

CHAPITRE 11

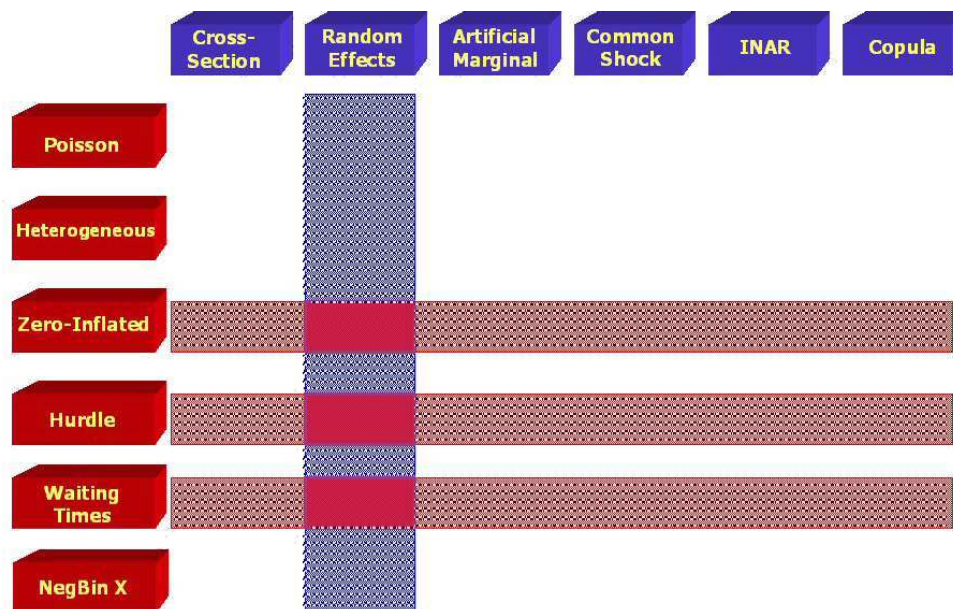
Conclusion, perspectives et recherches futures

Sommaire des travaux et interprétation des résultats

Cette thèse a permis non seulement de couvrir une panoplie de modèles pour ajuster le nombre de réclamations, mais aussi d'analyser différemment ces statistiques afin de mieux comprendre la composition du portefeuille d'assurance et le comportement des assurés. Puisque de nombreux sujets théoriques ont été abordés au travers des différents chapitres de la thèse, celle-ci se conclura en révisant les notions particulières abordées, touchant à la science actuarielle mais aussi à la théorie statistique.

Distributions de comptage

Divers modèles basés sur l'hypothèse que tous les contrats sont indépendants (*cross-section data*) ont été considérés dans le premier chapitre. Ceux-ci ont souvent été développés en généralisant diverses formes de la loi de Poisson. Similairement, dans le chapitre 4, en se basant sur les lois de Poisson et binomiale négative, un vaste choix de modèles ayant la flexibilité de supposer une dépendance entre les contrats d'un même assuré (*panel data*) a été développé. Graphiquement, l'ensemble des modèles analysés peut être illustré de la manière suivante :



Ainsi, les modèles à barrière (*hurdle models*, chapitre 1), les modèles gonflés à zéro (*zero-inflated models*, chapitre 1) et les modèles généralisant le temps d'attente entre les sinistres (*waiting times models*, chapitre 5) se justifiaient intuitivement pour les données avec indépendance temporelle, alors que les modèles avec effets aléatoires (*random effects*, chapitre 4) possédaient d'intéressantes

propriétés pour les données de panel. Ainsi, il semblait pertinent de combiner ces constructions, ce qui est illustré dans le graphique précédent par les croisements des lignes verticales et horizontales. L'idée générale de ces généralisations pour données de panel de ces trois types de modèles à effets aléatoires était d'introduire dans la distribution jointe des contrats d'un même assuré un ou plusieurs facteurs individuels ayant une distribution statistique quelconque. L'introduction de ces termes aléatoires communs à toutes les réalisations d'un assuré introduit donc une dépendance entre celles-ci.

Comparaisons et tests statistiques

Puisque de nombreux modèles sont possibles pour modéliser le nombre de réclamations, il était nécessaire de construire des méthodes permettant de comparer l'ajustement de ces modèles. Dans certaines situations, deux modèles à comparer peuvent être appelés *liés*, si l'un d'eux est équivalent à l'autre pour certaines restrictions paramétriques. Dans le cas inverse, les deux modèles sont considérés comme étant *non-liés* (chapitre 1).

Pour comparer deux modèles liés, il suffit de tester la valeur du paramètre provoquant l'égalité des modèles. Toutefois, parce que certaines situations impliquent le test de paramètres se situant sur la borne de leur domaine, les tests statistiques classiques à utiliser se devaient d'être corrigés, ou encore laissés de côté afin de choisir des tests étant invariables à cette situation. De cette manière, un système de comparaison des modèles liés avec indépendance temporelle a été établi (chapitre 1), alors qu'en généralisant ces tests de comparaison de modèles afin de les adapter aux modèles de données de panel, les modèles ajustant le mieux nos données d'assurance sous forme longitudinale (chapitres 4, 5, 7 et 8) ont pu être sélectionnés.

Lorsque tous les liens ont été testés, diverses méthodes de comparaison des modèles non-liés ont été étudiés, telles que le test de Vuong (chapitre 1), sa généralisation pour données de panel (chapitres 4 et 8) ou l'utilisation de critères d'information comme le AIC ou le BIC (chapitres 1, 4, 5, 7 et 8).

