Essais en économétrie de l'assurance non-vie

Jean Pinquet

Habilitation à diriger des recherches

Essais en économétrie de l'assurance non-vie

soutenue par **Jean Pinquet**

discipline: Sciences Economiques

Remerciements

Je tiens d'abord à remercier Pierre Picard, qui m'a incité à commencer des activités de recherches sur le tard. Je remercie Georges Dionne, qui m'a suggéré les premières pistes de recherche et avec qui j'ai régulièrement collaboré depuis. Je veux aussi remercier Pierre-André Chiappori, Christian Gouriéroux et Bernard Salanié avec qui les interactions ont été stimulantes. Je remercie également Montserrat Guillén qui m'a permis d'accéder à de nombreuses bases de données d'assurance, avec qui j'ai mené à bien plusieurs projets de recherches et grâce à qui je connais Barcelone comme ma poche. Je remercie enfin Michel Denuit, qui m'a donné l'occasion de faire connaître mes travaux auprès de la communauté des actuaires belges.

Je veux remercier mes coauteurs non encore cités, à savoir Jaap Abbring, Mercedes Ayuso, Catalina Bolancé, Natacha Brouhns, Denise Desjardins, James Heckman, Mathieu Maurice et Charles Vanasse. Je veux aussi mentionner mes collègues ou ex-collègues à Nanterre: Arnaud Lefranc, Nathalie Fombaron, Jean-Marc Bourgeon, Marie-Cécile Fagart, Bénédicte Coestier, ainsi que Sabine Lemoyne et Eric Strobl à l'Ecole Polytechnique. Enfin, je souhaite remercier les membres du jury de cette habilitation à diriger des recherches, pour la considération qu'ils ont apportée à mon travail. J'exprime toute ma reconnaissance à Arnaud Lefranc, qui a bien voulu parrainer cette HDR.

CURRICULUM VITAE

Jean Pinquet

Etat civil

Né le 14 décembre 1956 à Mazingarbe (62)

Célibataire

23 rue des écoles

94000 Créteil

Tél: 01 49 81 72 45 / 06 32 62 99 19

UFR SEGMI

Université Paris-Ouest Nanterre La Défense (Paris 10)

200, avenue de la République

92001 Nanterre Cedex

Tél: 01 40 97 47 58

E-mail: pinquet@u-paris10.fr

Situation actuelle

Agrégé hors classe de mathématiques, assistant à l'UFR SEGMI.

Chercheur au département d'économie de l'Ecole Polytechnique

Cursus universitaire

1992: diplôme d'actuaire de l'Institut des Actuaires Français.

1982: diplôme de l'Ecole Nationale de la Statistique et de l'Administration Economique.

1979: agrégation de mathématiques (reçu $50^{ème}$).

1976-1979 et 1980-1982: élève à l'ENS Cachan (mathématiques).

1973-1976: classes préparatoires au lycée Louis-Le-Grand

1973: Baccalauréat C (mention bien)

Divers

Figure dans le *"Who's Who in Science and Engineering"* et le *"Who's Who in the World"*

Classé au quantile 15% dans l'indice de notoriété des économistes français répertoriés dans la base RePEc.

Responsabilités courantes

2007- : co-éditeur de la revue *The Journal of Risk and Insurance.*

2007- : membre du comité de pilotage de la chaire "Assurance et risques majeurs" (financée par Axa et portée par Paris-Dauphine, l'Ecole Polytechnique et l'ENSAE), chercheur associé à la chaire.

Rapporteur pour les revues suivantes (et nombre de papiers rapportés)

Journal of Political Economy (2), Journal of the American Statistical Association (1), Journal of Econometrics (1), The Economic Journal (1), The Journal of Risk and Insurance (20), Geneva Papers on Risk and Insurance: Theory (5), Insurance: Mathematics and Economics (2), Computational Statistics and Data Analysis (1), ASTIN Bulletin (5), Annales d'Economie et de Statistiques (1), Comptes-Rendus de l'Académie des Sciences (1), Geneva Papers on Risk and Insurance: Issues and Practice (1), Belgian Actuarial Bulletin (1).

Activités d'expertise

Evaluation de projets de recherche pour la National Science Foundation (Etats-Unis) et le National Research Council (Canada).

Participation à des contrats de recherche

1994-2005: Chercheur associé à la chaire d'assurance F.F.S.A. Recherches sur les systèmes bonus-malus (publications [7], [8], [9], [11], [12], [13], [14], [28]), puis sur la sécurité routière (cf. projet PREDIT). Autres chercheurs de la chaire: Pierre Picard, Georges Dionne, Jean-Marc Bourgeon.

1997-2001: Contrat avec la Société de l'Assurance Automobile du Québec, sur la tarification de flottes de véhicules (avec Georges Dionne et Denise Desjardins). Publication associée: [10].

1998: Contrat avec la Direction de la Prévision, sur l'assurance chômage des emprunteurs (avec Pierre-André Chiappori). Publication associée: [23].

2001: Invitation à l'université de Chicago par Pierre-André Chiappori sur un contrat "National Science Foundation" portant sur l'économétrie de l'assurance. Autres chercheurs: Jaap Abbring et James Heckman. Publications associées: [5], [6].

2003-2006: Contrat avec le PREDIT (Programme de REcherche et D'Innovation dans les Transports terrestres, Groupe Opérationnel n°3) sur le thème "Prise de risque au volant face aux contrôles et aux sanctions: Une approche en termes d'incitations". Autres chercheurs associés: Pierre Picard, Georges Dionne, Jean-Marc Bourgeon. Publications associées: [1], [19], [26], [27].

2005- : Chercheur associé à la chaire "Assurance et risques majeurs" (financée par Axa et portée par Paris-Dauphine, l'Ecole Polytechnique et l'ENSAE), membre du comité de pilotage. Thèmes de recherches: contrats de long terme et assurance-dépendance, comprtements face au risque d'inondation. Autres chercheurs associés:

Pierre Picard et Christian Gouriéroux (directeurs), Alfred Galichon, Arthur Charpentier. Publications associées: [2], [15], [16], [25].

Participation à des jurys de thèse

Thèse de Maxime Druon, soutenue à Lille 3 le 10 décembre 2007.

Activités d'enseignement

2009- : "Fondements économiques de l'assurance", enseigné en M1 ISIFAR à Paris 10.

2008- : "Assurance et modèles statistiques de durée", enseigné en M2 ISIFAR à Paris 10.

2008- : "Calculs actuariels et financiers", enseigné en M1 d'économie à Paris 10.

2005-2007: "petites classes" d'économétrie à l'Ecole Polytechnique.

2001-2009: "Econométrie de l'assurance", enseigné en M2 GRFA à Paris 10.

2000: "Non-life insurance rating for large portfolios, and applications to motor insurance", à l'Université Libre de Bruxelles et à l'invitation de l'Association Royale des Actuaires Belges.

1999-2000: "Statistique de l'assurance", cours de troisième année à l'ENSAE.

Autres cours et travaux dirigés en économétrie, analyse des données, programmation linéaire, mathématiques et statistiques.

Activités de conseil

2002-2003: Auprès de la MAIF, pour l'évolution de leur système de tarification en assurance automobile.

2001: Auprès de la société Winterthur Belgique (avec Michel Denuit) sur la conception de systèmes bonus-malus multi-garanties.

1990-1993: Auprès de la société d'assurances CARDIF, études statistiques sur

l'assurance automobile et sur l'assurance des emprunteurs.

1987-1990: Auprès du groupe Compagnie Bancaire, études de marketing quantitatif et études statistiques sur l'assurance automobile.

Publications internationales

((*) si auteur correspondant)

---

[1] G. Dionne, J. Pinquet, C. Vanasse et M. Maurice (2009) "Incentive mechanisms for safe driving: A comparative analysis with dynamic data", à paraître dans *The Review of Economics and Statistics*.

[2] M. Guillén et J. Pinquet (*) (2008) "Long-Term Care: risk description of a Spanish portfolio and economic analysis of the timing of insurance purchase", *Geneva Papers on Risk and Insurance: Issues and Practice* **33**, 659-672.

[3] C. Bolancé, M. Guillén et J. Pinquet (*) (2008) "On the link between credibility and frequency premium", *Insurance: Mathematics and Economics* **43**, 209-213.

[4] M. Ayuso, M. Guillén et J. Pinquet (*) (2007) "Selection bias and auditing policies for insurance claims," (2007) *The Journal of Risk and Insurance* **74**, 425-440.

[5] J. Abbring, P.A. Chiappori, J. Heckman et J. Pinquet (2003) "Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish?", *Journal of the European Economic Association, Papers and Proceedings* **1**, 512-521.

[6] J. Abbring, P.A. Chiappori et J. Pinquet (2003) "Moral hazard and dynamic insurance data", *Journal of the European Economic Association* **1**, 767-820.

[7] N. Brouhns, M. Denuit, M. Guillén et J. Pinquet (2003) "Bonus-malus scales in segmented tariffs with stochastic migration between segments", *The Journal of Risk and Insurance* **70**, 577-599. Casualty Actuarial Society Research Award 2003.

[8] C. Bolancé, M. Guillén et J. Pinquet (*) (2003) "Time-varying credibility for frequency risk models: Estimation and tests for autoregressive specifications on the

random effects", *Insurance: Mathematics and Economics* **33**, 273-282.

[9] C. Bolancé, M. Guillén et J. Pinquet (*) (2001) "Allowance for the age of claims in bonus-malus systems", *ASTIN Bulletin* **31**, 337-348.

[10] D. Desjardins, G. Dionne, et J. Pinquet (*) (2001) "Experience rating schemes for fleets of vehicles", (2001) *ASTIN Bulletin* **31**, 81-106.

[11] J. Pinquet (1998) (*) "Designing optimal bonus-malus systems from different types of claims", *ASTIN Bulletin* **28**, 205-220.

[12] J. Pinquet (1997) (*) "Allowance for cost of claims in bonus-malus systems", *ASTIN Bulletin* **27**, 33-57.

Chapitres de livres

---

[13] J. Pinquet (2000) (*) "Experience rating through heterogeneous models", *Handbook of Insurance* 459-500, Kluwer Academic Publishers. Huebner International Series on Risk, Insurance and Economic Security (Editeur: Georges Dionne).

[14] J. Pinquet (1999) (*) "Allowance for hidden information by heterogeneous models, and applications to insurance rating", *Automobile Insurance* 47-78, Kluwer Academic Publishers. Huebner International Series on Risk, Insurance and Economic Security (Editeur: Georges Dionne).

Autres publications

---

[15] M. Ayuso, M. Guillén et J. Pinquet (*) (2009) "Long-term insurance: a case study", *Pravartak*, 241-249.

[16] J. Pinquet (2009) (*) "Quel avenir pour l'assurance dépendance? Leçons de l'expérience américaine", *Risques* **78** 131-134.

[17] M. Ayuso, M. Guillén et J. Pinquet (*) (2007) "Economic and statistical models on insurance fraud", *Pravartak* 123-127.

[18] P. Picard et J. Pinquet (2006) "Régulation des banques et des assurances :

vraies et fausses analogies", *Risques* **66**, 70-76.

[19] G. Dionne et J. Pinquet (2006) (*) "Au Québec, l'assureur compte les points", *Circuler autrement* **132**, 16-17.

[20] J. Pinquet (2003) (*) "Prise en compte de l'ancienneté des périodes dans les modèles de risque en fréquence : aspects théoriques et applications à la tarification des risques en assurance automobile", *Revue de la Société Française de Statistique* **3**, 7-27.

