classical approaches remain dominant despite their inadequacy for positive and asymmetric cost data. This study analyzes the determinants of material claim costs within the Société Nationale d'Assurance (SAA), based on a sample of 827 observations from the first quarter of 2025. A Gamma-Log generalized linear model is used to account for heteroskedasticity and data asymmetry. The results show that the driver's age, the vehicle's value, and its power significantly influence claim severity. Goodness-of-fit tests confirm the superiority of the Gamma-Log model compared to classical regression.

Keywords: Claim severity, Generalized Linear Model, Gamma distribution, automobile insurance, actuarial pricing.
JEL Classification : G22

RÉSUMÉ :
La modélisation de la sévérité des sinistres constitue un enjeu majeur en assurance automobile. En Algérie, les approches classiques restent dominantes malgré leur inadéquation aux données de coûts positives et asymétriques. Cette étude analyse les déterminants du coût des sinistres matériels au sein de la Société Nationale d'Assurance (SAA), à partir d'un échantillon de 827 observations au premier trimestre 2025. Un modèle linéaire généralisé Gamma-Log est mobilisé afin de tenir compte de

* Auteur correspondant

l'hétéroscédasticité et de l'asymétrie des données. Les résultats montrent que l'âge du conducteur, la valeur et la puissance du véhicule influencent significativement la sévérité. Les tests d'ajustement confirment la supériorité du modèle Gamma-Log par rapport à la régression classique.

Mots Clés: Sévérité des sinistres, Modèle Linéaire Généralisé, distribution Gamma, assurance automobile, tarification actuarielle.

I. INTRODUCTION

L'assurance automobile occupe une position structurante dans le portefeuille des compagnies d'assurance non-vie. À l'échelle internationale, cette branche représente une part déterminante des primes émises et constitue, pour l'assureur, l'une des principales sources de risque financier. En Algérie, le secteur des assurances a enregistré une croissance régulière de son activité automobile, la branche véhicules terrestres à moteur (VTM) représentant à elle seule plus de 55 % du chiffre d'affaires global du secteur (Conseil National des Assurances, 2024). Dans ce contexte, la maîtrise du coût des sinistres matériels constitue un impératif stratégique, non seulement pour la tarification et l'équilibre technique, mais également pour le provisionnement, la solvabilité et la gestion globale du risque.

La modélisation de la sévérité des sinistres, c'est-à-dire du coût individuel de chaque sinistre, s'inscrit dans le cadre plus général de la tarification actuarielle, qui vise à estimer la prime pure comme produit de la fréquence attendue et du coût moyen anticipé (Frees, 2010). Cette décomposition implique que toute imprécision dans l'estimation de la sévérité se répercute directement sur la prime facturée, engendrant soit une sous-tarification menaçante l'équilibre financier de la compagnie, soit une sur-tarification nuisant à sa compétitivité commerciale. La modélisation rigoureuse de cette composante est donc d'une importance opérationnelle considérable.

Or, les données de coût de sinistres présentent des propriétés statistiques spécifiques qui rendent les méthodes classiques de régression linéaire ordinaire (OLS) théoriquement inadaptées. Ces données sont strictement positives, généralement asymétriques à droite, avec une queue de distribution lourde et une variance qui croît avec la moyenne, propriété connue sous le nom d'hétéroscédasticité multiplicative (McCullagh & Nelder, 1989). Dans un tel contexte, le recours à un Modèle Linéaire Généralisé (GLM) de distribution Gamma et de lien logarithmique s'impose comme le cadre de référence de la littérature actuarielle contemporaine (Ohlsson & Johansson, 2010 ; Frees et al., 2014).

La littérature internationale sur la modélisation de la sévérité en assurance automobile est abondante. Des travaux fondateurs comme ceux de McCullagh et Nelder (1989) ont établi les bases théoriques des GLM, tandis que des contributions plus récentes ont démontré empiriquement la pertinence de la distribution Gamma dans des contextes variés (Boucher et al., 2007 ; Henckaerts et al., 2018). En revanche, peu d'études ont été consacrées aux marchés émergents en général, et au marché Algérien en particulier. Les caractéristiques structurelles de ce marché appellent une analyse empirique spécifique. En effet, le vieillissement du parc automobile, la prédominance de la responsabilité civile obligatoire dans l'activité, ainsi que

la forte hétérogénéité des profils de conducteurs modifient les mécanismes de formation du coût des sinistres et rendent nécessaire une étude adaptée à ce contexte.

C'est précisément ce gap que la présente étude vise à combler. À partir d'une base de données originale de 827 sinistres matériels collectée auprès de la Société Nationale d'Assurance (SAA) pour le premier trimestre 2025, l'objectif est double : identifier les principaux facteurs explicatifs du coût des sinistres matériels et évaluer la pertinence empirique du GLM Gamma-Log dans ce contexte institutionnel spécifique. La question de recherche centrale peut se formuler ainsi : quels sont les déterminants statistiquement significatifs de la sévérité des sinistres matériels en assurance automobile, et le GLM Gamma-Log constitue-t-il une spécification adaptée pour les données du marché algérien ?

Sur le plan des contributions, cette recherche génère une triple valeur. D'abord, elle produit un apport méthodologique, car elle met en œuvre un dispositif de modélisation structuré et exhaustif, adapté au contexte algérien. Ensuite, elle propose un apport empirique, en exploitant, pour la première fois, des données individuelles de sinistres issues de la SAA. Enfin, elle offre un apport managérial, dans la mesure où elle fournit des instruments opérationnels utiles à la tarification, au provisionnement et à la segmentation du risque. La suite de l'article s'articule de la manière suivante. La section 2 présente une synthèse de la littérature pertinente. La section 3 décrit les données et le cadre méthodologique. La section 4 expose et discute les résultats. La section 5 conclut et identifie les perspectives de recherche.