Statistiquement, ces méthodes permettent de déterminer quel est le modèle qui ajuste le mieux le nombre de sinistres rapportés à l'assureur. Pour les données avec indépendance temporelle, le modèle $NegBin_x$ et le modèle à barrière sont les modèles ajustant le mieux les données (chapitre 1). Pour les données de panel, une telle conclusion ne peut pas être énoncée aussi clairement puisque

plusieurs modèles n'ont pas été comparés directement. Néanmoins, il est à noter que le modèle NB1-Beta (chapitre 4), le modèle à barrière avec copula de Fréchet-Hoeffding (chapitre 7), le modèle Poisson gonflé à zéro avec hétérogénéité gamma (chapitre 8) et le modèle multivarié binomial négatif avec régresseur sur l'exposition au risque (chapitre 5) ont généré des ajustements intéressants et supérieurs à plusieurs autres modèles, notamment face au modèle classique multivarié binomial négatif.

Segmentation et effets des régresseurs

L'introduction des caractéristiques du risque assuré a été effectuée à l'aide de fonctions de score, à travers certains paramètres des distributions (chapitres 1-10). L'estimation des paramètres de cette fonction de score permet d'évaluer l'impact de ces caractéristiques du risque sur la fréquence de sinistres. Parce que la loi de Poisson est de la famille exponentielle, il a été prouvé que l'estimation des paramètres de moyenne sont consistents tant que la forme de la moyenne est bien définie, peu importe si la *vraie* distribution est Poisson (chapitre 1). Néanmoins, la justesse de l'écart-type de ces paramètres n'est aucunement justifiée (chapitres 1 et 2), ce qui peut avoir un impact dans le choix des régresseurs influents. En conséquence, le développement d'autres modèles pour ajuster les données est pertinent.

Malgré le fait que plusieurs bases de données aient été utilisées au travers des différents chapitres, certains résultats communs à chaque portefeuille d'assurance ont pu être observés. Par exemple, lorsque plusieurs autres caractéristiques du risque sont considérées, une conductrice (sexe féminin) produit davantage de réclamations qu'un conducteur masculin alors que les jeunes conducteurs montrent une expérience de sinistre pire que les plus vieux (chapitres 1, 2, 4, 5, 6, 7 et 8). Tout dépendant des modèles, l'impact de ces caractéristiques peut être interprété différemment, que ce soit à travers du nombre moyen de sinistres par année, à travers du paramètre de censure ou au travers du nombre moyen de sinistres lorsque l'assuré a réclamé au moins une fois (chapitres 1, 2, 3, 4, 5, 6, 7 et 8). L'impact des caractéristiques du risque dans l'analyse de la soif du bonus (chapitre 9), peut aussi être considéré.

Hétérogénéité et effets aléatoires

Malgré la présence de régresseurs pour expliquer la fréquence de réclamations, il existe toujours une différence de sinistralité dans chaque classe de risque. En effet, l'introduction des fonctions de score ne rend pas le portefeuille totalement homogène puisque de nombreuses caractéristiques im-

portantes du risque sont non-observées, telles que l'agressivité au volant, ou encore la rapidité des réflexes. Pour corriger cette situation, l'introduction d'un élément aléatoire, appelé hétérogénéité, a été effectuée pour plusieurs des modèles étudiés (chapitre 1).

Dans le chapitre 2, une analyse plus en détails a été effectuée afin de connaître davantage les propriétés de cette hétérogénéité pour la loi de Poisson. Contrairement aux modèles complètement paramétriques, cette analyse a permis de voir qu'il y a, proportionnellement à la moyenne, davantage de variation et d'hétérogénéité dans les risques ayant une petite moyenne de réclamations que dans les profils ayant une moyenne élevée.

Similairement, la même approche a été faite pour les données de panel, où le facteur d'hétérogénéité est commun à chaque contrat d'un même assuré (chapitres 3, 4, 5, 6, 7 et 8). En regardant davantage cette situation où le facteur d'hétérogénéité est plutôt déterministe (effet fixe), certains profils de risque ont été identifiés comme ayant une expérience de perte plus variable que d'autres (chapitre 3). Ainsi, les méthodes présentées aux chapitres 1 et 3 permettent d'identifier les situations où une caractéristique ou un profil de risque se présente comme plus hétérogène que les autres, donnant donc la chance aux compagnies d'assurance de modifier leurs critères de souscription pour ces assurés.

Mécanisme de réclamation

En dehors du fait que certains modèles permettent d'obtenir un meilleur ajustement des données, l'interprétation de ceux-ci permettent aussi de mieux comprendre le mécanisme de réclamation à l'assureur. Quelques cas pertinents méritent d'être cités et revus :

- L'effet des surcharges de primes (bonus-malus) données à l'assuré lorsque celui-ci réclame à l'assureur crée une censure au niveau des réclamations, appelé soit du bonus (chapitres 1 et 9). Les modèles gonflés à zéro et à barrière (chapitres 1, 6, 7, 8, 9 et 10) modélisent cet état de fait et expliquent peut-être le bon ajustement de ces modèles ;
- Parallèlement, la situation où une réclamation à l'assureur modifie la perception de conduite peut aussi expliquer le bon ajustement d'un modèle à barrière (chapitres 1, 6 et 7) ;
- L'idée générale qu'il existe deux types d'assurés, ceux qui réclament et ceux qui ne réclament pas, a été modélisée de plusieurs façons dans la thèse. Que ce soit au niveau des modèles semiparamétriques du deuxième chapitre ou encore des modèles $NegBin_X$ ou à barrière du premier chapitre ;

- La présence de dépendance d'occurrence, signifiant que la survenance d'une réclamation augmente la probabilité d'une future réclamation, a été modélisée par les distributions binomiale négative, Poisson-inverse Gaussienne ou Poisson-lognormale (chapitre 1) ;
- La dépendance de duration, signifiant qu'une réclamation diminue le temps espéré pour une prochaine réclamation, duquel les modèles généralisant le temps d'attente entre les réclamations, a montré un ajustement intéressant (chapitre 5) ;
- La modélisation temporelle de la dépendance entre les contrats d'un même assuré a semblé indiquer une dépendance apparente, à l'inverse d'une dépendance vraie, entre les réclamations (chapitre 4).