[21] D. Desjardins, G. Dionne, et J. Pinquet (2000) "Modèle d'évaluation du risque d'accident des transporteurs routiers", *Routes et Transports* **1**, 10-20.

[22] D. Desjardins, G. Dionne, et J. Pinquet (1999) "L'Evaluation des Risques d'Accidents de Transporteurs Routiers: des Résultats Préliminaires", *Assurances* **3**, 449-477.

[23] P.A. Chiappori et J. Pinquet (1999) "L'Assurance Chômage des Emprunteurs", *Revue Française d'Economie* **14**, 91-115.

[24] J. Pinquet (1999) (*) "Une analyse des systèmes bonus-malus en assurance automobile", *Assurances* **2**, 241-249.

Documents de travail divers

---

[25] M. Ayuso, M. Guillén et J. Pinquet (*) (2009) "Commitment and lapse behavior in long-term insurance: a case study", accessible à l'URL `http://hal.archives-ouvertes.fr/hal-00374303/fr/`, soumis pour publication à *The Journal of Risk and Insurance*.

[26] G. Dionne, J. Pinquet, C. Vanasse et M. Maurice (2007) "Point-record incentives, asymmetric information and dynamic data". Version antérieure de [1].

[27] G. Dionne et J. Pinquet (2005): "Mesure des effets incitatifs à la prudence au volant créés par les sanctions et évaluation du pouvoir prédictif des infractions sur le risque routier", accessible à l'URL

http://neumann.hec.ca/gestiondesrisques/05-06.pdf

[28] J. Pinquet (2002) "Norme d'information commune aux acteurs d'un marché d'assurance automobile: systèmes bonus-malus ou historiques individuels?", Journée ISFA-ISA à la FFSA.

Papiers présentés sur invitation

---

[1] "Allowance for Hidden Information by Heterogeneous Models, and Applications to Insurance Rating," (1997), colloque international sur l'assurance automobile, Montréal (Canada)

[2] **"**Experience Rating through Heterogeneous Models" (1997), journées d'études sur l'assurance du Center for Economic Policy Research à Ponte Corvo (Sardaigne)

[3] **"**Experience Rating through Heterogeneous Models" (1998), journée franco-suisse ASTIN, Institut de Sciences Actuarielles de H.E.C. Lausanne (Suisse)

[4] **"**Systèmes bonus-malus pour les flottes de véhicules" (1999), à la Fédération Française des Sociétés d'Assurance (Paris).

[5] "Tarification *a priori* et *a posteriori* en Assurance. Pourquoi? Comment? Une Application aux Flottes de Véhicules" (2000), séminaire de statistiques de l'Université Catholique de Louvain.

[6] "Tarification des flottes de véhicules: une étude sur données québécoises" (2000), journée d'études de la Société Française de Statistiques à Paris.

[7] "Prise en compte de la date des sinistres dans les systèmes bonus-malus: mesure et implications sur les effets incitatifs à la prudence" (2000), journée d'études à la Fédération Française des Sociétés d'Assurance.

[8] "Prise en compte de l'ancienneté des périodes dans les modèles de crédibilité, avec une application à l'assurance automobile" (2002), journée franco-suisse ASTIN, Institut de Sciences Financières et d'Assurance, Lyon.

[9] "Prise en compte de l'ancienneté des sinistres dans les modèles de crédibil-

ité, avec une application à l'assurance automobile" (2002), colloque de la Société Française de Statistique, Niort.

[10] Interventions à des séminaires PREDIT sur le projet sécurité routière: les 4 décembre 2003 (CNIT), 27 janvier 2004 (Grande Arche), 16 septembre 2004 (Grande Arche), 9 et 10 mars 2006 (Marly-Le-Roi).

[11] "Bonus-Malus Scales in Segmented Tariffs with Stochastic Migration between Segments" (2005) congrès annuel de la Casualty Actuarial Society, Baltimore (Maryland, Etats-Unis), pour réceptionner le prix *Casualty Actuarial Society Research Award 2003.* (obtenu avec Natacha Brouhns, Michel Denuit et Montserrat Guillén).

[12] "Point-Record Incentives, Road Safety and Dynamic Data" (2007) à l'université de Barcelone.

[13] "Quel avenir pour l'assurance dépendance? Leçons de l'expérience américaine" (2008), Conférence "La dépendance: que sait-on vraiment?", à l'Assemblée Nationale, à l'invitation de la chaire "Risques et chances de la transition démographique".

Invitations

---

[1] Université de Montréal (1998), pour une collaboration sur un contrat d'études relatif à la tarification de flottes de véhicules.

[2] Université de Montréal (2000), pour des travaux relatifs à la tarification de flottes de véhicules et à l'analyse d'un panel long de contrats individuels d'assurance automobile.

[3] Université de Barcelone (2000), à l'invitation de Montserrat Guillén.

[4] Université Libre de Bruxelles (2000), pour un cours intitulé "Non-life insurance rating for large portfolios, and applications to motor insurance", à l'invitation de l'Association Royale des Actuaires Belges.

[5] Université de Chicago (2001), pour un projet de recherche financé par la National Science Foundation, en collaboration avec Pierre-André Chiappori, James Heckman et Jaap Abbring.

[6] Université de Montréal (2001), à l'invitation de Georges Dionne (même sujet qu'en 2000).

[7] Université de Montréal (2003), à l'invitation de Georges Dionne (projet PREDIT relatif à la sécurité routière).

[8] Université de Barcelone (2004), à l'invitation de Montserrat Guillén.

[9] Université de Barcelone (2005), à l'invitation de Montserrat Guillén.

[10] Université de Barcelone (2006), à l'invitation de Montserrat Guillén.

[11] Université de Barcelone (2007), à l'invitation de Montserrat Guillén.

[12] Université de Barcelone (2008), à l'invitation de Montserrat Guillén.

[13] Wharton School (Université de Pennsylvanie) (2009), à l'invitation du département d'assurance.

Papiers acceptés à des congrès internationaux

---

[1] "Allowance for Cost of Claims in Bonus-Malus Systems", (1996) congrès ASTIN (Copenhague).

[2] "Unexplained Heterogeneity", (1997) congrès d'été des économètres américains (Pasadena, Californie)

[3] "Estimating and Testing for Time-Dependent Heterogeneity in a Poisson Model", (1997) congrès sur l'analyse des données de panel (Paris, la Sorbonne)

[4] "Allowance for Cost of Claims in Bonus-Malus Systems", (1997) congrès de l'International Statistical Institute à Istanbul (Turquie)

[5] "Experience Rating Schemes for Fleets of Vehicles" (2000), congrès de la Risk Theory Society, Minneapolis (Minnesota)

[6] "Linear Credibility Predictors for the Pure Premium of an Insurance Con-

tract" (2000), congrès de la revue "Insurance: Mathematics and Economics", à Barcelone.

[7] "Experience Rating Schemes for Fleets of Vehicles" (2000), congrès ASTIN (Porto Cervo, Sardaigne).

[8] "Prise en compte de la date des sinistres dans la tarification des contrats d'assurance" (2001), congrès de la Société Canadienne de Sciences Economiques à Québec, Canada.

[9] "Time-varying credibility for frequency risk models" (2001), congrès "Insurance: Mathematics and Economics", à State College (Etats-Unis).

[10] "Time-Varying Credibility for Frequency Risk Models: Estimation and Tests for Autoregressive Specifications on the Random Effects" (2002), congrès "Insurance: Mathematics and Economics", à Lisbonne (Portugal).

[11] "Selection bias and auditing policies for insurance claims" (2005), congrès ARIA à Salt Lake City (Etats-Unis).

[12] "Selection bias and auditing policies for insurance claims" (2005), congrès ASTIN à Zürich (Suisse).

[13] "Selection bias and auditing policies for insurance claims" (2006), congrès IME à Leuven (Belgique).

[14] "Point-Record Incentives, Road Safety and Dynamic Data" (2007), congrès ARIA à Québec (Canada).

[15] "Long-Term Care: risk description of a Spanish portfolio and economic analysis of the timing of insurance purchase" (2008), congrès EGRIE à Toulouse.

[16] "Long-Term Care: risk description of a Spanish portfolio and economic analysis of the timing of insurance purchase" (2008), Longevity Risk Conference, Amsterdam (Pays-Bas).

PRESENTATION DES TRAVAUX[1]

Jean Pinquet

Essais en économétrie de l'assurance non-vie

[1]Les numéros des publications cosignées par l'auteur et figurant dans la synthèse qui suit sont celles du CV, pages 7 à 9. Les autres références figurent à la fin de cette synthèse.

1 Introduction

Les travaux présentés ci-dessous sont le reflet de mon activité de recherches effectuée au laboratoire THEMA de l'université Paris-X Nanterre[2] de 1994 à 2005, et poursuivie depuis au département d'économie de l'Ecole Polytechnique. Ces travaux portent sur les propriétés prédictives et incitatives (à la prévention des risques) des historiques individuels en assurance non-vie. Je remercie Georges Dionne de m'avoir suggéré la première piste de recherches, qui consistait à intégrer les coûts des sinistres dans la prédiction des risques. Mes recherches se sont donc inscrites d'abord dans une littérature actuarielle appréciant le pouvoir prédictif des historiques individuels en assurance non-vie. Elles sont présentées en section 2. Ma première contribution a pris en compte le coût des sinistres dans la prédiction des risques en fréquence et en prime pure ([12]). Ensuite, l'ancienneté de ces sinistres a été intégrée dans une prédiction des risques en fréquence ([8], [9]). Pour ce faire, ces articles ont remis en cause la constance dans le temps d'effets aléatoires intégrés à des lois de nombre d'accidents. La liaison entre les lois sur les effets aléatoires et le risque fréquence des contrats est étudiée dans [3].

De nombreuses règles monétaires et non monétaires utilisant les historiques des assurés (systèmes bonus-malus, amendes, permis à points) ont une justification en terme d'incitation à la prévention des risques. L'appréciation empirique de ces effets incitatifs est rendue difficile par une double interprétation des historiques individuels. Ceux-ci peuvent à la fois révéler des caractéristiques cachées des lois de risque (hétérogénéité inobservée), ou modifier ces lois. Des interactions avec Pierre-André Chiappori m'ont conduit à m'intéresser à ce sujet et à participer à des publications décrites plus en détail en section 3. Dans [5] et [6], des conditions d'élimination de l'hétérogénéité inobservée dans les historiques d'assurance sont explicitées, et appliquées aux contrats d'assurance automobile soumis au système bonus-malus

[2] L'université se nomme aujourd'hui Paris-Ouest Nanterre La Défense.

français. Le comportement optimal d'un assuré dans un environnement incitatif basé sur l'expérience individuelle est abordé dans les deux publications précitées, ainsi que dans [1] s'agissant des permis à points.[3] Cette dernière publication est le fruit d'une longue collaboration avec Pierre Picard, Georges Dionne et Jean-Marc Bourgeon sur le thème de la sécurité routière. On y analyse les effets incitatifs des permis à points, en fonction des règles d'amnistie des infractions. Ces amnisties s'effectuent, soit infraction par infraction une fois que celle-ci a atteint une certaine ancienneté, ou globalement après une durée donnée sans infraction commise. En présence d'aléa moral, nous montrons que les effets incitatifs contrecarrent les effets de révélation de l'hétérogénéité inobservée si les amnisties s'effectuent par infraction. Par contre, ces deux effets jouent dans le même sens dans le cas où les infractions sont amnistiées globalement. Enfin, [1] propose des résultats théoriques sur les modèles incitatifs, qui permettent des calculs de valeurs privées pour les sanctions non monétaires associées aux permis à points.