II. REVUE DE LA LITTERATURE

La tarification actuarielle en assurance non-vie repose sur un principe fondamental : la prime perçue doit refléter le risque individuel de chaque assuré. Dans le cadre classique, la prime pure est définie comme l'espérance mathématique du coût total des sinistres, que l'on décompose en deux composantes, la fréquence et la sévérité selon la relation $P = E(N) \times E(S)$, où N désigne le nombre de sinistres et S leur coût individuel (Klugman et al., 2012). Cette décomposition, solidement ancrée dans la théorie du risque collectif de Cramér (1930) et formalisée par Bühlmann (1967) dans le cadre des modèles de crédibilité, constitue encore aujourd'hui le socle de la tarification moderne.

La théorie de la mutualisation des risques fonde la légitimité économique de l'assurance : par l'agrégation d'un grand nombre de risques indépendants, la variabilité individuelle se dilue et la compagnie peut faire face à ses engagements avec une probabilité proche de la certitude (Lemaire, 1985). Néanmoins, cette mutualisation n'est actuariellement équitable que si la prime reflète correctement le profil de risque de chaque assuré. L'intégration de variables tarifaires, relatives aux caractéristiques du conducteur et du véhicule, poursuit un objectif d'équité dans la tarification. En effet, ces variables agissent comme des facteurs de segmentation qui structurent le portefeuille en classes de risque plus homogènes, ce qui permet d'ajuster la prime au niveau de risque observé.

La régression linéaire ordinaire (OLS) a longtemps servi d'outil de référence en modélisation actuarielle. Toutefois, son application aux données de coûts révèle des limites importantes. L'hypothèse de normalité des résidus est souvent violée. La variance n'est pas constante lorsque le niveau de coût augmente. De plus, le modèle peut produire des valeurs prédites négatives, ce qui contredit la nature strictement positive des montants de sinistres. Ces contraintes ont conduit, de manière progressive, à privilégier le cadre des Modèles Linéaires Généralisés (GLM), mieux adaptés à la structure et aux mécanismes de ces données, introduit par McCullagh et Nelder (1989). Ce cadre généralise la régression linéaire en permettant à la variable réponse de suivre n'importe quelle distribution de la famille exponentielle, et en reliant l'espérance de cette variable aux prédicteurs via une fonction de lien non linéaire.

Nelder et Wedderburn (1972), dans leur article fondateur, ont démontré que l'estimation des GLM par maximum de vraisemblance pouvait être réalisée efficacement par un algorithme de moindres carrés repondérés itérativement (IRLS), rendant ces modèles opérationnellement accessibles. Dans le contexte de l'assurance automobile, Brockman et Wright (1992) ont été parmi les premiers à promouvoir l'adoption des GLM dans la pratique tarifaire, en montrant leur supériorité empirique sur les méthodes traditionnelles d'analyse par cases. Depuis lors, les GLM sont devenus le standard de la profession actuarielle (Ohlsson & Johansson, 2010).

Dans la littérature actuarielle contemporaine, la modélisation de la fréquence et de la sévérité est généralement conduite séparément, chaque composante faisant l'objet d'une spécification dédiée. Pour la fréquence, les distributions de Poisson et binomiale négative sont les plus couramment utilisées (Boucher et al., 2007). Pour la modélisation de la sévérité, la distribution Gamma s'impose comme une référence. Ce choix repose sur des propriétés statistiques adaptées à la nature des coûts. D'une part, le support est strictement positif. D'autre part, la distribution présente une asymétrie à droite, ce qui reflète la présence de sinistres élevés mais rares. Enfin, la variance évolue en fonction du carré de la moyenne, ce qui traduit un mécanisme de dispersion croissante avec le niveau de coût. Ces caractéristiques sont cohérentes avec les observations empiriques des sinistres automobiles (Frees et al., 2014).

Henckaerts et al. (2018) ont proposé une approche intégrée de la modélisation actuarielle combinant GLM et arbres de décision, appliquée à des données d'assurance automobile belge. Leurs résultats soulignent l'importance de la sélection des variables explicatives et de la détection des interactions non linéaires dans la modélisation de la sévérité. Dans une perspective différente, Antonio et Valdez (2012) ont étudié l'effet des caractéristiques du conducteur sur le coût des sinistres en Belgique, confirmant le rôle prépondérant de l'âge et de l'expérience de conduite. Parmi les variables relatives au véhicule, la valeur assurée et la puissance sont systématiquement identifiées comme des déterminants significatifs du coût des sinistres matériels (Denuit et al., 2019).

La revue des travaux existants fait apparaître deux constats majeurs. D'une part, le GLM Gamma-Log s'est imposé comme la spécification la plus cohérente pour la modélisation de la sévérité des sinistres automobiles, confirmée par de nombreuses applications empiriques sur des marchés développés. D'autre part, les marchés émergents et

le marché Algérien en particulier restent très peu explorés dans la littérature académique. Les rares études disponibles sur l'Afrique du Nord se limitent à des analyses descriptives ou à des modèles de fréquence sans traitement rigoureux de la composante sévérité, ce qui constitue la motivation principale de la présente contribution.

III. DONNEES ET METHODOLOGIE

3.1. Présentation des données

Les données mobilisées dans cette étude proviennent de la Société Nationale d'Assurance (SAA), principale compagnie d'assurance du marché algérien. La base de données constituée est homogène, elle contient exclusivement les coûts individuels des sinistres matériels relatifs aux véhicules particuliers couverts par la garantie responsabilité civile (RC) et, le cas échéant, par des garanties complémentaires. L'échantillon comprend 827 sinistres individuels enregistrés au cours du premier trimestre de l'année 2025. Seuls les sinistres pour lesquels l'ensemble des variables explicatives sont renseignées ont été retenus, garantissant ainsi la complétude de la base et l'absence de valeurs manquantes.