Analyse et primes prédictives

Les modèles avec effets aléatoires ont été privilégiés pour modéliser la dépendance entre les contrats d'un même assuré, signifiant l'adoption du point de vue des statistiques bayésiennes. Conséquemment, le facteur individuel a pu être mise-à-jour annuellement selon l'expérience de réclamation de l'assuré. L'analyse *a posteriori* de ce terme aléatoire permet d'ajuster la prime annuelle de l'assuré, révélant ainsi partiellement l'effet des caractéristiques non observées sur la fréquence de réclamations (chapitres 4, 5, 6, 7, 8, 9 et 10).

Parce qu'il arrive fréquemment que le développement de ces distributions prédictives soit complexe, différentes méthodes peuvent être envisagées pour obtenir les primes futures. En dehors du calcul de la distribution prédictive exacte (chapitres 4, 6, 7 et 8), nous avons utilisé les simulations Monte-Carlo par chaînes de Markov (chapitres 5 et 7), de même que la théorie de la crédibilité linéaire (chapitre 10). L'analyse plus en détails de ces primes prédictives pour le modèle gonflé à zéro nous a permis d'interpréter le problème de la soif du bonus sous un autre angle (chapitre 9).

Considérations statistiques

Des résultats statistiques intéressants ont aussi été obtenus au cours de la thèse. Bien que ceux-ci n'aient pas d'impact clairement direct sur la tarification de la prime annuelle, ils peuvent toutefois être vus comme un avancement scientifique pertinent pouvant potentiellement permettre une évolution des statistiques et de la science actuarielle :

- L'évolution de la dépendance *a posteriori* entre les effets aléatoires du modèle à barrière avec copula Gaussienne (chapitre 7) permet d'observer des résultats intéressants. En effet, celle-ci, évaluée par le tau de Kendall, évolue de manière linéaire selon la durée de l'historique

de sinistres, alors que la pente de cette évolution est sensiblement la même pour toutes les situations d'historique de pertes ;

- La généralisation de la méthode semiparamétrique d'évaluation de la variance de la loi de Poisson, en retirant le lien entre la moyenne et la variance a aussi été analysée. Dans le chapitre 2, l'étude par la fonction de score, couplée à l'utilisation d'estimateurs de densité par noyau (*kernel*) a généré des résultats intéressants, tout en offrant un angle de vue beaucoup plus adapté que l'ancienne méthode par estimateurs des voisins les plus proches (*nearest neighbors*). Cette méthode offre de nombreuses généralisations possibles, telle que l'analyse pour données de panel ;
- La comparaison de l'ajustement par ventilation annuelle des critères d'information (AIC, BIC, etc.) du chapitre 8 est aussi pertinente pour des analyses futures. En effet, cette méthode d'observation de l'ajustement des données de panel permet de voir l'évolution de l'ajustement au cours du temps, en identifiant clairement quels modèles offrent une meilleure capacité de prédiction (chapitre 8) ;
- Le lien entre l'analyse par effets fixes et effets aléatoires, permettant de voir que les modèles avec effets aléatoires biaisent les paramètres de moyenne, tout en représentant néanmoins l'effet apparent des caractéristiques du risque sur la fréquence de sinistre, a permis d'aborder les problèmes de données de panel différemment (chapitre 3) ;
- L'approximation du modèle avec temps d'attente Weibull par le modèle classique Poisson sur lequel un régresseur supplémentaire pour la durée d'exposition au risque était aussi originale et facilitait les calculs (chapitre 5) ;
- La fonction de perte exponentielle utilisée pour calculer la prime future d'un modèle gonflé à zéro (chapitre 9) offre une meilleure flexibilité dans la détermination de la prime future ;
- Même si la plupart des modèles étudiés dans les chapitres 1, 4 et 5 a été introduit en actuariat pour la première fois, ils ont été empruntés de la biostatistique, de l'économétrie ou de divers autres domaines de statistiques appliquées. A l'opposé, les trois modèles généralisés pour effets aléatoires (modèles gonflés à zéro, à barrière ou généralisant le temps d'attente entre les réclamations) sont tout-à-fait nouveaux (chapitres 5, 6, 7 et 8).

Les innovations statistiques de cette thèse permettent aussi d'utiliser ces modèles et ces conclusions dans d'autres domaines que l'actuariat. Ainsi, des applications économétriques, biostatistiques ou autres, peuvent être envisagées.

Conclusion générale, perspectives et recherches futures

La thèse a permis de couvrir une étendue assez large du domaine de la segmentation du nombre de sinistres en assurance. Comme cité précédemment, la thèse a présenté plusieurs modèles pour le nombre de réclamations, et aussi tenté de construire des méthodes de comparaison théoriquement justes. Néanmoins, certains modèles n'ont pas été testés directement l'un envers l'autre et il pourrait être justifié de se demander, finalement, quel est le meilleur modèle pour ajuster la fréquence de sinistres. De plus, bien que des méthodes statistiquement viables aient été construites et expliquées, il reste important de ne pas oublier que plusieurs autres considérations plus subjectives peuvent, et doivent être considérées dans le choix du modèle. A mon avis, il n'existe pas de *vrai* modèle ajustant les données, ce qui signifie que ces considérations personnelles, couplées aux résultats statistiques, soient importantes. Parmi celles-ci, notons l'interprétation intuitive du modèle, la simplicité du modèle, la possibilité de généralisation dans diverses directions, ou encore le fait que ce modèle offre des résultats différents des modèles classiques. Ainsi, clairement, il semble que les modèles basés sur la généralisation du modèle de Poisson gonflé à zéro possède un avantage sur les autres familles de modèles. En effet, le modèle peut s'exprimer de manière fermée pour la distribution jointe et la distribution prédictive, un seul effet aléatoire peut être utilisé, l'interprétation du modèle est intuitive et le modèle peut être utilisé sous l'angle de la soif du bonus.