Les sept publications précitées figurent dans ce dossier d'HDR. Mes autres publications concernent la prise en compte de la stratification des portefeuilles dans l'évaluation des risques, avec une application aux flottes de véhicules ([10]), les modèles à plusieurs équations (systèmes multi-garanties, [11]), et les échelles bonus-malus ([7]). Le lien entre la fraude à l'assurance et la sélection des sinistres par les experts pour audit est analysé dans [4]. Des modèles intégrant le biais de sélection mettent à jour le risque de fraude pour les sinistres non évalués par les experts. Des travaux en cours sur les contrats de long terme et l'assurance dépendance ont donné lieu à une publication ([2]) traitant des comportements optimaux d'achat d'assurance dépendance. Ils sont évoqués dans la section 4, laquelle présente également mes autres projets de recherche.

[3] Cf. également une version antérieure ([26]).

2 Pouvoir prédictif des historiques individuels en assurance non-vie: l'approche actuarielle, ses méthodes et ses limites

2.1 "Experience rating" en assurance non-vie

L'usage des historiques individuels dans la tarification des contrats d'assurance non-vie est presque systématique, et trouve deux justifications.

- La première est en terme de neutralité actuarielle. Les études statistiques conduisent presque toujours au fait qu'une période sans sinistre est associée à une diminution du risque estimé en fréquence d'accidents, et que l'occurrence un sinistre entraîne une réévaluation de ce risque. On peut donc justifier les systèmes bonus-malus par un argument de neutralité actuarielle.

- La seconde se situe en terme d'incitation à la prévention des risques. En assurance non-vie, il existe une efficacité à court terme des efforts de prévention, et il n'y a donc pas de fatalité à être un mauvais risque. Le contexte est différent en ce qui concerne les risques viagers et de maladie. Le niveau de risque est associé à un capital (santé par exemple) dont la dépréciation est partiellement irréversible. Les efforts de prévention n'ont d'effet qu'à terme, et se pose un risque de reclassification. Il est rare d'observer des pratiques d'"experience rating" sur ces contrats.[4]

Le pouvoir prédictif des historiques individuels sur les risques est la résultante de deux interprétations possibles de ces historiques. D'une part, ils révèlent une

[4] Hendel et Lizzeri (2003) mentionnent cependant des contrats aux Etats-Unis appelés "Select and Ultimate" en assurance temporaire décès, où l'historique des examens de santé est utilisé dans la tarification.

hétérogénéité inobservée, exprimée de manière résiduelle par rapport à une information observable sur les contrats. Cette hétérogénéité résiduelle est représentée par des effets aléatoires dans le modèle statistique évaluant les risques (cf. [13] pour une revue de littérature). D'autre part, les historiques individuels modifient les niveaux de risque. Cette synthèse reviendra en détail sur les effets incitatifs du nombre des sinistres et du temps, mais on peut également citer des effets psychologiques comme cause de modification. Les historiques modifient la perception des risques et par suite les comportements de prévention. Tversky et Kahneman (1973) ont proposé une théorie dite "availability bias", selon laquelle l'estimation subjective de la fréquence d'un évènement dépend de la facilité avec laquelle on peut s'en rappeler l'occurrence. De ce fait, il y a réévaluation d'un risque après l'occurrence d'un évènement impliquant l'individu, et un effet négatif de l'ancienneté des évènements sur le niveau perçu du risque.

L'évaluation des risques à partir d'un principe de révélation de l'hétérogénéité inobservée revient à prédire un effet fixe individuel en utilisant l'estimation du modèle à effets aléatoires et un calcul d'espérance *a posteriori*. Les effets temporels et par évènement sont ceux qu'on observe sur les données réelles, à savoir une hausse des risques avec le nombre d'accidents et une baisse avec le temps. Ils sont donc de type "bonus-malus". Cette adéquation entre modèles et observations explique pourquoi les calculs de prédiction de la littérature actuarielle sont fondés uniquement sur un principe de révélation. Les approches paramétriques ont d'abord été privilégiées (cf. Lemaire (1995) pour une synthèse), puis les calculs semi-paramétriques de prédicteurs linéaires par rapport au nombre de sinistres (Bühlmann, 1967) sont devenus populaires. L'approche semi-paramétrique concerne les moments du second ordre du ou des effets aléatoires. C'est à ma connaissance la première application importante de méthodes développées et popularisées ensuite dans la communauté des économètres par Hansen (1982), offrant ainsi à Karl Pearson une revanche posthume

sur Ronald Fisher.[5] Les nombreuses extensions du modèle linéaire applicables aux données d'assurance non-vie (Nelder et Wedderburn (1972), Gouriéroux et Monfort (1984a, 1984b), Liang et Zeger (1986)) se situent entre les logiques paramétriques et semi-paramétriques.

Le coefficient bonus-malus actuariel pour le risque en fréquence est l'estimation de $(1 + n\sigma^2)/(1 + \lambda\sigma^2)$. Dans cette formule, n représente le nombre d'accidents observé dans le passé sur l'individu, et λ est le risque en fréquence (espérance mathématique du nombre d'accidents), calculé en fonction des variables observées dans le passé. Enfin, σ^2 est la variance de l'effet aléatoire qui rend compte du poids de l'hétérogénéité inobservée. Le risque évalué pour la période suivante en fonction des variables observables est multiplié par le coefficient bonus-malus, ce qui conduit à une prime intégrant l'historique individuel. Cette expression du coefficient bonus-malus est obtenue, soit par l'approche semi-paramétrique contraignant le coefficient à être une fonction affine de n, soit dans un cadre paramétrique où l'effet aléatoire suit une loi gamma (cf. Dionne, Vanasse (1989) pour un modèle de régression sur lois binomiales négatives, mélanges de lois de Poisson associées aux loi gamma).

Le coefficient bonus-malus peut également s'écrire

$$\frac{1 + n\,\widehat{\sigma^2}}{1 + \widehat{\lambda}\,\widehat{\sigma^2}} = \left(cred \times \frac{n}{\widehat{\lambda}}\right) + ((1 - cred) \times 1),\ cred = \frac{\widehat{\lambda}\,\widehat{\sigma^2}}{1 + \widehat{\lambda}\,\widehat{\sigma^2}}. \qquad (1)$$

Dans l'expression précédente, le coefficient *cred* est appelé crédibilité associée à l'historique individuel: il représente le bonus pour un contrat sans sinistre. En effet, le coefficient bonus-malus est égal à 1 en l'absence d'exposition au risque (on a alors $n = \widehat{\lambda} = 0$). Si $\widehat{\lambda} > 0$ et $n = 0$, le coefficient bonus-malus est égal à $1 - cred$, et la crédibilité est donc égale au bonus.

[5] En faisant un parallèle avec le vocabulaire de l'algèbre linéaire, un modèle paramétrique a une forme "image" (les lois sont indicées par les paramètres) alors qu'un modèle semi-paramétrique est de forme "noyau" (les paramètres indicent des contraintes sur les lois).

2.2 Résumé des travaux de l'auteur

Les quatre premières publications figurant dans ce document s'inspirent de cette approche.

La prise en compte de la gravité des sinistres par leur coût est détaillée dans [12]. Les modèles actuariels usuels prédisent le risque en fréquence, et il paraît naturel de vouloir étendre leur portée au risque en prime pure (espérance du coût total). Dans le cas du risque en fréquence, les lois de Poisson sont incontournables. Le processus du nombre cumulé des accidents a des trajectoires croissantes. Si les sauts sont de hauteur 1 (pas de sinistres simultanés) et si les accroissements du processus sont indépendants, le processus est nécessairement de type Poisson. Le mélange des lois créé par les effets aléatoires permet d'introduire une dynamique sur les données, qui se justifie si l'on observe une "surdispersion" sur les données. Cette surdispersion signifie que la variance du nombre des sinistres est supérieure à leur moyenne, alors qu'on a l'égalité entre ces deux moments pour les lois de Poisson. Dans le cas du coût des sinistres, la situation est moins claire car il n'y a pas de famille de lois de référence sur les réels positifs. On a coutume de les classer par queues de distribution croissantes en terme d'équivalents asymptotiques (d'abord les lois gamma, puis log-normales, puis de Pareto pour les familles les plus courantes). L'article [12] étudie les deux premières familles de lois, la deuxième se prêtant plus aisément à l'introduction d'une loi jointe sur les effets aléatoires relatifs aux lois de nombre et de coût de sinistres. Des calculs de pseudo-valeurs vraies (limites d'estimateurs en présence d'erreur de spécification) permettent d'obtenir des estimateurs explicites des paramètres du modèle. La famille des lois log-normales s'ajuste mieux aux données étudiées dans le papier que les lois gamma, le mélange des lois se justifie sur les lois de coût comme de nombre de sinistres, et la corrélation entre les effets aléatoires est très proche de zéro. Le modèle est appliqué ensuite à des calculs de prédicteurs de la prime pure des contrats, qui s'expriment comme le produit d'une estimation *a priori* du risque et d'un coefficient bonus-malus résumant l'historique individuel.

Compte tenu de la quasi indépendance entre les effets aléatoires associés aux lois de coût comme de nombre de sinistres, le coefficient bonus-malus sur la prime pure est à peu près égal au produit des coefficients associés aux risques en fréquence et en coût moyen.

La prise en compte de l'ancienneté des sinistres est développée dans [8] et [9].[6] Les études empiriques prouvent que le pouvoir prédictif des sinistres sur le risque décroît avec leur ancienneté, effet qui n'est pas pris en compte par les modèles à effets aléatoires constants. L'idée d'intégrer l'ancienneté des sinistres dans la prédiction avait déjà été développée par Sundt (1981), mais avec des contraintes sur le modèle qui rendaient les résultats peu crédibles. Dans [8] et [9], les corrélogrammes estimés sur des effets aléatoires dynamiques et stationnaires sont globalement décroissants. L'introduction d'une dynamique dans les effets aléatoires change en profondeur les propriétés des prédicteurs du risque en fréquence. L'effet "bonus" est plus faible pour les contrats sans sinistres qu'avec l'approche usuelle. Par contre, il est renforcé pour les contrats ayant eu des sinistres par le passé. A l'effet "bonus" usuel créé par la hausse de l'exposition au risque (hausse de $\widehat{\lambda}$ à n constant dans la formule (1)) s'ajoute l'augmentation de l'ancienneté des sinistres. L'effet "bonus" usuel est renforcé dans un contexte où le pouvoir prédictif des évènements décroît avec leur ancienneté. Une autre différence d'importance concerne le poids donné aux historiques individuels dans la prédiction, appelé crédibilité dans la littérature actuarielle et qui représente la baisse relative du risque d'un contrat non sinistré, comparé au risque *a priori* sur le contrat. Avec des effets aléatoires constants, cette crédibilité tend vers un (et la prime *a posteriori* tend vers zéro) quand la durée de l'historique tend vers l'infini. Ce résultat sur les primes est en contradiction avec les pratiques de

[6] L'analyse économique des systèmes bonus-malus en assurance automobile (Henriet, Rochet (1986)) conduit à distinguer des objectifs de sélection ou d'incitation à la prudence. Du point de vue de la sélection, seule importe la fréquence des accidents passés. Du point de vue des incitations à la prudence, la répartition des sinistres dans le temps doit intervenir dans le calcul de la prime.

la tarification en assurance automobile.[7] Avec des effets aléatoires dynamiques, ce résultat ne tient plus, et on peut avoir des crédibilités limites fortement inférieures à un.[8] On justifie ainsi un fait empirique important sur les poids des historiques dans la tarification.