La variable dépendante est le coût du sinistre matériel (coût.sini), exprimé en dinars algériens (DA). Cette variable est continue, strictement positive, et présente une distribution empiriquement asymétrique à droite, avec la présence de quelques sinistres de grande intensité susceptibles d'allonger la queue droite de la distribution. Ces propriétés sont caractéristiques des données de sévérité en assurance non-vie et justifient le recours à un GLM de distribution Gamma (Klugman et al., 2012).

3.2. Description des variables explicatives

Les variables explicatives sont regroupées en deux blocs : les caractéristiques du conducteur et les caractéristiques du véhicule. Pour chaque variable, une discrétisation en classes homogènes a été appliquée afin de capturer d'éventuels effets non linéaires et de faciliter l'interprétation actuarielle des résultats.

Tableau 01 : Description et codage des variables de l'étude.

Variable	Code	Nature	Modalités
Coût du sinistre matériel	coût.sini	Continue (DA)	Variable dépendante
Sexe du conducteur	sexe.cond	Binaire	$\begin{cases} 1 & \text{si Homme} \\ 0 & \text{si Femme} \end{cases}$
Âge du conducteur	âge.cond	Ordinale (4 classes)	$\begin{cases} 1 & \text{si } 18 \leq \text{âge} \leq 33 \\ 2 & \text{si } 33 < \text{âge} \leq 48 \\ 3 & \text{si } 48 < \text{âge} \leq 63 \\ 4 & \text{si âge} > 63 \end{cases}$
Historique des sinistres	hist.sini.cond	Ordinale (3 classes)	$\begin{cases} 1 & \text{si sinistres} \leq 5 \\ 2 & \text{si } 5 < \text{sinistres} \leq 10 \\ 3 & \text{si sinistres} > 10 \end{cases}$
Âge du véhicule	âge.véhi	Ordinale (5 classes)	$\begin{cases} 1 & \text{si âge} \leq 5 \\ 2 & \text{si } 5 < \text{âge} \leq 10 \\ 3 & \text{si } 10 < \text{âge} \leq 15 \\ 4 & \text{si } 15 < \text{âge} \leq 20 \\ 5 & \text{si âge} > 20 \end{cases}$
Valeur du véhicule	valeur.véhi	Ordinale (4 classes)	$\begin{cases} 1 & \text{si } Valeur \leq 1{,}5 \text{ M DA} \\ 2 & \text{si } 1{,}5 < Valeur \leq 3 \text{ M DA} \\ 3 & \text{si } 3 < Valeur \leq 4{,}5 \text{ M DA} \\ 4 & \text{si } Valeur > 4{,}5 \text{ M DA} \end{cases}$
Puissance du véhicule	puissance.véhi	Ordinale (4 classes)	$\begin{cases} 1 & \text{si } Puissance \leq 10 \text{ CV} \\ 2 & \text{si } 10 < Puissance \leq 20 \text{ CV} \\ 3 & \text{si } 20 < Puissance \leq 30 \text{ CV} \\ 4 & \text{si } Puissance > 30 \text{ CV} \end{cases}$

Source : Réalisé par les auteurs à partir des données de la SAA (2025).

3.3. Cadre théorique du GLM Gamma-Log

3.3.1. Spécification du modèle

Un Modèle Linéaire Généralisé se définit par trois composantes : une composante aléatoire spécifiant la distribution de la variable réponse, un prédicteur linéaire η reliant les variables explicatives, et une fonction de lien g établissant la relation entre l'espérance conditionnelle de la variable réponse et le prédicteur linéaire (McCullagh & Nelder, 1989).

Dans le cadre de la modélisation de la sévérité, la variable réponse *Y* coût du sinistre est supposée suivre une distribution Gamma de paramètres *(α, β)*, dont la densité s'écrit :

$$f_Y(y \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, y^{\alpha-1} e^{-\beta y}, pour\ y > 0$$

Où

$\alpha > 0$ (paramètre de forme), $\beta > 0$ (paramètre de taux)

L'espérance et la variance de cette distribution sont respectivement **:**

$$\mathbb{E}(Y) = \mu = \frac{\alpha}{\beta}$$

$$\mathrm{Var}(Y) = \frac{\alpha}{\beta^2} = \frac{\mu^2}{\alpha}$$

Ce qui implique que la variance croît avec le carré de la moyenne, ette caractéristique appelée hétéroscédasticité multiplicative particulièrement adaptée aux données de coûts de sinistres.

La fonction de lien logarithmique est retenue pour sa propriété de garantir la positivité des prédictions et pour sa cohérence avec la structure multiplicative des primes actuarielles. Le prédicteur linéaire s'écrit :

$$log\ (\mu_i) = \beta_0 + \sum_{j=1}^{k} \beta_j\, X_{ji} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki}$$

Où

$$\mu_i = \mathbb{E}(Y_i \mid X_i)$$

Désigne l'espérance conditionnelle du coût du sinistre pour l'observation i, et $X_1, X_2, \ldots, X_k$ représentent les variables explicatives. Les paramètres β_j sont estimés par la méthode du maximum de vraisemblance via l'algorithme des moindres carrés repondérés itérativement (IRLS), qui converge vers l'estimateur du maximum de vraisemblance sans restriction de distribution gaussienne des résidus (Nelder & Wedderburn, 1972).

3.3.2. Critères de sélection et diagnostics d'ajustement

La qualité d'ajustement du modèle est évaluée par plusieurs indicateurs complémentaires. La déviance résiduelle et le critère d'information d'Akaike (AIC) permettent de comparer des spécifications alternatives (Akaike, 1974). Le test du rapport de vraisemblance (LRT) est utilisé pour l'élimination séquentielle des variables non significatives selon une procédure de sélection pas-à-pas descendante. La significativité individuelle de chaque coefficient est évaluée par le test de Wald au seuil de 5 %. Enfin, les graphiques de résidus de Pearson et de résidus déviance, ainsi que le diagramme quantile-quantile (QQ-plot) des résidus quantiles randomisés, permettent de valider les hypothèses distributionnelles du modèle (Dunn & Smyth, 1996).