Au niveau des possibilités de recherche, il semble que l'analyse statistique des données d'assurance non-vie est rendue au point de considérer plusieurs variables aléatoires simultanément. Diverses statistiques telles que le nombre de réclamations pour diverses couvertures d'assurance ou encore le coût moyen par sinistre de ces couvertures doivent être analysées de manière globale afin d'obtenir une prime d'assurance totale par client. Ce point de vue d'analyse peut se justifier de diverses façons. En effet, nous pouvons penser à la situation où une même réclamation en responsabilité civile automobile peut aussi impliquer une réclamation pour les autres couvertures d'assurance. Similairement, on peut supposer qu'un mauvais assuré en assurance automobile peut aussi être mauvais en assurance habitation (négligence, distraction, etc.). Finalement, un assuré qui cause de nombreux accidents automobiles peut peut-être générer des sévérités de réclamations beaucoup plus élevées. Pour ces raisons, et pour plusieurs autres, il semble justifié de considérer la dépendance entre toutes ces statistiques.

Non seulement il semble plus juste d'analyser globalement ces statistiques, mais ces nouveaux

modèles permettent d'analyser d'une toute autre façon certains problèmes classiques en actuariat, tels que l'impact des franchises et des limites de couverture, mais surtout la soif du bonus. De plus, notons que ces recherches futures permettront aussi de se pencher sur l'analyse du type de dépendance supposé entre ces statistiques d'assurance, ce qui en fait un domaine potentiellement très riche au niveau théorique.

Bibliographie

- ABRAMOWITZ, M. and STEGUN, I. A. (1968). *Handbook of Mathematical Functions*. National Bureau of Standards, Applied Mathematics Series Nr. 55, Washington, D.C.
- AKAIKE, H. (1973). *Information Theory and an Extension of the Maximum Likelihood Principle*. in B.N. Petrov and F. Csaki, eds. Second International Symposium on Information Theory, 267-281, Budapest Akademia Kaido.
- AL-OSH, M. A. and ALZAID, A. A. (1987). First-order Integer-valued Autoregressive (INAR(1)) Process. *Journal of Time Series Analysis*, 8 :261–275.
- AL-OSH, M. A. and ALZAID, A. A. (1993). Some Gamma Processes Based on the Dirichlet-Gamma Transformation. *Communications in Statistics : Stochastic Models*, 9 :123–143.
- ALLISON, P. and WATERMAN, R. (2002). Fixed-Effects Negative Binomial Regression Models. *Sociological Methodology*, 32 :247–265.
- ANDERSEN, E. (1970). Asymptotic Properties of Conditional Maximum Likelihood Estimators. *Journal of the Royal Statistical Society B.*, 32 :283–301.
- ATKINSON, A. (1981). Likelihood Ratios, Posterior Odds and Information Criteria. *Journal of Econometrics*, 16 :15–20.
- BAILEY, R. A. (1963). Insurance Rates with Minimum Bias. *PCAS L.*, pp. 4–11.
- BAILEY, R. A. and SIMON, L. (1960). Two Studies in Automobile Insurance Ratemaking. *ASTIN Bulletin*, 1 :192–212.
- BATES, G. and NEYMAN, J. (1951). Contributions to the Theory of Accident Proneness. II : True or False Contagion. *University of California Publications in Statistics*, pp. 215–253.
- BERKHOUT, P. and PLUG, E. (2004). A Bivariate Poisson Count Data Model using Conditional Probabilities. *Statistica Neerlandica*, 58 :349–364.

- BESSON, J.-L. and PARTRAT, C. (1992). Trend et systèmes de bonus-malus. *ASTIN Bulletin*, 22 :11–31.
- BÜHLMANN, H. (1967). Experience Rating and Credibility. *ASTIN Bulletin*, 4 :199–207.
- BÜHLMANN, H. and STRAUB, E. (1970). Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, 70 :111–133.
- BOCKENHOLT, U. (1999). Mixed INAR(1) Poisson Regressions Models : Analyzing Heterogeneity and Serial Dependence in Longitudinal Count Data. *Journal of Econometrics*, 89 :317–338.
- BOCKENHOLT, U. (2003). Analysing State Dependences in Emotional Experiences by Dynamic Count Data Models. *Applied Statistics*, 52 :213–226.
- BORTKEWISCH, L. v. (1898). *Das Gesetz der kleinen Zahlen*. Teubner Verlag, Leipzig.
- BOUCHER, J.-P. and DENUIT, M. (2006). Fixed versus Random Effects in Poisson Regression Models for Claim Counts : Case Study with Motor Insurance. *ASTIN Bulletin*, 36 :285–301.
- BOUCHER, J.-P. and DENUIT, M. (2007a). Crédibilité linéaire multivariée utilisant le nombre de périodes avec réclamations : modèles de Poisson, modèles à barrière et modèles gonflés à zéro. *Soumis pour publication à Assurances et gestion des risques*.
- BOUCHER, J.-P. and DENUIT, M. (2007b). Credibility Premiums for the Zero-Inflated Model and New Hunger for Bonus Interpretation. *Soumis pour publication à Accident Analysis & Prevention*.
- BOUCHER, J.-P., DENUIT, M., and GUILLÉN, M. (2006a). Independent and Correlated Random Effects for Hurdle Models Applied to Panel Data Count. *Working Paper*.
- BOUCHER, J.-P., DENUIT, M., and GUILLÉN, M. (2006b). Models of Insurance Claim Counts with Time Dependence Based on Generalisation of Poisson and Negative Binomial Distributions. *Working Paper*.
- BOUCHER, J.-P., DENUIT, M., and GUILLÉN, M. (2006c). Risk Classification for Claim Counts : A Comparative Analysis of Various Zero-Inflated Mixed Poisson and Hurdle Models. *Working Paper*.
- BOUCHER, J.-P., DENUIT, M., and GUILLÉN, M. (2006d). Zero-inflated Poisson Models for Panel Data. *Working Paper*.