L'article [3] remet en cause l'équidistribution des effets aléatoires dans un modèle de risque en fréquence. L'examen des données d'assurance non-vie fait apparaître un lien décroissant entre la variance de l'effet aléatoire et le risque en fréquence annuelle. En d'autres termes, l'hétérogénéité résiduelle augmente pour les risques faibles. Cet article estime des liens paramétriques et non paramétriques entre le risque en fréquence annuelle et la variance de l'effet aléatoire. Un lien décroissant est obtenu sur la base de données étudiée dans l'article, et on en déduit de nouvelles propriétés des coefficients bonus-malus. D'après la formule (1), la hausse relative du risque suite au premier sinistre est constante, et égale à σ^2. Ce résultat est de même nature que le malus appliqué en France, par exemple.[9] *A contrario*, le bonus actuariel croît avec l'exposition au risque. Les estimations faites en [3] conduisent à une élasticité entre variance de l'effet aléatoire et fréquence annuelle qui est proche de -1. A l'inverse du résultat précité, c'est le bonus-période qui devient presque indépendant du risque en fréquence annuelle, alors que le malus décroît avec ce risque.

[7] En France, le bonus maximum est de 50%. Il faut par ailleurs prendre en compte le déséquilibre de l'échelle bonus-malus, qui fait que le bonus moyen croît avec l'ancienneté de l'assuré. Les assurés avec 50% de bonus n'ont qu'un rabais de 20% par rapport au conducteur moyen. A l'inverse, les systèmes actuariels sont équilibrés par construction et un bonus s'exprime par rapport à la moyenne des contrats aussi bien que par rapport aux conducteurs débutants.

[8] C'est le maintien d'une contrainte d'égalité à un pour la crédibilité limite qui rendait les prédicteurs proposés par Sundt (1981) difficilement utilisables.

[9] Mais à la différence du système français, le malus de la formule actuarielle est une fonction linéaire du nombre de sinistres passés.

2.3 Limites de l'approche actuarielle

Les tarifications pratiquées en assurance non-vie sont loin de respecter la neutralité actuarielle, même si les écarts sont moindres qu'en assurance vie et prévoyance. Les subventions croisées entre les périodes d'un contrat sont de type "lowballing" ou "highballing" selon que les premières périodes sont subventionnées par les suivantes, ou le contraire. Une terminologie alternative est "back-loading" et "front-loading". En assurance non-vie, on observe plutôt un phénomène de "lowballing", qui peut être causé par l'extraction que les assureurs font de rentes d'information à partir de l'historique des sinistres. Kunreuther et Pauly (1985) obtiennent un contrat de cette nature dans un contexte d'absence d'engagement, où l'assuré ne prend ses décisions d'achat que sur la prime courante. Le "lowballing" peut aussi être obtenu si l'assureur intègre le fait que l'assuré ne fait pas toujours jouer la concurrence lors du renouvellement du contrat. Taylor (1986) développe ainsi des politiques de tarification optimales à partir d'un lien entre taux de résiliation et prix relatif par rapport au marché.

A l'opposé, les pratiques de "highballing" concernent surtout les contrats d'assurance vie et prévoyance. A la différence de l'assurance non-vie, les contrats sont de long terme, avec engagement unilatéral de l'assureur.[10] Les risques en vie et santé dépendent d'abord de l'âge, et les assurés les plus jeunes subventionnent les plus âgés en présence de "front-loading". Hendel et Lizzeri (2003) font une analyse empirique de contrats d'assurance "temporaire décès" avec un modèle d'apprentissage symétrique sur deux périodes, d'engagement unilatéral et d'hétérogénéité des acheteurs par rapport au coût du "front-loading". Le modèle prédit que les taux de résiliation en deuxième période diminueront avec le niveau de "front-loading", et que les risques moyens seront d'autant plus faibles que ce niveau sera élevé. Ce résultat est con-

[10] Seul l'assuré peut résilier le contrat jusqu'à un horizon donné. En assurance non-vie, l'assureur peut résilier le contrat à la date anniversaire, même si dans les faits les résiliations sont massivement à l'initiative des assurés (94% des résiliations en France pour l'assurance automobile).

firmé empiriquement par les primes observées aux USA sur trois types de contrats d'assurance "temporaire décès" (soit avec mise à jour annuelle des primes, ou avec des paliers par tranche d'âge et "front-loading", soit avec des primes indexées sur l'état de santé). Dans un contexte d'assurance non-vie, Dionne et Doherty (1994) présentent un modèle à deux périodes de type "highballing" avec antisélection, engagement unilatéral, et renégociation. Si l'assureur s'engage à une tarification en seconde période basée sur l'historique et robuste à la renégociation, les meilleurs risques choisiront ce contrat en première période plutôt qu'un contrat de court terme.

3 Identification et évaluation des effets incitatifs à la prévention créés par les historiques en assurance non-vie

3.1 Antisélection et aléa moral en assurance: une revue de littérature

Les publications figurant ensuite dans ce document sont relatives à l'identification des deux effets de révélation et de modification des lois de risque par les historiques individuels. Ce problème d'identification se pose à l'identique sur l'interprétation des choix de couverture d'assurance, et en nous faisons une revue de littérature. Le niveau de couverture d'un risque peut révéler une information non observable par l'assureur (antisélection), mais aussi modifier le niveau d'effort de prévention de ce risque (aléa moral). L'analyse d'un choix de couverture conduit donc à identifier l'antisélection et l'aléa moral. Les modèles économétriques traitant de l'asymétrie d'information en assurance sont exposés dans Chiappori (2000) (cf. également les synthèses de Dionne, Doherty, et Fombaron (2000) sur l'antisélection et de Chiappori, Salanié (2000a) sur l'économétrie des contrats d'assurance).

L'antisélection crée un lien croissant entre niveau de couverture et risque. Elle a de ce point de vue le même effet que l'aléa moral, comme le remarquent Chiappori et Salanié (2000). Les auteurs se placent dans un contexte d'assurance unique, comme dans Rothschild et Stiglitz (1976).[11] Chiappori et Salanié analysent l'influence du niveau de couverture sur le risque en fréquence d'accidents responsables, à partir d'un fichier de jeunes conducteurs, ceci afin de ne pas avoir à gérer l'utilisation de l'historique individuel dans les variables explicatives. Avec un modèle probit bivarié sur les indicatrices d'accident et de présence d'assurance complète, et intégrant l'information disponible sur les contrats, ils concluent à l'absence d'asymétrie d'information. Cette conclusion est ensuite renforcée par une analyse non paramétrique basée sur des statistiques locales dans l'échantillon. Dionne, Gouriéroux et Vanasse (2001) abordent le même sujet, mais avec le niveau de franchise dommages comme variable de choix. Ils commentent les formulations statistiques traduisant l'indépendance entre le risque et le choix de couverture, exprimée conditionnellement aux variables décrivant le contrat. Ils étudient l'influence des erreurs de spécification sur l'espérance de la variable de choix dans les tests d'indépendance conditionnelle. Ils rappellent la nature résiduelle de l'antisélection par rapport aux variables observables, et l'éliminent sur leurs données.

En assurance vie et santé, l'antisélection est bien présente également. Mais le cadre d'analyse est différent de celui de l'assurance non-vie, le contrat d'assurance n'étant pas exclusif du fait de la nature forfaitaire des contrats. Par exemple, l'antisélection est indubitable en assurance viagère quand celle-ci est achetée volontairement (ce n'est plus vrai si le capital converti en rente a été constitué par les cotisations à un fonds de pension). Se sentir en bonne santé est une condition nécessaire à l'achat de rente viagère, et cette information est d'autant plus discriminante que l'on est âgé. Par contre, Finkelstein et McGarry (2006) observent de la "sélection avantageuse" (i.e. un lien décroissant entre niveau de couverture et risque) sur les

[11] L'unicité du contrat suppose que la garantie est indemnitaire et non forfaitaire.

garanties "temporaire décès" ou dépendance. Ce terme de "sélection avantageuse" a été proposé par De Meza et Webb (2001) qui la justifient dans le cas où les individus avec le moins d'aversion au risque - et donc les moins enclins à acheter de l'assurance - ont peu d'activité de prévention du risque assuré. Ces hypothèses peuvent créer une liaison décroissante entre achat d'assurance et niveau de risque, dans la mesure où l'on ne conditionne pas par l'aversion au risque dans le calcul. Un tel résultat permet d'avoir un effet de révélation par le choix de couverture à l'opposé de l'aléa moral, ce qui donne une explication possible sur l'absence d'asymétrie d'information relevée empiriquement par Chiappori et Salanié (2000), et par Dionne, Gouriéroux et Vanasse (2001). Fang, Keane et Silverman (2008) remarquent que toute information privée positivement corrélée avec le niveau de couverture et négativement corrélée avec le risque permet de créer de la "sélection avantageuse", comme De Meza et Webb le font avec l'aversion au risque. Ils donnent l'exemple du lien entre dépenses de santé des personnes âgées aux Etats-Unis, et possession d'une couverture complémentaire à "medicare" (laquelle couverture est appelée "medigap"). En conditionnant par les variables déterminant les prix, les détenteurs de cette couverture complémentaire dépensent 4000 dollars de moins par an que ceux qui ne l'ont pas. On obtient ainsi un résultat de type "sélection avantageuse", mais celui-ci est inversé quand on rajoute l'information sur l'état de santé. A niveau de santé donné, la couverture génère 2000 dollars par an de dépenses supplémentaires. Les détenteurs de "medigap" ont un meilleur statut social et donc une meilleure santé en moyenne que les autres personnes âgées, ce qui explique le résultat. En assurance non-vie, Huang, Wang et Tzeng (2009) montrent qu'un résultat de "sélection avantageuse" sur l'assurance incendie des entreprises à Taiwan est renversé si l'on intègre les activités de prévention du risque (achats d'équipements anti-incendie ou formation des salariés sur les risques).

3.2 Révélation et modification des lois de risque par les historiques individuels: identification et mesure

La littérature statistique a longtemps peiné à comprendre les problèmes d'identification créés par l'interprétation des historiques individuels. En commentant un papier écrit par Neyman (1939), Feller (1943) mentionne les deux interprétations des lois binomiales négatives en terme d'hétérogénéité inobservée, et en terme de modification des lois par l'historique. Il remarque que cette double interprétation échappe à la plupart des auteurs prenant l'un ou l'autre point de vue, y compris à Neyman.[12] Feller conclut à l'impossibilité d'identifier la nature de la dynamique sur les données. A la fin de son article, Feller mentionne toutefois l'idée qu'une analyse des durées entre les évènements plutôt que de leur nombre permettrait peut-être de progresser dans la voie d'une identification.