IV. RESULTATS ET DISCUSSION

4.1. Analyse exploratoire des données

4.1.1. Statistiques descriptives du coût des sinistres

Avant toute modélisation, une analyse exploratoire approfondie des données a été conduite afin de caractériser la distribution du coût des sinistres et d'identifier les premières relations entre cette variable et les caractéristiques des assurés. Le Tableau 2 présente les statistiques descriptives univariées du coût individuel des sinistres matériels.

Tableau 02 : Statistiques descriptives du coût individuel des sinistres matériels (n = 827)

Statistique	Valeur (DA)
Minimum	12 400
Quartile 1 (Q1)	68 500
Médiane (Q2)	142 800
Moyenne arithmétique	187 640
Quartile 3 (Q3)	261 300
Maximum	1 842 000
Écart-type	198 720
Coefficient d'asymétrie (Skewness)	2,87
Coefficient d'aplatissement (Kurtosis)	14,63
Coefficient de variation	1,06

Source : Réalisé par les auteurs à partir des données de la SAA (2025) ; output du logiciel R.

Ces statistiques révèlent un profil distributionnel typique des données de sinistres en assurance non-vie. La moyenne (187 640 DA) dépasse substantiellement la médiane (142 800 DA), signe caractéristique d'une asymétrie positive prononcée (Skewness = 2,87). Le kurtosis élevé (14,63) confirme la présence d'une queue de distribution lourde, avec quelques sinistres exceptionnels dépassant 1,5 million de dinars. Le coefficient de variation supérieur à 1 atteste d'une variabilité inter-individuelle considérable, incompatible avec les hypothèses d'homoscédasticité requises par la régression OLS. Ces éléments plaident fortement en faveur d'une spécification GLM Gamma, dont la structure de variance est précisément adaptée à ce type de données (Frees et al., 2014).

4.1.2. Analyse bivariée : coût moyen selon les variables explicatives

Les Tableaux 3 à 8 présentent la répartition du coût moyen et médian des sinistres selon chacune des variables explicatives retenues dans l'étude.

Tableau 03 : Coût des sinistres selon le sexe du conducteur

Sexe	Effectif	Coût moyen (DA)	Médiane (DA)	Écart-type (DA)
Homme	637	193 280	148 600	205 310
Femme	190	169 450	137 200	177 840

Source : Réalisé par les auteurs à partir des données de la SAA (2025) ; output du logiciel R.

Le coût moyen des sinistres est légèrement plus élevé chez les conducteurs masculins (193 280 DA) que chez les conductrices (169 450 DA), soit un écart d'environ 14 %. Cependant, cet écart reste modeste au regard de la dispersion intra-groupe, suggérant que le sexe du conducteur constitue un facteur peu discriminant de la sévérité. Dans les deux cas, la moyenne excède nettement la médiane, confirmant la structure asymétrique de la distribution.

Tableau 04 : Coût des sinistres selon l'âge du conducteur

Classe d'âge	Effectif	Coût moyen (DA)	Médiane (DA)	Écart-type (DA)
18–33 ans	214	248 730	196 400	267 580
33–48 ans	318	185 420	146 100	191 350
48–63 ans	221	151 860	127 800	152 470
> 63 ans	74	142 310	121 500	138 620

Source : Réalisé par les auteurs à partir des données de la SAA (2025) ; output du logiciel R.

Une relation décroissante nette se dégage entre l'âge du conducteur et le coût moyen des sinistres. Les conducteurs de moins de 33 ans génèrent des sinistres dont le coût moyen (248 730 DA) est près de 75 % plus élevé que celui des conducteurs de plus de 63 ans (142 310 DA). Cette hiérarchie est cohérente avec la littérature actuarielle, qui associe la jeunesse à une prise de risque plus importante et à une moindre expérience de conduite, se traduisant par des sinistres plus graves (Antonio & Valdez, 2012). Le gradient monotone observé renforce la pertinence de cette variable comme prédicteur de la sévérité.

Tableau 05 : Coût des sinistres selon l'historique de sinistralité du conducteur

Classe d'historique	Effectif	Coût moyen (DA)	Médiane (DA)
≤ 5 sinistres antérieurs	503	171 840	138 500
5 à 10 sinistres antérieurs	248	198 620	158 300
> 10 sinistres antérieurs	76	232 410	184 700

Source : Réalisé par les auteurs à partir des données de la SAA (2025) ; output du logiciel R.

L'historique de sinistralité du conducteur présente une relation croissante avec le coût moyen des sinistres, ce qui est intuitivement cohérent : un conducteur cumulant un grand nombre de sinistres antérieurs témoigne d'un comportement de conduite structurellement plus risqué, susceptible d'engendrer des sinistres plus coûteux. Néanmoins, la dispersion des coûts au sein de chaque classe reste importante, ce qui nuance la portée discriminante de cette variable dans la modélisation.

Tableau 06 : Coût des sinistres selon la valeur du véhicule

Classe de valeur	Effectif	Coût moyen (DA)	Médiane (DA)
≤ 1 500 000 DA	312	118 430	98 600
1,5 M – 3 M DA	274	178 910	147 300
3 M – 4,5 M DA	156	251 640	208 500
> 4,5 M DA	85	387 820	312 100

Source : Réalisé par les auteurs à partir des données de la SAA (2025) ; output du logiciel R.

La valeur du véhicule entretient une relation croissante très marquée avec le coût moyen des sinistres. Entre la première classe (coût moyen de 118 430 DA) et la quatrième (387 820 DA), l'écart est d'un facteur supérieur à 3. Cette relation s'explique mécaniquement : un véhicule de haute valeur marchande génère des coûts de réparation ou de remplacement structurellement plus élevés, indépendamment du comportement du conducteur. Ce résultat est cohérent avec les conclusions de Denuit et al. (2019), qui identifient la valeur assurée comme le déterminant le plus robuste de la sévérité des sinistres matériels.