- BOYER, M., DIONNE, G., and VANASSE, C. (1992). *Econometric Models of Accident Distribution*. in Contributions to Insurance Economics, G. Dionne, Boston : Kluwer Academic Press.
- BRADLOW, E., FADER, P., ADRIAN, M., and MCSHANE, B. (2006). Count Models Based on Weibull Interarrival Times. *Working Paper, Wharton School, University of Pennsylvania*.
- BRÄNNAS, K. (1995). Explanatory Variables in the AR(1) Model. *Umea Economic Studies No. 381, University of Umeå*.
- BRÄNNAS, K. and ROSENQVIST, G. (1994). Semiparametric Estimation of Heterogeneous Count Data Model. *European Journal of Operational Research*, 76 :247–258.
- BROECK, J. v. D. (1995). A Score Test for Zero Inflation in a Poisson Distribution. *Biometrics*, 51 :738–743.
- BROWN, R. L. (1988). Minimum Bias with Generalized Linear Models. *PCAS*, LXXV :187–217.
- BUHLMANN, H. and GISLER, A. (2005). *A Course in Credibility Theory and its Applications*. Springer-Verlag, Berlin.
- BURNHAM, K. and ANDERSON, D. (2002). *Model Selection and Multi-Model Inference : A Practical Information-Theoretic Approach*. Springer-Verlag, New York.
- CAMERON, A., LI, T., TRIVEDI, P., and ZIMMER, D. (2004). Modelling the Differences in Counted Outcomes Using Bivariate Copula Models with Application to Mismeasured Counts. *Econometrics Journal*, 7 :566–584.
- CAMERON, A. and TRIVEDI, P. (1986). Econometric Models Based on Count Data : Comparisons and Applications of some Estimators. *Journal of Applied Econometrics*, 46 :347–364.
- CAMERON, A. C. and TRIVEDI, P. K. (1998). *Regression Analysis of Count Data*. New York : Cambridge University Press.
- CHANT, D. (1974). On Asymptotic Tests of Composite Hypotheses in Nonstandard Conditions. *Biometrika*, 62 :291–298.
- CHERNOFF, H. (1954). On the Distribution of the Log-likelihood Ratio. *Ann. Math. Statist.*, 25 :573–578.

- CHOW, G. (1981). A Comparison of the Information and Posterior Probability Criteria for Model Selection. *Journal of Econometrics*, 16 :21–33.
- CONSUL, P. (1989). *Generalized Poisson Distributions : Properties and Applications*. New York, Marcel Dekker.
- COX, D. (1961). *Tests of Separate Families of Hypotheses*. in Proceedings of the Fourth Berkeley Symposium of Mathematical Statistics and Probability, 105-123, Berkeley, University of California Press.
- CREPON, B. and DUGUET, E. (1997). Research and Development, Competition and Innovation Pseudo-maximum Likelihood and Simulated Maximum Likelihood Methods Applied to Count Data Models with Heterogeneity. *Journal of Econometrics*, 79 :355–378.
- DEAN, C. (1992). Testing for Overdispersion in Poisson and Binomial Regression Models. *Journal of the American Statistical Association*, 87 :451–457.
- DEAN, C., LAWLESS, J., and WILLMOT, G. (1989). A Mixed Poisson-Inverse Gaussian Regression Model. *Canadian Journal of Statistics*, 17 :171–182.
- DEB, P. and TRIVEDI, P. (2002). The Structure of Demand for Health Care : Latent Class versus Two-Part Models. *Journal of Health Economics*, 21 :601–625.
- DELGADO, M. and KNIESNER, T. (1997). Count Data Models with Variance of Unknown Form : An Application to a Hedonic Model of Worker Absenteeism. *The Review of Economics and Statistics*, 79 :41–49.
- DENUIT, M. (1997). A New Distribution of Poisson-Type for the number of Claims. *ASTIN Bulletin*, 27 :229–242.
- DENUIT, M. and DHAENE, J. (2001). Bonus-Malus Scales using Exponential Loss Functions. *German Actuarial Bulletin*, 25 :13–27.
- DENUIT, M. and LAMBERT, P. (2001). Smotthed NPML Estimation of the Risk Distribution Underlying Bonus-Malus System. *Proceedings of the CAS, No.169*.
- DENUIT, M. and LANG, S. (2004). Nonlife Ratemaking with Bayesian GAM's. *Insurance : Mathematics and Economics*, 35 :627–647.

- DENUIT, M., MARÉCHAL, X., PITREBOIS, S., and WALHIN, J.-F. (2007). *Actuarial Modelling of Claim Counts : Risk Classification, Credibility and Bonus-Malus Scales*. Wiley : New York.
- DENUIT, M., PITREBOIS, S., and WALHIN, J.-F. (2003). Tarification Automobile sur Données de Panel. *Bulletin of the Swiss Association of Actuaries*, pp. 51–81.
- DIONNE, G., ARTIS, M., and GUILLÉN, M. (1996). Count Data Models for a Credit Scoring System. *Journal of Empirical Finance*, 3 :303–325.
- DIONNE, G. and VANASSE, C. (1989). A Generalization of Automobile Insurance Rating Models : The Negative Binomial Distribution with Regression Component. *ASTIN Bulletin*, 19 :199–212.
- DIONNE, G. and VANASSE, C. (1992). Automobile Insurance Ratemaking in the presence of asymmetrical information. *Journal of Applied Econometrics*, 7 :149–165.
- EGGENBERGER, F. and POLYA, G. (1923). Über die Statistik Verketteter Vorgänge. *Zeitschrift für Angewandte Mathematik und Mechanik (Journal of Applied Mathematics and Mechanics)*, 1 :279–289.
- FELDBLUM, S. and BROSIUS, E. J. (2002). The Minimum Bias Procedure : A Practitioner's Guide. *Casualty Actuarial Society Forum*, Fall :591–684.
- FELLER, W. (1943). On a General Class of 'Contagious' Distributions. *Annals of Mathematical Statistics*, 14 :389–400.
- FERREIRA, J. (1977). Identifying Equitable Insurance Premiums for Risk Classes : an Alternative to the Classical Approach. *Lecture presented at the 23th international meeting of the Institute of Management Sciences, Athens, Greece*.
- FREELAND, R. K. and MCCABE, B. P. M. (2004). Forecasting Discrete Valued Low Count Time Series. *International Journal of Forecasting*, 20 :427–434.
- FREES, E., YOUNG, V., and LUO, Y. (1999). A longitudinal data analysis interpretation of credibility models. *Insurance : Mathematics and Economics*, 24 :229–247.
- FREES, E. W. and VALDEZ, E. A. (1999). Understanding Relationships Using Copulas. *North American Actuarial Journal*, 2 :1–25.