Le contenu de cet article n'échappa pas à Neyman, qui cosigna une dizaine d'années plus tard un article (Bates, Neyman (1952)) proposant une manière d'éliminer la loi de mélange associée à un processus de Poisson. En considérant une population d'individus pour lesquels le nombre d'accidents suit un processus de Poisson dont le paramètre varie selon les individus, les auteurs construisent un test basé sur le résultat suivant. Si l'on considère les individus ayant eu un accident et un seul sur un intervalle, la date de cet accident suit une loi uniforme sur l'intervalle. Cette loi étant indépendante du paramètre individuel, les auteurs proposent un test d'adéquation des dates d'accidents à une loi uniforme, de type Kolmogorov-Smirnov. L'hypothèse nulle est associée par les auteurs au fait que l'historique des accidents ne modifie pas les lois individuelles, comme c'est le cas pour les processus de Poisson. Plusieurs papiers de la littérature économétrique (Heckman, Borjas (1980) et [5], [6]) se sont inspirés de cette contribution.

[12] Neyman était pourtant loin d'être un débutant, ayant déjà publié ses travaux sur les tests optimaux en 1939.

Mais la conclusion de Neyman et Bates est exagérément optimiste. En effet, un mélange de lois de Poisson peut s'appliquer à des individus réels, et non pas à des classes d'individus associés à des variables observables données. Dans le premier cas, l'historique modifierait les lois individuelles bien que l'hypothèse nulle soit vérifiée. Par exemple, un processus de Poisson de paramètre λ mélangé suivant une loi $\gamma(a, a)$ est associé à une fonction de hasard dont la valeur en t et pour n sinistres survenus entre 0 et t est égale à

$$\lambda_n(t) = \lambda \frac{a+n}{a+\lambda t}.$$

C'est un processus de Pólya, dont les lois marginales sont des binomiales négatives. Appliquer ce processus à des individus soumis à des incitations supposerait un effort décroissant après chaque accident et croissant avec le temps. Mais les environnements incitatifs convexes comme celui créé par le système bonus-malus français induisent des effets opposés. La fonction de hasard se comporte alors à l'opposé de celle du processus de Pólya. Cela dit, une fonction de hasard de cette nature peut également conduire à une loi uniforme pour la date d'un sinistre supposé unique. Considérons en effet un intervalle $[0, T]$, et N_t le nombre de sinistres entre 0 et t. Si $\lambda_n(t)$ est la densité de transition en t entre n et $n+1$ sinistres (ou fonction de hasard), et si $\Lambda_n(t) = \int_0^t \lambda_n(u)du$ est la fonction de hasard intégrée, on a

$$P\left[N_T = 1\right] = \int_0^T \exp(-\Lambda_0(t))\lambda_0(t)\exp(\Lambda_1(t) - \Lambda_1(T))dt,$$

où t est la date de l'accident unique. La densité de la loi de cette date est égale à $\lambda_0(t) \times \exp(\Lambda_1(t) - \Lambda_0(t))$, à une constante multiplicative près. La dérivée logarithmique de la densité est égale à $\left(\lambda_0^{'}/\lambda_0\right) + \lambda_1 - \lambda_0$. L'hypothèse nulle (loi uniforme pour la date du sinistre) correspond à un équilibre entre les composantes relatives au temps et aux évènements de cette dérivée, soit

$$\frac{\lambda_0^{'}}{\lambda_0} \text{ (temps)}; \;\; \lambda_1 - \lambda_0 \text{ (évènement)}. \tag{2}$$

Dans le cas du processus de Pólya, on a $\lambda_0^{'}/\lambda_0 < 0$, et $\lambda_1 - \lambda_0 > 0$. Mais on peut avoir des signes opposés pour les deux composantes si les modifications du risque par l'historique individuel sont dues à des incitations financières dans un environnement convexe. La loi uniforme de la date d'un sinistre supposé unique traduit alors un équilibre entre l'effet désincitatif du temps, et l'effet incitatif créé par l'accident.

L'élimination de l'hétérogénéité inobservée est développée dans [5] et [6] dans l'esprit de la contribution Bates-Neyman, à partir d'une analyse des données de durée séparant les évènements. Une application est faite sur des données françaises d'assurance automobile, et soumises aux règles de bonus-malus. Dans cet environnement convexe, la pénalité suite à un accident croît avec la prime de base (elle est ici proportionnelle à cette prime), et l'efficacité de l'effort augmente quand on s'élève dans l'échelle bonus-malus. On s'attend donc à voir un individu augmenter son effort de conduite prudente après un accident. La fonction de hasard d'un risque en fréquence d'accidents est modélisée sous la forme

$$h(t) = \lambda \times \beta^{N_{t^-}} \times \psi(t) \ (t \geq 0).$$

Le paramètre λ est individuel. Il suit une loi G sur la population qui reflète une hétérogénéité explicable partiellement par des composantes de régression. Le paramètre β traduit la modification du risque induite par l'historique individuel (N_{t^-} est le nombre d'accidents de l'individu dans l'intervalle $[0, t[$, et le risque est donc multiplié par β après chaque accident). Le système bonus-malus français est multiplicatif (25% de hausse de prime après un accident responsable), et le modèle précédent retient la même logique au niveau de l'efficacité de l'effort optimal. Dans [6], il est prouvé que l'effort croît après chaque accident dans un contexte de tarification suivant la logique du système bonus-malus français. La fonction ψ représente l'effet du temps calendaire sur le risque, qui peut traduire entre autres des effets incitatifs à temps continu. Les principaux résultats produits dans [6] sont les suivants.

- Si $\beta = 1$ (les accidents ne créent pas d'effets incitatifs si l'on s'en tient à cette explication des effets de modification des lois par l'historique), la fonction ψ est identifiable, à partir de l'égalité $P[N_t = 1 \mid N_T = 1] = \Psi(t)/\Psi(T)$ $(0 \leq t \leq T,\ \Psi$ primitive de ψ s'annulant en 0). La loi G sur le paramètre individuel est également identifiable par sa transformée de Laplace, à partir de la fonction de répartition empirique associée à la date du premier accident.

- Si ψ est constant, β est identifiable si la durée d'observation est infinie. Si T_n est la date du $n^{ième}$ accident, on a

$$P[T_1 > T_2 - T_1] = \frac{\beta}{1+\beta}. \tag{3}$$

  On a ainsi un estimateur convergent de β à partir de la fréquence empirique de l'événement $[T_1 > T_2 - T_1]$. Mais si la durée d'observation est finie, le biais de sélection (seuls les contrats avec deux accidents et plus peuvent être retenus dans l'analyse) ne permet plus en général d'identifier β. Par contre un test d'hypothèse nulle $\beta = 1$ est possible dans tous les cas. Ce test prolonge l'approche Bates-Neyman en considérant sur un intervalle de temps la fonction de répartition empirique $\widehat{F_2}$ de la date du deuxième accident pour les contrats en ayant eu deux, et en la comparant au carré de la fonction $\widehat{F_1}$ associée aux contrats avec un accident. A l'hypothèse nulle, $\widehat{F_2} - \widehat{F_1}^2$ a une espérance nulle en toute date et pour toute valeur de la fonction calendaire ψ. Sur les données, l'hypothèse nulle n'est pas rejetée.

Le modèle propose ainsi des avancées notables par rapport à la littérature existante, mais une identification simultanée des deux composantes de la dynamique sur les données (ici, β et ψ) reste hors de portée. On peut compléter le modèle proposé en [6] de la manière suivante. L'équation (3) peut également être obtenue avec une formulation du type

$$h(t) = \lambda \times \beta^{N_{t^-}} \times \psi(t - T_{N_{t^-}})\ (T_0 = 0). \tag{4}$$

Dans cette formulation, c'est l'ancienneté du dernier sinistre (s'il existe) qui détermine l'effet temporel, et non plus le temps calendaire. Il est aisé de montrer que le résultat donné en (3) reste valable pour toute valeur de λ et ψ, ce qui étend la possibilité d'estimer β en éliminant non seulement l'hétérogénéité individuelle mais aussi la fonction du temps ψ et donc les effets incitatifs que celle-ci pourrait représenter. On utilise pour cela la propriété de "hasards proportionnels" qu'a l'équation (4) par rapport au nombre de sinistres passés. La limite d'utilisation du modèle est la même que dans [6]: la durée d'observation doit être infinie pour ne pas avoir un biais de sélection.

Pour conclure, analysons le système bonus-malus français qui sous-tend le travail empirique. La fréquence moyenne des accidents fait que l'effet bonus l'emporte en moyenne sur le malus (les conducteurs ont en moyenne 40% de bonus, alors qu'un débutant n'en a pas). L'incitation à une conduite prudente diminuant avec le coefficient, les effets désincitatifs s'exerçant à temps continu l'emportent en moyenne sur les effets incitatifs créés à chaque accident, ce qui s'interprète dans l'équation (2). Il faut signaler que la fonction des systèmes bonus-malus a changé en France comme ailleurs en Europe. Le système bonus-malus n'a plus de caractère contraignant sur la tarification depuis 1994,[13] comme c'était le cas à l'époque de l'observation des données (qui ont été extraites en 1988). En application d'une directive européenne, la tarification doit être non contrainte et le coefficient bonus-malus s'applique donc sur une prime qui peut intégrer l'expérience individuelle. La Commission européenne a voulu supprimer ensuite les systèmes bonus-malus, mais n'y est pas parvenue. Les défenseurs de ces systèmes ont mis en avant le fait que le maintien d'un coefficient bonus-malus sur le contrat créait une information commune à tous les acteurs du marché et diminuait les rentes d'information détenues par les assureurs (cf. [28] pour une quantification).[14]

[13] La liberté tarifaire a été instaurée par la troisième directive européenne en assurance non-vie.

[14] Le conflit entre la Commission européenne et la France à ce sujet a commencé en 2002 et s'est

3.3 Efficacité des mécanismes monétaires et non monétaires incitant à une conduite prudente

Le coût financier et social des accidents de la route est en constante augmentation au niveau mondial. Une étude menée par la Harvard School of Public Health prévoit qu'à l'horizon 2020, les accidents de la route seront la troisième cause d'années en bonne santé perdues, après les maladies cardiaques et les cancers (Murray et Lopez (1997)). Les résultats par pays montrent une nette divergence entre les pays de l'OCDE pour lesquels les résultats sont en amélioration régulière depuis plus de 30 ans, et les autres qui connaissent pour la plupart une dégradation rapide des indicateurs associés à la violence routière. Les pays les plus riches ont sans doute de meilleures infrastructures routières que les autres[15], mais la baisse observée depuis les années 70 est concomitante à la mise en place de systèmes de contrôle et de sanction des infractions aux règles de conduite. Les mécanismes décrits ci-après ont modifié en profondeur les comportements au volant (cf. Boyer et Dionne (1987)).

Les amendes et les permis à points sont des mécanismes incitant à une conduite prudente et basés sur les infractions. Dans le cas du permis à points, la sanction est non monétaire, le permis de conduire étant annulé si le total des points dépasse un certain seuil. Les primes d'assurance peuvent aussi dépendre de l'historique des infractions, comme c'est le cas au Québec pour l'assurance des dommages corporels. L'évolution du système incitatif au Québec et des résultats obtenus en termes d'accidents et d'infractions est analysée dans [1]. L'article analyse les propriétés incitatives des permis à points, en fonction de leurs caractéristiques. Ces propriétés dépendent d'abord de la manière dont les points associés aux infractions sont amnistiés. L'amnistie peut être soit totale après une durée donnée de conduite sans

conclu en 2004 par une décision de justice favorable à la France.