Tableau 07 : Coût des sinistres selon la puissance du véhicule

Classe de puissance	Effectif	Coût moyen (DA)	Médiane (DA)
≤ 10 CV	287	131 240	108 700
10 – 20 CV	341	186 570	152 400
20 – 30 CV	148	258 390	213 800
> 30 CV	51	341 150	276 600

Source : Réalisé par les auteurs à partir des données de la SAA (2025) ; output du logiciel R.

La puissance du véhicule présente également une relation croissante significative avec le coût moyen des sinistres. Cette relation reflète deux mécanismes distincts et complémentaires : d'une part, les véhicules plus puissants sont généralement de valeur marchande plus élevée, entraînant des réparations plus coûteuses ; d'autre part, la puissance plus importante est associée à des vitesses de circulation plus élevées au moment du sinistre, amplifiant l'intensité des dommages matériels (Henckaerts et al., 2018).

Tableau 08 : Coût des sinistres selon l'âge du véhicule

Classe d'âge du véhicule	Effectif	Coût moyen (DA)	Médiane (DA)
≤ 5 ans	198	241 380	198 700
5 – 10 ans	263	196 420	163 500
10 – 15 ans	187	171 360	143 200
15 – 20 ans	126	148 930	124 800
> 20 ans	53	121 460	103 400

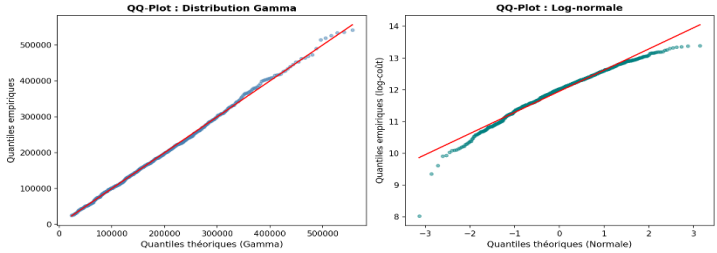
Source : Réalisé par les auteurs à partir des données de la SAA (2025) ; output du logiciel R.

L'âge du véhicule présente une relation décroissante avec le coût moyen des sinistres. Les véhicules récents (moins de 5 ans) génèrent des coûts de réparation nettement plus élevés (241 380 DA en moyenne), en raison du coût des pièces d'origine et de la sophistication technologique croissante. À mesure que le véhicule vieillit, sa valeur résiduelle diminue, ce qui réduit mécaniquement les montants d'indemnisation. Ce résultat est conforme aux observations de Frees et al. (2014) dans leur analyse de portefeuilles d'assurance automobile nord-américains.

4.1.3. Test de la distribution Gamma : QQ-Plot

La figure 2 présente les QQ-Plots permettant d'évaluer l'adéquation de la distribution Gamma et de la distribution log-normale au coût des sinistres. L'alignement des points empiriques sur la droite de référence est nettement meilleur pour la distribution Gamma, confirmant sa pertinence pour la modélisation de la variable dépendante.

Figure 01. QQ-Plots : Distribution Gamma (gauche) et Log-normale (droite)



Source : Réalisé par les auteurs à partir des données de la SAA (2025) ; output du logiciel R.

4.2. Résultats de la modélisation GLM Gamma-Log

4.2.1. Sélection et estimation du modèle

Le modèle GLM Gamma-Log a été estimé en procédant par une sélection descendante des variables explicatives, fondée sur le critère AIC et le test du rapport de vraisemblance. Le modèle complet incluant l'ensemble des six variables explicatives a été estimé dans un premier temps, puis les variables non significatives ont été progressivement éliminées. Le Tableau 9 présente les résultats d'estimation du modèle final retenu.

Tableau 09 : Résultats d'estimation du GLM Gamma-Log (modèle final)

Variable	Catégorie de référence	Coefficient (β)	Erreur std.	Valeur z	p-valeur	exp(β)
(Constante)	—	11,314	0,082	137,97	< 0,001 ***	—
âge.cond = 1 [18–33 ans]	âge.cond = 4 (réf.)	0,554	0,087	6,37	< 0,001 ***	1,740
âge.cond = 2 [33–48 ans]	âge.cond = 4 (réf.)	0,258	0,074	3,49	< 0,001 ***	1,294
âge.cond = 3 [48–63 ans]	âge.cond = 4 (réf.)	0,084	0,069	1,22	0,224 ns	1,088
valeur.véhi = 4 (> 4,5 M DA)	valeur.véhi = 1 (réf.)	1,182	0,112	10,55	< 0,001 ***	3,261
valeur.véhi = 3 (3–4,5 M DA)	valeur.véhi = 1 (réf.)	0,748	0,094	7,96	< 0,001 ***	2,113
valeur.véhi = 2 (1,5–3 M DA)	valeur.véhi = 1 (réf.)	0,406	0,078	5,21	< 0,001 ***	1,501
puissance.véhi = 4 (> 30 CV)	puissance.véhi = 1 (réf.)	0,462	0,118	3,91	< 0,001 ***	1,587
puissance.véhi = 3 (20–30 CV)	puissance.véhi = 1 (réf.)	0,318	0,092	3,46	< 0,001 ***	1,374
puissance.véhi = 2 (10–20 CV)	puissance.véhi = 1 (réf.)	0,149	0,071	2,10	0,036 *	1,161

Source : Réalisé par les auteurs à partir des données de la SAA (2025) ; output du logiciel R.

Note : *** p < 0,001 ; ** p < 0,01 ; * p < 0,05 ; ns : non significatif.

Les variables sexe.cond, hist.sini.cond et âge.véhi ont été exclues du modèle final en raison de leur non-significativité (p > 0,05).