- FRENCH, E. and JONES, J. B. (2004). On the Distribution and Dynamics of Health Care Costs. *Journal of Applied Econometrics*, 19 :705–722.
- GALLANT, A. and NYCHKA, D. (1987). Semi-Nonparametric Maximum Likelihood Estimation. *Econometrica*, 55 :363–390.
- GERBER, H. (1972). Discussion of Hewitt(1971). *Proceedings of the Casualty Actuarial Society*, 58 :25–27.
- GERBER, H. and JONES, D. (1975). Credibility Formulas of the Updating Type. *Transactions of the Society of Actuaries*.
- GOLDEN, R. (2003). Discrepancy Risk Model Selection Test for Comparing Possibly Misspecified or Nonnested Models. *Psychometrika*, 68 :229–249.
- GOSSIAUX, A. M. and LEMAIRE, J. (1981). Méthodes d’ajustement de distributions de sinistres. *Bulletin of the Swiss Association of Actuaries*, pp. 87–95.
- GOULET, V., FORGUES, A., and LU, J. (2006). Credibility for Severity Revisited. *North American Actuarial Journal*, 10 :49–62.
- GOURIEROUX, C. (1999). The Econometrics of Risk Classification in Insurance. *The Geneva Papers on Risk and Insurance Theory*, 24 :119–137.
- GOURIEROUX, C. and JASIAK, J. (2004). Heterogeneous INAR(1) Model with Application to Car Insurance. *Insurance : Mathematics and Economics*, 34 :177–192.
- GOURIEROUX, C. and MONFORT, A. (1995). *Statistics and Econometric Models, vol. 1 and 2*. Cambridge University Press.
- GOURIEROUX, C., MONFORT, A., and TROGNON, A. (1984a). Pseudo-Maximum Likelihood Methods : Application to Poisson Models. *Econometrica*, 52 :701–720.
- GOURIEROUX, C., MONFORT, A., and TROGNON, A. (1984b). Pseudo-Maximum Likelihood Methods : Theory. *Econometrica*, 52 :681–700.
- GREENWOOD, M. and YULE, G. (1920). An Inquiry into the Nature of Frequency Distributions Representative of multiple Happenings with Particular References to the Occurrence of Multiple Attacks of Disease or Repeated Accidents. *Journal of the Royal Statistical Society A*, 83 :255–279.

- GROGGER, J. and CARSON, R. (1991). Models for Truncated Counts. *Journal of Applied Econometrics*, 6 :225–238.
- GROOTENDORST, P. (1995). A Comparison of Alternative Models of Prescription Drug Utilization. *Health Economics*, 4 :183–198.
- GURMU, S. (1998). Generalized Hurdle Count Data Regression Models. *Economics Letters*, 58 :263–268.
- GURMU, S., RILSTONE, P., and STERN, S. (1999). Semiparametric Estimation of Count Regression Models. *Journal of Econometrics*, 88(1) :123–150.
- GURMU, S. and TRIVEDI, P. (1992). Overdispersion Tests for Truncated Poisson Regression Models. *Journal of Econometrics*, 54 :347–370.
- HACHEMEISTER, C. (1975). *Credibility for Regression Models with Applications to Trend*. in *Credibility : Theory and Applications*, ed. P. M. Kahn, Academic Press, New York, 129–163.
- HALL, D. and BERENHAUT, K. (2002). Score Tests for Heterogeneity and Overdispersion in Zero-inflated Poisson and Binomial Regression Models. *The Canadian Journal of Statistics*, 30 :1–16.
- HAUSMAN, J. (1978). Specification Tests in Econometrics. *Econometrica*, 46 :1251–1271.
- HAUSMAN, J., HALL, B., and GRILICHES, Z. (1984). Econometric Models for Count Data with Application to the Patents-R and D Relationship. *Econometrica*, 52 :909–938.
- HECKMAN, J. and SINGER, B. (1984). A Method of Minimizing the Distributional Impact in Econometric Models for Duration Data. *Econometrica*, 52 :271–320.
- HINDE, J. (1982). Compound Poisson Regression Models. in R. Gilchrist, ed., *GLIM 82 : Proceeding of the International Conference on Generalised Linear Models*, New York, Springer-Verlag.
- HOLGATE, P. (1964). Estimation for the Bivariate Poisson distribution. *Biometrika*, 51 :241–245.
- HOLGATE, P. (1970). The Modality of Some Compound Poisson Distributions. *Biometrika*, 57 :666–667.
- HOLLA, M. (1966). On a Poisson-Inverse Gaussian Distribution. *Metrika*, 11 :115–121.