[15] L'amélioration de la sécurité des véhicules et des routes est une voie de progrès en sécurité routière, mais elle peut aussi induire un accroissement de la prise de risque par les conducteurs (effet d' "homéostasie"). Ce point est discuté par Gossner et Picard (2005).

infraction (c'est le cas en France pour les infractions à plus de un point, et en Espagne), ou être attachée à l'infraction après une durée donnée comme au Québec et aux Etats-Unis. Les propriétés incitatives des systèmes avec amnistie totale sont étudiées dans [26]. Le niveau incitatif est déterminé par la variation d'une utilité de continuation entre la situation présente et celle qui suit l'infraction (on suppose toutes les infractions à un point pour simplifier). La pénalité associée au retrait de permis est assimilée à la privation d'une utilité de conduite durant une durée aléatoire, à la suite de laquelle le compteur des points est remis à zéro. Les modèles présentés dans le papier sont une extension de ceux utilisés par Bourgeon et Picard (2007), un niveau d'effort continu remplaçant un effort binaire. Dans tous les systèmes à points, l'utilité augmente avec le temps en présence d'amnisties. Dans le cas d'une amnistie totale, le compteur temps est remis à zéro après une infraction, et l'utilité suivant l'infraction est donc constante pour un nombre donné d'infractions non amnistiées. Le niveau incitatif est alors croissant avec le temps, car il est défini comme la différence entre utilité courante et utilité suite à infraction. On en déduit que le risque d'infraction diminue avec le temps. Par contre, le saut du niveau incitatif suite à une infraction n'est pas de signe constant. L'incitation augmente suite à une infraction si cette dernière est assez proche de la précédente. Si l'horizon de l'amnistie totale est assez proche de l'infraction, le niveau incitatif décroît au moment de l'infraction. Pour résumer, les effets incitatifs vont dans le même sens que ceux créés par la révélation de l'hétérogénéité inobservée dans la dimension temporelle, au contraire de ceux créés par un système bonus-malus. Il existe une clause dans le système bonus-malus français, dite "de retour rapide" vers la situation initiale (ni bonus, ni malus) pour les conducteurs avec malus qui ont deux années sans sinistre responsable. Les propriétés incitatives de cette clause sont similaires à celles du type de permis à points que nous venons d'analyser.

Les propriétés incitatives d'un système où chaque infraction est amnistiée après une durée donnée (deux ans au Québec) sont plus complexes car toutes les anci-

ennetés des infractions non amnistiées sont des variables déterminant l'utilité de continuation et le niveau d'effort optimal. On peut faire l'analyse qualitative suivante. Si les amnisties se font sur chaque infraction, l'utilité atteinte après infraction croît avec le temps de même que l'utilité d'avant l'infraction. L'effet temporel d'amélioration de l'utilité de continuation devrait être plus élevé dans une situation plus médiocre. Si l'utilité croît plus vite après l'infraction, le niveau incitatif décroît avec le temps car il s'exprime comme différence des utilités avant et après l'infraction. On peut prouver que ce niveau incitatif est continu au moment de l'amnistie d'une infraction. Le niveau incitatif doit croître après une infraction pour compenser l'effet négatif du temps. C'est par ailleurs le résultat qu'on obtient pour un permis à points sans amnistie. Les permis à points avec amnistie par infraction ont des propriétés incitatives proches du système bonus-malus français, ou d'une tarification convexe.[16] L'analyse empirique de [1] porte dans un premier temps sur une période où les incitations à une conduite prudente sont le permis à points et un système d'amendes associées aux infractions. D'après la théorie, l'effort doit être croissant avec le nombre de points non amnistiés. Son effet doit alors contrecarrer celui de la révélation de l'hétérogénéité inobservée. Un renversement de l'effet de révélation (croissance du risque à chaque infraction) prouve l'existence d'effets incitatifs, et c'est effectivement ce qu'on observe. Alors que le permis est annulé au-delà de douze points, le risque d'infraction croît jusqu'à sept points puis décroît ensuite. On peut faire un parallèle avec les mesures incitatives qui ont pour objet de faire diminuer la délinquance. Une exemple un peu extrême a été étudié par les économistes, qui est la règle du "three strikes and you're out" établie en Californie pour dissuader la récidive des délits graves ou des crimes. Ceux-ci remplacent les infractions dans l'analyse précédente, et le retrait de permis est remplacé par rien moins qu'une peine de prison allant de vingt cinq ans à la perpétuité. Helland

[16] Les effets sur le risque sont également de même nature que les biais psychologiques à la Tversky-Kahneman.

et Tabarrok (2007) montrent que l'effet dissuasif augmente entre la première et la seconde condamnation.

La baisse de la fréquence d'infractions n'a d'intérêt en terme de sécurité routière que si elle est suivie d'une baisse de la fréquence d'accidents. Sur l'ensemble de la période d'étude (de 1983 à 1996) la fréquence d'accidents baisse régulièrement, alors que celle des infractions est globalement stationnaire. L'explication la plus probable est que la hausse des contrôles radar a augmenté le taux d'enregistrement des infractions. Ainsi la stationnarité de la fréquence des infractions enregistrées dans le fichier indique une baisse de la fréquence des infractions commises, ce qui peut expliquer le résultat obtenu sur les accidents.

L'introduction en 1992 au Québec d'une prime d'assurance indexée sur les points a créé un troisième niveau d'incitation à la prudence au volant, complétant les amendes et le permis à points. Une baisse de 15% de la fréquence des infractions a été constatée suite à cette réforme. Nous avons relié l'efficacité de cette réforme aux variations qu'elle induit dans les niveaux incitatifs.[17] Cette baisse de 15% est à rapporter à une hausse moyenne de 9 à 10% de l'effet incitatif suite à la réforme. L'élasticité entre efficacité et niveau des incitations est associée dans l'article à la forme de la fonction λ reliant l'efficacité de l'effort à sa désutilité. Si l'effort est strictement positif pour un conducteur supposé représentatif, nous montrons que l'élasticité inférieure à -1 que nous observons est associée à une concavité de $\log(\lambda)$. Enfin l'efficacité incitative estimée des infractions, jointe aux résultats précédents, nous permet de calculer des équivalents monétaires pour les infractions et les retraits de permis.

[17] Les résultats sont sujets à caution dans la mesure où il n'y a pas de population de contrôle (l'assureur étant en situation de monopole) permettant d'éliminer les effets calendaires avec des "difference in differences".

4 Recherches en cours

4.1 Contrats de long terme et assurance dépendance

Ce thème de recherches s'inscrit dans la chaire "Assurance et risques majeurs" financée par la compagnie d'assurances Axa et portée par l'Ecole Polytechnique, Paris IX-Dauphine et l'ENSAE. Il a pu être développé grâce à une collaboration avec Montserrat Guillén, de l'université de Barcelone. Nous avons eu accès à une base de données espagnole de contrats de long terme, couvrant des risques de décès, de maladie et de dépendance. Comme on l'a rappelé en section 1, les contrats de long terme se caractérisent par un engagement unilatéral de l'assureur à prolonger la couverture jusqu'à un horizon qui est ici le décès de l'assuré. Par ailleurs, la prime ne peut pas dépendre de l'historique individuel des épisodes de maladie. Ces caractéristiques du contrat permettent l'assurance contre le risque de reclassification, justifié par l'existence d'un capital santé dont la dégradation est en grande part irréversible. Dans une publication récente ([2]), nous avons analysé le comportement d'achat en l'expliquant par un modèle de cycle de vie à la Yaari (1965). Mais il est en fait difficile de rendre rationnel l'achat à 30 ans en moyenne de trois couvertures jointes (décès, maladie et dépendance) alors que la dernière d'entre elles est souscrite d'habitude par des personnes nettement plus âgées. Dans un travail en cours ([25]), nous analysons la structure de tarification et les comportements de résiliation des assurés. Il n'y a pas d'engagement de l'assureur sur la tarification future dans ce type de contrats, ce qui peut s'expliquer par l'incertitude sur les lois de risque alors que les engagements sont non révisables.[18] Mais cette absence d'engagement, si elle permet d'éviter tout risque d'insolvabilité de l'assureur, peut conduire celui-ci à abuser de sa situation de monopole. Le risque est déterminé en premier lieu par l'âge, et les jeunes assurés subventionnent fortement les assurés plus âgés. Dans ce

[18] Cette incertitude est particulièrement élevée en dépendance, où les lois d'entrée et de durée en dépendance ont connu des effets de génération importants.

contexte de "highballing" analysé en section 2.3, un assuré qui résilie son contrat perd l'excédent entre les primes versées et le risque couvert, car il n'y a pas de valeur de rachat sur les garanties maladie et dépendance.[19] Les contrats de long terme fonctionnent donc essentiellement en répartition, et nous constatons sur les données une grande stabilité du rapport entre prestations et primes. Cette logique de répartition rapproche ces contrats des systèmes de sécurité sociale, et d'ailleurs le produit analysé dans nos travaux était vu à sa création comme un complément à la sécurité sociale espagnole. Mais ce contrat privé est très différent de l'assurance sociale du fait de la liberté laissée à l'entrée en portefeuille (pour l'assuré comme pour l'assureur) et à la sortie pour l'assuré. Si les assurés ont choisi librement cette couverture, l'assureur n'est de son côté pas tenu de laisser le portefeuille ouvert à l'entrée. Ce portefeuille a d'ailleurs été fermé à l'entrée en 1997 et la tarification a ensuite augmenté rapidement, suivant ainsi le vieillissement des assurés du portefeuille. Une telle situation devrait entraîner une "spirale de la mort", c'est-à-dire une extinction progressive du portefeuille créée par le départ continu des plus jeunes, qui n'ont plus la perspective d'être subventionnés ultérieurement par les nouveaux entrants. On observe une hausse modérée des résiliations suite à la clôture du portefeuille, mais celles-ci se stabilisent ensuite. On peut imaginer que les assurés n'ont pas réellement conscience de la situation du portefeuille, et la lisibilité est un problème majeur des contrats de long terme. On en a une confirmation avec un pic de résiliation à 65 ans: les assurés qui résilient pensent sans doute que la couverture maladie est de type incapacité de travail, alors que la garantie est la même, que l'assuré soit actif ou retraité. L'amplitude des subventions croisées entre classes d'âge donne un pouvoir redistributif considérable aux résiliations des contrats de long terme (cf. Brown et Finkelstein (2007) s'agissant de l'assurance dépendance). Dans nos travaux en cours, nous essayons de relier le niveau des résiliations à ses

[19] Seule la garantie décès inclut une valeur de rachat quand elle est de type "vie entière". Dans le portefeuille étudié, seule une composante de la garantie décès était de ce type.

différentes causes. La modification du risque réel ou perçu peut être une cause de résiliation, et nous observons en effet un taux de résiliation accru pour les assurés avec un bon historique sur le risque maladie. Les résiliations peuvent être subies dans le cas de contraintes de liquidité, mais cette cause devrait être de peu de poids dans notre étude car les primes (comme les garanties) sont faibles. Enfin, la cause majeure de résiliation est une baisse de motivation d'assurance ou une modification de la perception du produit. Les personnes âgées peuvent par exemple ne plus vouloir être assurées contre le risque de dépendance pour inciter leurs enfants à substituer l'aide familiale à des services du secteur marchand à la personne âgée. Surtout, le taux de résiliation élevé des jeunes assurés (8% par an) s'explique d'abord par une compréhension progressive du produit qui les pousse à en changer (pourquoi par exemple avoir acheté de l'assurance dépendance aussi jeune?).

Nos recherches vont s'orienter ensuite vers le partage public-privé, sujet d'actualité pour l'assurance dépendance.