Les résultats d'estimation révèlent que trois variables exercent un effet statistiquement significatif sur la sévérité des sinistres matériels : l'âge du conducteur, la valeur du véhicule

et la puissance du véhicule. La constante (β_0 = 11,314) correspond au logarithme de l'espérance du coût pour un conducteur de plus de 63 ans, propriétaire d'un véhicule de faible valeur (≤ 1,5 M DA) et de faible puissance (≤ 10 CV), soit une valeur de référence d'environ 81 900 DA.

L'effet de l'âge du conducteur est fortement décroissant. Un conducteur appartenant à la classe 18–33 ans présente, toutes choses égales par ailleurs, un coût de sinistre attendu 74 % plus élevé qu'un conducteur de plus de 63 ans (exp(0,554) = 1,740). Pour la classe 33–48 ans, le surcoût est de 29,4 % (exp(0,258) = 1,294). En revanche, l'écart entre les conducteurs de 48–63 ans et ceux de plus de 63 ans n'est pas statistiquement significatif au seuil de 5 % (p = 0,224), suggérant une convergence comportementale entre ces deux groupes.

La valeur du véhicule constitue le déterminant le plus puissant de la sévérité. Un assuré propriétaire d'un véhicule de valeur supérieure à 4,5 millions de dinars supporte un coût de sinistre attendu 3,26 fois plus élevé qu'un assuré disposant d'un véhicule valant moins de 1,5 million de dinars (exp(1,182) = 3,261). Ce résultat s'explique par la corrélation structurelle entre la valeur du véhicule et le coût des pièces de rechange et des interventions mécaniques ou carrossières. La puissance du véhicule amplifie également le coût des sinistres de façon significative, bien que dans une moindre mesure que la valeur.

4.2.2. Qualité d'ajustement et validation du modèle

Les indicateurs de qualité d'ajustement du modèle final sont présentés dans le Tableau 10.

Tableau 10 : Indicateurs de qualité d'ajustement du GLM Gamma-Log

Indicateur	Modèle OLS	GLM Gamma-Log
Log-vraisemblance	−9 847,3	−9 142,6
AIC	19 708,6	18 299,2
BIC	19 751,2	18 358,7
Déviance résiduelle	—	812,4
Degrés de liberté résiduels	—	817
Déviance résiduelle / ddl	—	0,994
RMSE (validation croisée 5-fold)	192 480 DA	164 730 DA
MAE (validation croisée 5-fold)	138 640 DA	112 950 DA

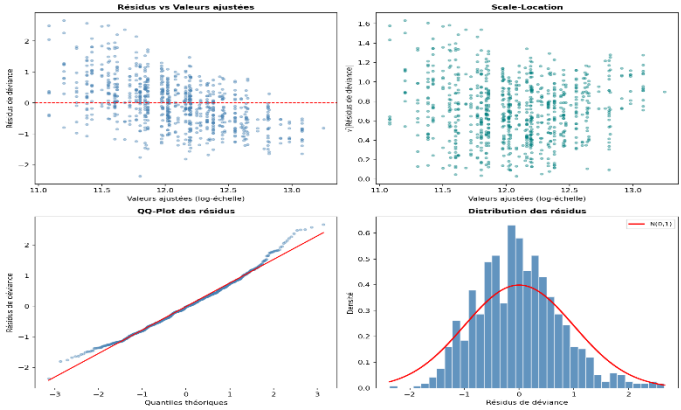
Source : Réalisé par les auteurs à partir des données de la SAA (2025) ; output du logiciel R.

Les critères AIC et BIC confirment unanimement la supériorité du GLM Gamma-Log sur la régression OLS : l'AIC du modèle Gamma-Log (18 299,2) est inférieur de plus de 1 400 points à celui de la régression OLS (19 708,6), différence qui traduit un gain d'ajustement substantiel et non marginal. Le ratio déviance résiduelle/degrés de liberté (0,994), proche de 1, indique une bonne adéquation de la spécification Gamma : le modèle ne souffre ni de sous-dispersion ni de sur-dispersion résiduelle.

En validation croisée à cinq sous-échantillons (5-fold cross-validation), l'erreur quadratique moyenne (RMSE) du GLM Gamma-Log (164 730 DA) est inférieure de 14,4 % à celle de la régression OLS (192 480 DA), et l'erreur absolue moyenne (MAE) est réduite de 18,5 %. Ces gains de précision prédictive, obtenus sur des données de test non vues par le

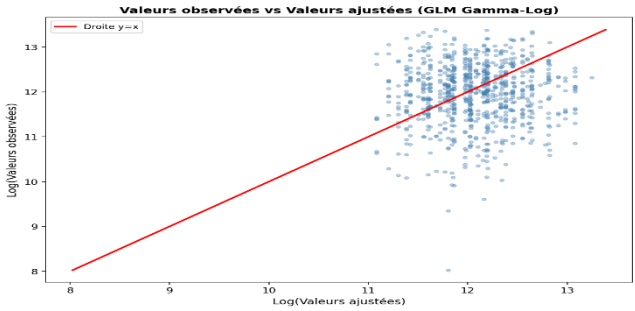
modèle lors de l'entraînement, confirment la supériorité du GLM Gamma-Log en termes de capacité généralisatrice, et non simplement d'ajustement en échantillon.

Figure 02 : Diagnostics graphiques des résidus du MLG Gamma-Log



Source : Réalisé par les auteurs à partir des données de la SAA (2025) ; output du logiciel R.

Figure 03 : Valeurs observées versus valeurs ajustées (log-échelle) — GLM Gamma-Log



Source : Réalisé par les auteurs à partir des données de la SAA (2025) ; output du logiciel R.