- HOLLER, K. D., SOMMER, D., and TRAHAI, G. (1999). Something Old, Something New in Classification Ratemaking with a Novel Use of GLMs for Credit Insurance. *Casualty Actuarial Society Forum*, Winter :31–84.
- HSIAO, C. (2003). *Analysis of Panel Data*. Cambridge : Cambridge University Press, 2nd ed.
- ISLAM, M. and CONSUL, P. (1992). A Probabilistic Model for Automobile Claims. *Bulletin of the Swiss Association of Actuaries*, pp. 85–93.
- JEWELL, W. (1975). The Use of Collateral Data in Credibility Theory : A Hierarchical Model. *Giornale dell Istituto Italiano degli Attuari*, 38 :1–16.
- JOE, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, New York.
- JOHNSON, N., KOTZ, S., and BALAKRISHNAN, N. (1996). *Discrete Multivariate Distributions*. New York : Wiley, 2nd ed.
- JOHNSON, N., KOTZ, S., and KEMP, A. (1992). *Univariate Discrete Distributions*. New York : Wiley, 2nd ed.
- JORGENSEN, B. (1997). *The Theory of Dispersion Models*. London, Chapman and Hall.
- KARLIS, D. and MELIGKOTSIDOU, L. (2005). Multivariate Poisson Regression with Covariance Structure. *Statistics and Computing*, 15 :255–265.
- KARLIS, D. and XEKALAKI, E. (2006). The Polygonal Distribution. *International Conference on Mathematical and Statistical Modeling in Honor of Enrique Castillo. June 28-30, 2006*.
- KEIDING, N., C., A., and P., F. (1995). The Cox Regression Model for Claims Data in Non-Life Insurance. *Astin Bulletin*, 28 :95–118.
- KIMBERLING, C. (1974). A Probabilistic Interpretation of Complete Monotonicity. *Aequationes Math.*, 10 :152–164.
- KLUGMAN, S., PANJER, H., and WILLMOT, G. (2004). *Loss Models : From Data to Decisions*. Wiley, New York.
- KOCHERLATOKA, S. and KOCHERLATOKA, K. (1992). *Bivariate Discrete Distributions*. Marcel Dekker : New York.

- KUHA, J. (2004). AIC and BIC Comparisons of Assumptions and Performance. *Sociological Methods and Research*, 33 :188–229.
- LAIRD, N. (1978). Non-Parametric Maximum Likelihood Estimation of a Mixing Distribution. *Journal of the American Statistical Association*, 73 :805–811.
- LAMBERT, D. (1992). Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing. *Technometrics*, 34 :1–14.
- LAWLESS, J. (1987). Negative Binomial and Mixed Poisson Regression. *The Canadian Journal of Statistics* 15 : 209–225, 15 :209–225.
- LEE, L. (1983). Generalized Econometric Models with Selectivity. *Econometrica*, 51 :507–512.
- LEMAIRE, J. (1977). La Soif du Bonus. *ASTIN Bulletin*, 9 :181–190.
- LEMAIRE, J. (1995). *Bonus-Malus Systems in Automobile Insurance*. Boston : Kluwer Academic Publishers.
- LIANG, K. and ZEGER, S. (1986). Longitudinal Data Analysis using Generalized Linear Models. *Biometrika*, 73 :13–22.
- MARSHALL, A. and OLKIN, I. (1990). *Multivariate Distributions Generated from Mixtures of Convolution and Product Families*. in Topics in Statistical Dependence, H.W. Block, A.R. Sampson and T.H. Savits, Lecture Notes-Monograph Series, Vol. 16 : 371–393.
- MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*. London : Chapman and Hall, 2nd ed.
- MCKENZIE, E. (1986). Autoregressive Moving-Average Processes with Negative-Binomial and Geometric Marginal Distributions. *Advances in Applied Probability*, 18 :679–705.
- MCKENZIE, E. (1988). Some ARMA Models for Dependent Sequence of Poisson Counts. *Advances in Applied Probability*, 20 :822–835.
- MILDENHALL, S. (1999). A Systematic Relationship between Minimum Bias and Generalized Linear Models. *Proceedings of the Casualty Actuarial Society*, LXXXVI.
- MORAN, P. (1971). Maximum Likelihood Estimation in Non-Standard Conditions. *Proceeding of the Cambridge Philosophical Society*, 70 :441–450.

- MORILLO, I. and BERMUDEZ, L. (2003). Bonus-Malus System using an Exponential Loss Function with an Inverse Gaussian Distribution. *Insurance : Mathematics and Economics*, 33 :49–57.
- MULLAHY, J. (1986). Specification and Testing in some Modified Count Data Models. *Journal of Econometrics*, 33 :341–365.
- MULLAHY, J. (1997). Instrumental Variable Estimation of Count Data Models : Applications to Models of Cigarette Smoking Behavior. *Review of Economics and Statistics*, 79 :586–593.
- MULLAHY, J. (1998). Much Ado about Two : Reconsidering Retransformation and the Two-Part Model in Health Econometrics. *Journal of Health Economics*, 17 :247–281.
- MUNDLAK, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46 :69–85.
- NELDER, J. and WEDDERBURN, R. (1972). Generalized Linear Models. *Journal of the Royal Society. Series 1*, 135 :370–384.
- NELSEN, R. (1999). *An Introduction to Copulas*. Springer-Verlag, New York.
- NELSON, K., LIPSITZ, S., FITZMAURICE, G., IBRAHIM, J., PARZEN, M., and STRAWDERMAN, R. (2006). Use of the Probability Integral Transformation to Fit Nonlinear Mixed-Effects Models with Nonnormal Random Effects. *Journal of Computational and Graphical Statistics*, 15 :39–57.
- NEYMAN, J. (1965). *Certain Chance Mecanisms Involving Discrete Distributions*. in G.N. Patil, ed., *Classical and Contagious Discrete Distributions*, 1-14, Calcutta, Statistical Publishing House.
- PALMGREN, J. (1981). The Fisher Information Matrix for Log-Linear Models Arguing Conditionally in Observed Explanatory Variables. *Biometrika*, 68 :563–566.
- PESARAN, M. (1974). On the General Problem of Model Selection. *Review of Economic Studies*, 41 :153–171.
- PHILIPSON, C. (1960). The Swedish System of Bonus. *ASTIN Bulletin*, 1 :134–141.
- PINQUET, J. (1997). Estimating and Testing for Time Dependent Heterogeneity in a Poisson Model. *Thema 9705*.
- PINQUET, J. (2000). Experience Rating through Heterogeneous Models. In *Handbook of Insurance*, edited by G. Dionne. Kluwer Academic Publishers, pp. 459-500.