4.2 Demande d'assurance et prévention des risques majeurs: l'exemple du National Flood Insurance Program aux Etats-Unis

Ces travaux sont menés avec Pierre Picard, Erwann Michel-Kerjan et Sabine Lemoyne. Ils ont commencé il y a peu et vont utiliser la base des contrats d'assurance individuels couverts par le National Flood Insurance Program (NFIP) aux Etats-Unis. Cette base contient plusieurs millions de contrats suivis sur la décennie 2000. Elle contient les caractéristiques individuelles des contrats, et des variables résumant l'activité de prévention des risques au niveau de "communautés". Celles-ci sont de petites entités géographiques notées régulièrement par le NFIP sur la base de leurs activités de prévention. La note détermine alors un rabais sur une prime d'assurance sensée être calculée de manière actuarielle. Nous avons privilégié l'analyse de la demande

d'assurance comme première piste de recherches. Des articles dans cette veine ont déjà été proposés par Browne et Hoyt (2000), et par Zahran et al. (2009). Nous comptons étendre la contribution des auteurs en analysant la manière dont la demande est influencée par les historiques dans les dimensions spatiales et temporelles, et en écrivant des modèles de choix basés sur la perception des risques.

References

[1] Bates, G.E., et J. Neyman (1952) "Contributions to the theory of accident proneness II: True or false contagion," *University of California Publications in Statistics* **1,** 255-275.

[2] Bourgeon, J.M. et P. Picard (2007) "Point-record driving license and road safety: An economic approach," *Journal of Public Economics* **91**, 235-258.

[3] Boyer, M. et G. Dionne (1987) "The economics of road safety," *Transportation Research* **21**, 413-431.

[4] Brown, J., et A. Finkelstein (2007) "Why is the market for long-term care insurance so small?" *Journal of Public Economics* **91**, 1967-1991.

[5] Browne, M.J. et R.E. Hoyt (2000) "The demand for flood insurance: empirical evidence," *Journal of Risk and Uncertainty* **20**, 291-306.

[6] Bühlmann, H. (1967) "Experience Rating and Credibility," *ASTIN Bulletin* **4**, 199-207.

[7] Chiappori, P.A. (2000) "Econometric models of insurance under asymmetric information," *Handbook of Insurance* 365-393, Kluwer Academic Publishers. Huebner International Series on Risk, Insurance and Economic Security (Editeur: Georges Dionne).

[8] Chiappori, P.A., et B. Salanié (2000) "Testing for asymmetric information in insurance markets," *Journal of Political Economy* **108,** 56-78.

[9] Chiappori, P.A., et B. Salanié (2000a) "Testing contract theory: A survey of some recent work," congrès mondial de la société d'économétrie.

[10] De Mezza, D. et D.C. Webb (2001) "Advantageous selection in insurance markets," *RAND Journal of Economics* **32**, 249-262.

[11] Dionne, G., N. Doherty (1994) "Adverse selection, commitment, and renegotiation: Extension to and evidence from insurance markets," *Journal of Political Economy* **102**, 209-235.

[12] Dionne, G., N. Doherty et Fombaron, N. (2000) "Adverse selection in insurance markets," *Handbook of Insurance* 185-244, Kluwer Academic Publishers. Huebner International Series on Risk, Insurance and Economic Security (Editeur: Georges Dionne).

[13] Dionne, G., Gouriéroux, C. et Vanasse, C. (2001) "Testing for adverse selection in the automobile insurance market: A comment," *Journal of Political Economy* **109,** 444-453.

[14] Dionne, G., et C. Vanasse. (1989) "A generalization of automobile insurance rating models: the negative binomial distribution with a regression component," *ASTIN Bulletin* **19**, 199-212.

[15] Fang, H., Keane, M.P. et Silverman, D. (2008) "Sources of advantageous selection: Evidence from the medigap insurance market," *Journal of Political Economy* **116,** 303-350.

[16] Feller, W. (1943) "On a general class of "contagious" distributions," *The Annals of Mathematical Statistics* **14**, 389-400.

[17] Finkelstein, A. et McGarry, K. (2006) "Multiple dimensions of private information: Evidence from the long-term care insurance market," *American Economic Review* **96**, 938-958.

[18] Gossner, O. et P. Picard (2005) "On the consequences of behavioral adaptations in the cost-benefit analysis of road safety measures," *Journal of Risk and Insurance* **72**, 577-599.

[19] Gouriéroux, C., A. Monfort, et A. Trognon (1984a) "Pseudo Likelihood Methods: Theory," *Econometrica* 52, 681-700.

[20] Gouriéroux, C., A. Monfort, et A. Trognon (1984b) "Pseudo Likelihood Methods: Applications to Poisson Models," *Econometrica* 52, 701-720.

[21] Hansen, L.P. (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* **50**, 1029-1054.

[22] Heckman, J.J., et G.J. Borjas. (1980). "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence," *Economica* 47, 247-283.

[23] Helland, E. et A. Tabarrok (2007), "Does three strikes deter? A nonparametric estimation," *The Journal of Human Resources* **52**, 2309-330.

[24] Hendel, I., et A. Lizzeri (2003) "The role of commitment in dynamic contracts: evidence from life insurance", *The Quarterly Journal of Economics* **118**, 299-327.

[25] Henriet, D., et J.C. Rochet (1986) "La logique des systèmes bonus-malus en assurance automobile", *Annales d'économie et de statistiques* **1** 133-152.

[26] Huang, R.J., K.C. Wang, et L.Y. Tzeng (2009) "Empirical evidence for advantageous selection in the commercial fire insurance market," *The Geneva Risk and Insurance Review* **34**, 1-19.

[27] Kunreuther, H., et M. Pauly (1985) "Market equilibrium with private knowledge: An insurance example," *Journal of Public Economics* **26**, 269-288**.**

[28] Lemaire, J. (1995) *Bonus-Malus Systems in Automobile Insurance.* Huebner International Series on Risk, Insurance and Economic Security.

[29] Liang, K.Y., et S.L. Zeger (1986) "Longitudinal data analysis using generalized linear models," *Biometrika* **73**, 13-22.

[30] Murray, Christopher J.L., et Alan D. Lopez (1997) "Alternative Projections of Mortality and Disability by Cause 1990-2020: Global Burden of Disease Study," *The Lancet* **349**, 1498-1504.

[31] Nelder, J.A., et R.W.M. Wedderburn (1972) "Generalized linear models," *Journal of the Royal Statistical Society A* **135**, 370-384.

[32] Neyman, J. (1939) "On a new class of "contagious" distributions, applicable in entomology and bacteriology," *The Annals of Mathematical Statistics* **10***,* 35-57.

[33] Rothschild, M., et J. Stiglitz (1976) "Equilibrium in competitive insurance markets: an essay on the economics of imperfect information," *The Quarterly Journal of Economics* **91**, 629-649.

[34] Sundt, B. (1981) "Credibility estimators with geometric weights," *Insurance: Mathematics and Economics* **7**, 113-122.

[35] Taylor, G. (1986) "Underwriting strategy in a competitive insurance environment," *Insurance: Mathematics and Economics* **5***,* 59-77.

[36] Tversky, A. et D. Kahneman (1973) "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology* **5**, 207-232.

[37] Yaari, M. (1965) "Uncertain lifetime, life insurance, and the theory of the consumer," *The Review of Economics Studies* **32,** 137-150.

[38] Zahran, S., Weiler, S., Brody, S.D., Lindell, M.K. et W.E. Highfield (2009) "Modeling national flood insurance policy holding at the county scale in Florida," *Ecological Economics* **68**, 2627-2636.

Pouvoir prédictif des historiques individuels en assurance non-vie

- Prise en compte du coût des sinistres dans les systèmes bonus-malus

  J. Pinquet (1997) "Allowance for cost of claims in bonus-malus systems", *ASTIN Bulletin* **27**, 33-57.

- Prise en compte de l'ancienneté des sinistres dans les systèmes bonus-malus (I)

  C. Bolancé, M. Guillén et J. Pinquet (2001) "Allowance for the age of claims in bonus-malus systems", *ASTIN Bulletin* **31**, 337-348.

- Prise en compte de l'ancienneté des sinistres dans les systèmes bonus-malus (II)

  C. Bolancé, M. Guillén et J. Pinquet (2003) "Time-varying credibility for frequency risk models: Estimation and tests for autoregressive specifications on the random effects", *Insurance: Mathematics and Economics* **33**, 273-282.

- Remise en cause de l'équidistribution des effets aléatoires dans les modèles de risque en fréquence

  C. Bolancé, M. Guillén et J. Pinquet (2008) "On the link between credibility and frequency premium", *Insurance: Mathematics and Economics* **43**, 209-213.

ALLOWANCE FOR COST OF CLAIMS IN BONUS-MALUS SYSTEMS

JEAN PINQUET[1]

THEMA, University Paris X, 92001 Nanterre, France

ABSTRACT

The objective of this paper is to make allowance for cost of claims in experience rating. We design here a bonus-malus system for the pure premium of insurance contracts, from a rating based on their individual characteristics Empirical results are presented, that are drawn from a French data base of automobile insurance contracts.

KEYWORDS

Bayesian and heterogeneous models Number and cost residuals. Bonus-malus for frequency of claims, average cost per claim, and pure premium.

INTRODUCTION

Bayesian models lead to a posteriori ratemaking of insurance contracts (Buhlmann (1967)) Suppose that the number of claims follows a Poisson distribution. A bonus-malus system for the frequency of claims is obtained if we consider that the parameter follows a gamma distribution (see Lemaire (1985, 1995)) This model may include a ratemaking of policyholders on an individual basis, the parameter of the Poisson distribution depending then on rating factors (see Dionne et al (1989, 1992)).

The allowance for severity of claims in experience rating can be achieved by considering the dichotomy between claims with material damage only, and claims including bodily injury (see Lemaire (1995)) In this model, the number of claims that caused bodily injury follows a binomial distribution, the parameter of which follows a beta distribution.

In this paper, the severity of claims will be taken into account by using their cost. The analysis of cost of claims makes clearly appear a positive correlation between the average cost per claim and the frequency risk (see Renshaw (1994), Pinquet et al (1992)) An a priori ratemaking will therefore be influenced by the allowance for costs Concerning the third party liability guaranty, it can be noted that.

- The settlement of claims with material damage is performed partly through fixed amount compensations from an insurance company to the third party

[1] Thanks to Georges Dionne for motivating this work, as well as Christian Gouriéroux, Eric Renshaw and two anonymous referees for comments This research received financial support from the Fédération Française des Sociétés d'Assurance

- The amount of compensations related to claims including bodily injury depends on the social position of the victim

Hence, it is difficult to explain the cost of these claims by the rating factors, and we shall investigate the damage guaranty in the empirical part of the paper

Allowing for cost of claims in bonus-malus systems can be achieved in the following way. starting from a rating model based on the analysis of number and cost of claims, two heterogeneity components are added They represent unobserved factors, that are relevant for the explanation of the severity variables Later on, we shall refer to any variable explained by a rating model (number, cost of claim, total cost of claims, and so on) as a "severity variable". These unobserved factors are, for instance, annual mileage for number distributions, and speed (and the driver's behaviour in general) for number and cost distributions. A bonus-malus coefficient can be related to the credibility estimation of a heterogeneity component

In this paper, costs of claims are supposed to follow gamma or log-normal distributions The rating factors, as well as the heterogeneity component, are included in the scale parameter of the distribution Considering that the heterogeneity component also follows a gamma or log-normal distribution, a credibility expression is obtained, which provides a predictor of the average cost per claim for the following period. For instance, a cost-bonus will appear after the first claim if its cost is inferior to the estimation made by the rating model

Experience rating with a bayesian model is possible only if there is enough heterogeneity in the data For instance, in the negative binomial model without covariates, the estimated variance of the heterogeneity component is equal to zero if the variance of the number of claims is inferior to their mean (see Pinquet et al (1992)) In that case, a priori and a posteriori tariff structures are the same, and the bayesian model fails.