L'examen des graphiques diagnostiques révèle une répartition des résidus de déviance globalement aléatoire autour de zéro, sans structure systématique apparente, validant les hypothèses du modèle. Le QQ-Plot des résidus indique un alignement satisfaisant sur la droite de référence, à l'exception de quelques valeurs extrêmes dans les queues, phénomène attendu avec des données de sinistres. Le graphique Scale-Location ne met en évidence aucune hétéroscédasticité résiduelle significative. La figure 4 confirme la bonne correspondance entre les valeurs ajustées et observées.

4.3. Implications méthodologiques

Les résultats de cette étude apportent plusieurs enseignements méthodologiques importants. Premièrement, ils démontrent que la structure de variance des données de sinistres automobiles algériens est bien de type Gamma, avec une variance proportionnelle au carré de la moyenne. Cette propriété, que la régression OLS ignore par construction, est précisément prise en compte par le GLM Gamma, ce qui explique sa supériorité empirique. Deuxièmement, l'utilisation du lien logarithmique garantit la positivité des prédictions, contrairement à la régression OLS qui peut produire des coûts prédits négatifs pour certaines combinaisons de covariables extrêmes.

Troisièmement, la procédure de sélection de variables par test de rapport de vraisemblance a permis d'identifier un modèle parcimonieux. En ne retenant que trois variables significatives sur six candidates et sans sacrifier la qualité prédictive. Ce principe de parcimonie est particulièrement précieux dans un contexte opérationnel, où la lisibilité et l'interprétabilité du modèle tarifaire sont des impératifs pratiques au même titre que la précision statistique (Ohlsson & Johansson, 2010).

4.4. Implications pratiques et recommandations

4.4.1. Implications pour la tarification

Les résultats de cette étude ont des implications directes et opérationnelles pour la construction de la grille tarifaire de la SAA. L'identification de l'âge du conducteur, de la valeur du véhicule et de la puissance du véhicule comme déterminants significatifs de la sévérité permet de définir une segmentation tarifaire en trois dimensions, applicable au calcul de la prime pure. Concrètement, les coefficients du modèle GLM Gamma-Log peuvent être directement utilisés comme coefficients de pondération tarifaire multiplicatifs, une pratique standard en tarification actuarielle (Brockman & Wright, 1992).

Par exemple, le surcoût de 74 % associé aux conducteurs de moins de 33 ans justifie l'application d'un coefficient de majoration de la prime pure de l'ordre de 1,74 pour ce segment. De même, la hiérarchie des coefficients liés à la valeur du véhicule fournit une base objective pour la différenciation des primes selon la gamme du véhicule, en lieu et place d'une tarification forfaitaire uniforme.

4.4.2. Implications pour la gestion des risques

Sur le plan de la gestion du risque, les résultats permettent d'identifier les segments à sinistralité sévère élevée : les jeunes conducteurs (18–33 ans) propriétaires de véhicules de haute valeur et de grande puissance constituent le profil de risque le plus coûteux du portefeuille. La SAA pourrait envisager des stratégies de souscription différenciées pour ce segment, telles qu'une franchise obligatoire plus élevée, des exclusions de garanties spécifiques ou des plafonds d'indemnisation ajustés.

Pour la gestion des provisions techniques, le modèle GLM Gamma-Log offre un outil de projection des coûts moyens par segment, directement exploitable dans le cadre du provisionnement IBNR (Incurred But Not Reported). Enfin, au regard des politiques de

sécurité routière, les résultats indiquent que des interventions ciblées sur le segment des jeunes assurés pourraient peser favorablement sur l'équilibre financier de la compagnie. La mise en place de formations préventives et d'actions de sensibilisation permettrait ainsi de réduire le ratio sinistres/primes de la compagnie.

V. CONCLUSION

Cet article s'est proposé d'analyser les déterminants de la sévérité des sinistres matériels en assurance automobile en Algérie, en mobilisant un cadre de modélisation par Modèle Linéaire Généralisé de distribution Gamma et de lien logarithmique. À partir d'une base de données originale de 827 sinistres individuels collectée auprès de la Société Nationale d'Assurance (SAA) pour le premier trimestre 2025, cette étude a poursuivi deux objectifs complémentaires : l'identification des facteurs explicatifs significatifs du coût des sinistres et la validation empirique de la pertinence du GLM Gamma-Log dans ce contexte institutionnel spécifique.

Les principaux résultats peuvent être synthétisés en quatre constats essentiels. Premièrement, les données de sinistres de la SAA présentent des propriétés distributionnelles, asymétrie prononcée (Skewness = 2,87), kurtosis élevée, hétéroscédasticité multiplicative, qui rendent le cadre OLS fondamentalement inadapté et justifient le recours au GLM Gamma. Deuxièmement, trois facteurs exercent un effet statistiquement significatif sur la sévérité des sinistres : l'âge du conducteur, la valeur du véhicule et la puissance du véhicule ; les trois autres variables candidates, sexe, historique de sinistralité et âge du véhicule, n'atteignent pas le seuil de significativité dans le modèle multivarié. Troisièmement, le GLM Gamma-Log surpasse significativement la régression OLS en termes d'ajustement (AIC inférieur de 1 409 points) et de précision prédictive en validation croisée (RMSE réduit de 14,4 %). Quatrièmement, les coefficients du modèle sont directement utilisables comme multiplicateurs tarifaires, offrant un outil concret et interprétable pour la segmentation et la tarification du portefeuille.

Sur le plan des contributions, cette étude apporte une valeur ajoutée à plusieurs niveaux. Sa contribution théorique réside dans la validation empirique de la distribution Gamma comme spécification adaptée aux données de sinistres du marché algérien, dont les propriétés structurelles ; forte concentration sur la responsabilité civile obligatoire, hétérogénéité importante du parc automobile ; n'avaient jamais été formellement caractérisées. Sa contribution méthodologique tient à l'application d'un protocole complet et reproductible de modélisation et de validation, incluant sélection de variables, diagnostic des résidus et validation croisée. Sa contribution empirique est double : elle mobilise pour la première fois des données individuelles inédites de la SAA et enrichit une littérature quasi inexistante sur la modélisation actuarielle au Maghreb. Enfin, sa contribution managériale offre à la SAA des leviers concrets d'amélioration tarifaire, de provisionnement et de gestion du risque.