- POHLMEIER, W. and ULRICH, V. (1995). An Econometric Model of the Two-Part Decision Making Process in the Demand for Health Care. *Journal of Human Ressources*, 30 :339–361.
- POISSON, S.-D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et matière civile*. Paris, Bachelier.
- RIDOUT, M., HINDE, J., and DEMETRIO, C. G. B. (2001). A Score Test for Testing a Zero-Inflated Poisson Regression Model Against Zero-Inflated Negative Binomial Alternatives. *Biometrics*, 57 :219–223.
- RIVERS, D. and VUONG, Q. (2002). Model Selection Tests for Nonlinear Dynamic Models. *Econometrics Journal*, 5 :1–39.
- ROBINSON, P. (1987). Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form. *Econometrica*, 55(4) :875–891.
- ROSENBLATT, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, 27 :832–837.
- SANTOS SILVA, J. (2003). A Note on the Estimation of Mixture Models under Exogeneous Sampling. *Econometrics Journal*, 6 :46–52.
- SANTOS SILVA, J. and WINDMEIJER, F. (2001). Two-part Multiple Spell Models for Health Care Demand. *Journal of Econometrics*, 104 :67–89.
- SCHWARTZ, E. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6 :461–464.
- SCOLLNIK, D. (2001). Actuarial Modeling With MCMC and BUGS. *North American Actuarial Journal*, 5 :96–125.
- SHAKED, M. (1980). On Mixtures from Exponential Families. *Journal of Royal Statistical Society B*, 42 :415–433.
- SHOUKRI, M., ASYALI, M., VANDORP, R., and KELTON, D. (2004). The Poisson Inverse Gaussian Regression Model in the Analysis of Clustered Counts Data. *Journal of Data Science*, 2 :17–32.
- SIMAR, L. (1976). Maximum Likelihood Estimation of a Compound Poisson Process. *Annals of Statistics*, 4 :1200–1209.

- SKLAR, A. (1959). Fonctions de Répartition à n Dimensions et leurs Marges. *Publ. Inst. Statist. Univ. Paris*, 8 :229–231.
- SMITH, A. and ROBERTS, G. (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods (with discussion). *Journal of the Royal Statistical Society, Series B*, 55 :3–23.
- SONG, J. (2005). Zero-Inflated Poisson Regression to Analyze Lengths of Hospital Stays Adjusting for Intra-Center Correlation. *Communications in Statistics-Simulation and Computation*, 11 :59–67.
- STEUTEL, F. W. and HARN, K. (1979). Discrete Analogues of Self-Decomposability and Stability. *Annals of Probability*, 7 :893–899.
- STODDART, G. and BARER, M. (1981). *Analyses of the Demand and Utilization Through Episodes of Medical Care*. in van der Gaag, J., Perlman, M. (Eds.), Health, Economics and health Economics. Elsevier, Amsterdam.
- SUNDT, B. (1988). Credibility Estimators with Geometric Weights. *Insurance : Mathematics and Economics*, 7 :113–122.
- TREMBLAY, L. (1992). Using the Poisson Inverse Gaussian in Bonus-Malus Systems. *ASTIN Bulletin*, 22 :97–106.
- VAN OPHEM, H. (1999). A General Method to Estimate Correlated Discrete Random Variables. *Econometric Theory*, 15 :228–237.
- VANDENHENDE, F. and LAMBERT, P. (2000). Modeling Repeated Ordered Categorical Data using Copulas. *Institute of Statistics, UCL, Belgium*.
- VUONG, Q. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57 :307–333.
- WALHIN, J.-F. (2001). Bivariate ZIP Models. *Biometrical Journal*, 43 :147–160.
- WALHIN, J.-F. and PARIS, J. (1999). Using Mixed Poisson Distributions in Connection with Bonus-Malus Systems. *ASTIN Bulletin*, 29 :81–99.

- WALHIN, J.-F. and PARIS, J. (2000). The True Claim Amount and Frequency Distributions Within a Bonus-Malus System. *ASTIN Bulletin*, 30 :391–403.
- WEDEL, M., DESARBO, W., BULT, J., and RAMASWAMY, V. (1993). A Latent Class Poisson Regression Model for Heterogeneous Count Data. *Journal of Applied Econometrics*, 8 :397–411.
- WILLMOT, G. (1987). The Poisson-Inverse Gaussian Distribution as an Alternative to the Negative Binomial. *Scandinavian Actuarial Journal*, 87 :113–127.
- WINKELMANN, R. (2000). Seemingly Unrelated Negative Binomial Distribution. *Oxford Bulletin of Economics and Statistics*, 62 :553–560.
- WINKELMANN, R. (2003a). *Econometric of Count Data*. Springer-Verlag, Berlin, 4th ed.
- WINKELMANN, R. (2003b). Health Care Reform and the Number of Doctor Visits - an Econometric Analysis. *Journal of Applied Econometrics*, 19 :455–472.
- WINKELMANN, R. and ZIMMERMANN, K. (1991). A New Approach for Modeling Economic Count Data. *Economics Letters*, 37 :139–143.
- WINKELMANN, R. and ZIMMERMANN, K. (1995). Recent Development in Count Data Modelling : Theory and Application. *Journal of Economic Surveys*, 9 :1–24.
- YIP, K. and YAU, K. (2005). On Modeling claim Frequency Data in General Insurance With Extra Zeros. *Insurance : Mathematics and Economics*, 36 :153–163.
- YOUNG, V. (1996). Credibility and Persistency. *ASTIN Bulletin*, 26 :53–69.
- YOUNG, V. (1998). Credibility using a Loss Function from Spline Theory : Parametric Models with a One-Dimensional Sufficient Statistic. *North American Actuarial Journal*, 2 :101–117.
- YOUNG, V. and DE VYLDER, F. (2000). Credibility in Favor of Unlucky Insureds. *North American Actuarial Journal*, 4 :107–113.
- ZEGER, S. and LIANG, K. (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, 42 :121–130.