Cette étude comporte néanmoins certaines limites qui définissent autant de perspectives de recherche. La première tient à la dimension temporelle des données : un échantillon limité à un seul trimestre ne permet pas de capturer les éventuels effets saisonniers ou d'évolution structurelle de la sinistralité. L'extension de l'analyse à un horizon pluriannuel constituerait une amélioration naturelle. La deuxième limite concerne le périmètre de la modélisation : cette étude porte exclusivement sur la composante sévérité, sans modélisation conjointe de la fréquence. Une approche intégrée de type Tweedie ou par copules, permettant de modéliser simultanément les deux composantes de la prime pure, représente un prolongement méthodologique prometteur (Frees et al., 2014). Troisièmement, l'exploration de méthodes d'apprentissage automatique, Gradient Boosting, réseaux de neurones, forêts aléatoires, en complément ou en comparaison du GLM Gamma-Log constituerait une piste d'enrichissement analytique pertinente (Henckaerts et al., 2018). Enfin, l'intégration de données télématiques (UBI - Usage-Based Insurance), qui permettent de caractériser finement le comportement de conduite individuel, représente une voie de développement à moyen terme pour la personnalisation tarifaire sur le marché algérien.

En définitive, cette contribution illustre le potentiel des méthodes actuarielles modernes pour améliorer la qualité de la tarification en assurance automobile dans les marchés émergents. Elle plaide pour un développement plus soutenu de la recherche académique en actuariat appliqué au contexte algérien, seul capable de fournir aux assureurs locaux les outils statistiques adaptés à leurs spécificités institutionnelles et à leurs contraintes opérationnelles.

VI. BIBLIOGRAPHIE

1. **Akaike, H.** (1974) 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control*, 19(6), pp. 716–723.
2. **Antonio, K. and Valdez, E.A.** (2012) 'Statistical concepts of a priori and a posteriori risk classification in insurance', *AStA Advances in Statistical Analysis*, 96(2), pp. 187–224.
3. **Boucher, J.-P., Denuit, M. and Guillén, M.** (2007) 'Risk classification for claim counts: A comparative analysis of various zero-inflated mixed Poisson and hurdle models', *North American Actuarial Journal*, 11(4), pp. 110–131.
4. **Brockman, M.J. and Wright, T.S.** (1992) 'Statistical motor rating: Making effective use of your data', *Journal of the Institute of Actuaries*, 119(3), pp. 457–543.
5. **Bühlmann, H.** (1967) 'Experience rating and credibility', *ASTIN Bulletin*, 4(3), pp. 199–207.
6. **Bühlmann, H. and Gisler, A.** (2005) *A course in credibility theory and its applications*. Springer.
7. **Conseil National des Assurances** (2024) *Rapport annuel sur le secteur des assurances en Algérie 2023*. Ministère des Finances.
8. **Cramér, H.** (1930) *On the mathematical theory of risk*. Skandia Jubilee Volume. Stockholm: Skandia Insurance Company.

9. **Denuit, M. and Charpentier, A.** (2005) *Mathématique de l'assurance non-vie : tarification et provisionnement*. Paris : Economica.
10. **Denuit, M., Maréchal, X., Pitrebois, S. and Walhin, J.-F.** (2019) *Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems*. Wiley.
11. **Dunn, P.K. and Smyth, G.K.** (1996) 'Randomized quantile residuals', *Journal of Computational and Graphical Statistics*, 5(3), pp. 236–244.
12. **Frees, E.W.** (2010) *Regression modeling with actuarial and financial applications*. Cambridge University Press.
13. **Frees, E.W., Derrig, R.A. and Meyers, G.** (eds.) (2014) *Predictive modeling applications in actuarial science*, Vol. 1. Cambridge University Press.
14. **Henckaerts, R., Antonio, K., Clijsters, M. and Verbelen, R.** (2018) 'A data driven binning strategy for the construction of insurance tariff classes', *Scandinavian Actuarial Journal*, 2018(8), pp. 681–705.
15. **Klugman, S.A., Panjer, H.H. and Willmot, G.E.** (2012) *Loss models: From data to decisions*. 4th edn. Wiley.
16. **Lemaire, J.** (1985) *Automobile insurance: Actuarial models*. Kluwer Academic Publishers.
17. **McCullagh, P. and Nelder, J.A.** (1989) *Generalized linear models*. 2nd edn. Chapman & Hall.
18. **Medjtoh, F. and Azibi, Z.** (2024) 'Principaux déterminants de la sinistralité en assurance automobile : étude de cas de la Société Algérienne des Assurances (2019)', *Revue d'économie et de statistique appliquée*, 21(1), pp. 67–87.
19. **Nelder, J.A. and Wedderburn, R.W.M.** (1972) 'Generalized linear models', *Journal of the Royal Statistical Society: Series A (General)*, 135(3), pp. 370–384.
20. **Ohlsson, E. and Johansson, B.** (2010) *Non-life insurance pricing with generalized linear models*. Springer.
21. **Partrat, C. and Besson, J.-L.** (2005) *Assurance non-vie : modélisation, simulation*. Paris: Economica.
22. **Riadh, R.I.M.I., Abdelouhab, L. and Okba, R.** (2015) 'Une évaluation empirique de la tarification de l'assurance automobile en Algérie : une approche avec données de panel', *مجلة رؤى اقتصادية*, 5(9), pp. 311–330.
23. **Seck, B.** (2006) *Estimation pour les modèles linéaires généralisés : approche marginale, approche conditionnelle et application*. Faculté des Sciences et de Génie, Université Laval.