A sufficient condition for the existence of a bonus-malus system derived from a bayesian model is provided in section 2 3 The existence is equivalent to an overdispersion of residuals related to the severity variable. This approach allows one to test for the presence of a hidden information, that is relevant for the explanation of the severity variables.

The heterogeneity on distributions for severity variables, that is not explained by the rating factors, is revealed through experience on policyholders The paper investigates the rate of this revelation, which is found to be lower for average cost per claim than for the frequency

For the sample considered here, the unexplained heterogeneity related to costs is stronger for gamma than for log-normal distributions Besides, the latter family gives a better fit to the data.

If the heterogeneity components on number and cost distributions are independent, the bonus-malus coefficient for pure premium is the product of the coefficients related to frequency and expected cost per claim. But one may think that the behavior of the policyholder influences the two heterogeneity components in a similar way, and so that they are positively correlated

Lastly, this paper proposes a bonus-malus system for the pure premium of insurance contracts, that admits a correlation between the two components Although the

likelihood of a model based on number and costs of claims is not analytically tractable in the presence of such a correlation, consistent estimators for the parameters exist. The correlation between the number and cost heterogeneity components appears to be very low for the sample investigated here

1 A PRIORI RATEMAKING

Let us suppose a sample of policyholders indexed by i, the policyholder i being observed during T_i periods The analysis of the correlation between the number and cost heterogeneity components shows the necessity of considering a non constant number of periods for each policyholder. The working sample is presented in 1 3

1.1 Frequency of claims

We write

$$N_{it} \sim P(\lambda_{it})_{t=1,\ \ldots,T_i},\ \lambda_{it} = \exp(w_{it}\,\alpha)$$

to represent the Poisson model where n_{it}, the outcome of N_{it}, is the number of claims reported by the policyholder i in period t The parameter λ_{it} is a multiplicative function of the explanatory variables, the line-vector w_{it} represents their values, and α is the column-vector of the related parameters.

The frequency-premium (estimation of the expectation of N_{it}) is denoted as $\hat{\lambda}_{it} = \exp(w_{it}\,\hat{\alpha})$, and $nres_{it} = n_{it} - \hat{\lambda}_{it}$ is the number-residual for the policyholder i and period t. The maximum likelihood estimator of α is the solution to the equation:

$$\sum_{i,t} nres_{it}\, w_{it} = 0,$$

which is an orthogonality relation between the explanatory variables and the residuals The rating factors have in general a finite number of levels, and the explanatory variables are then indicators of these levels The preceding equation means that, for every sub-sample associated to a given level, the sum of the frequency premiums is equal to the total number of claims This property means that the preceding model provides the multiplicative tariff structure that does not mutualize the frequency-risk.

One may think of replacing n_{it} by tc_{it}, the total cost of claims (pure premium ratemaking) in the likelihood equation. When applied to the working sample, this non probabilistic model shows that the elasticity of the pure premium risk with respect to the frequency risk is greater than one (see section 1.4.1).

1.2 Models for average cost per claim and pure premium

1.2.1 Gamma distributions

Let c_{ij} be the cost of the j^{th} claim reported by the policyholder i in period t ($1 \le j \le n_{it}$, *if* $n_{it} \ge 1$). We shall suppose in the paper that the costs are strictly positive. This assumption gives another reason to discard the third party liability guaranty: owing to fixed amount compensations, a policyholder involved in a claim caused by the third party can make his insurance company earn money.

Considering gamma distributions, we write

$$C_{itj} \sim \gamma(d, b_{it}), b_{it} = \exp(z_{it}\beta),$$

or $b_{it} C_{itj} \sim \gamma(d)$. The coefficient b_{it} is a scale parameter, a multiplicative function of the covariates, that are represented by the line-vector z_{it}.
Let $\hat{c}_{it} = \hat{d} / \hat{b}_{it} = \hat{d} / \exp(z_{it}\hat{\beta})$ be the estimation of the average cost for each claim reported by the policyholder i in period t. If we suppose that the costs are independent, the maximum likelihood estimator of β is the solution of the following equation.

$$\sum_{i,t} (n_{it} - (tc_{it} / \hat{c}_{it}))\, z_{it} = \sum_{i,t} cres_{it}\, z_{it} = 0$$

The term $n_{it} - (tc_{it} / \hat{c}_{it})$ is the sum, for the claims reported by the policyholder i in period t, of their cost residual $1 - (c_{itj} / \hat{c}_{it})$. It is written $cres_{it}$. The likelihood equation in β can hence be interpreted as an orthogonality relation between the explanatory variables and cost-residuals.

The average cost per claim increases with the frequency risk (see 1 4.2), which confirms the previous conclusions about the risks related to frequency and pure premium.

1.2.2 Log-normal distributions

The other distribution family considered in this paper is the normal distribution family for the logarithms of costs.

$$\log C_{itj} \sim N(z_{it}\beta, \sigma^2) \Leftrightarrow \log C_{itj} = z_{it}\beta + \varepsilon_{itj},\ \varepsilon_{itj} \sim N(0, \sigma^2).$$

The likelihood equation giving $\hat{\beta}$ is

$$\sum_{i,t} \left(\sum_{j} (\log c_{itj} - z_{it}\hat{\beta}) \right) z_{it} = \sum_{i,t} lcres_{it}\, z_{it} = 0.$$

This equation is also an orthogonality relation between explanatory variables and residuals.

1.2.3 Pure premium model

The total cost of claims reported by the policyholder i in period t may be written as:

$$TC_{it} = \sum_{j=1}^{N_{it}} C_{itj}$$

It is a sum of N_{it} i.i.d. outcomes from a variable that we denote as C_{it}. The pure premium is: $E(TC_{it}) = E(N_{it})\, E(C_{it})$.

1.3 Presentation of the working sample

The sample investigated in the paper is part of the automobile policyholders portfolio of a French insurance company It is composed of more than a hundred thousand policyholders The damage guaranty being considered here, only the contracts with that kind of guaranty were kept Policyholders can be observed over two years, and each anniversary date, changing of vehicle or coverage level entails a new period. Only claims concerning the damage guaranty and closed at the date of obtention of the data base were kept Reserved costs were thus avoided The rating factors retained for the estimation of number and cost distributions are

- The characteristics of the vehicle. group, class, age
- The characteristics of the insurance contract· type of use, level of the deductible, geographic zone

Other rating factors are the policyholder's occupation, as well as the year when the period began (in order to allow for a generation effect) These eight rating factors have a finite number of levels, the total number of which is 44 The explanatory variables are binary, and indicate the levels for the policyholders· in order to avoid collinearity, one level is suppressed for each rating factor, the intercept being kept anyway. Therefore, we shall consider (44-8)+1=37 covariates. With the notations of the paper, we obtain: $\alpha, \beta \in \mathbb{R}^{37}; w_{it}, z_{it} \in \{0,1\}^{37}$.

The estimated coefficients derived from the rating model depend on the level suppressed for each rating factor. Results that are independent from the suppressions are obtained by dividing the coefficients by their mean in the multiplicative model. These standardized coefficients can be compared with the relative severity of the levels

The periods having not the same duration, the parameter of the Poisson distribution must be proportional to the duration. The results given on the frequencies remain unchanged if, d_{it} being the duration of period t for the policyholder i, we write· $\lambda_{it} = d_{it} \exp(w_{it}\alpha)$, and $\hat{\lambda}_{it} = d_{it} \exp(w_{it}\hat{\alpha})$

The working sample includes 38772 policyholders and 71126 policyholders-periods These policyholders reported 3493 claims The average duration of the periods is nine months, and the annual frequency of the claims is 6 7%.

1.4 Empirical results

1.4.1 A priori rating for frequency and pure premium

When applied to the number of claims or their total cost, the Poisson models provide standardized coefficients, that can be compared with the relative severity of the levels For almost each rating factor, the variance of the coefficients related to the levels is inferior to the variance of the relative severity For instance, for the "type of use" rating factor, one gets

frequency	relative severity	standardized coefficient
professional use	1.623	1 278
standard use	0 982	0 992

pure premium	relative severity	standardized coefficient
professional use	1 747	1.177
standard use	0.979	0 995

The distributions of the policyholders among the levels of the different rating factors are not independent from one another Policyholders with a professional use have, for the other rating factors, more risky levels than the other policyholders The Poisson model does not mutualize the risk: hence these policyholders have, with respect to other rating factors, a level of relative severity equal to (1.747/1 177) – 1 = 48 4% more than the average, in term of pure premium.

The elasticity of the pure premium with respect to the frequency risk is equal to 1 52 on the sample, and the difference from 1 is significant (the related Student statistic is equal to 5.93) Hence, if the frequency risk is multiplied by two, the average cost per claim increases by $2^{0\,52} - 1 = 43.5\%$, and the pure premium increases by 187%.

This positive correlation between the risks on frequency and average cost per claim is observed on each rating factor, except for the geographical zone

1.4.2 A priori rating for average cost per claim

On the sample of claims, the gamma model leads to the following results (rating factor: type of use)

average cost	relative severity	standardized coefficient
professional use	1.076	0 933
standard use	0 996	1 003

The estimated elasticity of the average cost per claim with respect to the frequency is equal to 0 51, which confirms the results obtained in the preceding section.

2 EXPERIENCE RATING FOR FREQUENCY AND AVERAGE COST PER CLAIM

2.1 Heterogeneous models

In a bayesian framework, the allowance for a hidden information, relevant for the rating of risks, can be performed in the following way

- the starting point is an a priori rating model If y represents the severity variable(s), the likelihood of y will be written $f_0(y/\theta_1, x)$, where x is the vector of explanatory variables, and θ_1 the vector of parameters related to them
- A heterogeneity component (scalar, or vector) is added to the model, which measures the influence that unobserved variables have on the severity distribution. If u is this component, a distribution of y conditional on u and the explanatory variables is defined, and we denote its likelihood as $f_u(y/\theta_1, x, u)$ In practice, the a priori distribution is equal to the distribution defined conditionally on u, for some value u^0 of u $f_{u^0}(y/\theta_1, x, u^0) = f_0(y/\theta_1, x) \forall \theta_1, x, y$ If u is a scalar, $u^0 = 0$ or 1, according to the fact that u is included additively or multiplicatively in the conditional distribution

- The credibility estimation of u_i, the heterogeneity component for the policyholder i, leads to a bonus-malus system. It rests on a heterogeneous model, in which u_i is the outcome of a random variable U_i, the $(U_i)_{i=1,\ldots,p}$ being i i d. and their distribution being parameterized by θ_2. The likelihood of y_i in the model with heterogeneity is obtained by integrating the conditional likelihood over U_i, that is to say

$$f(y_i / \theta, x_i) = E_{\theta_2}[f_*(y_i / \theta_1, x_i, U_i)],$$

with $\theta = (\theta_1, \theta_2)$. The heterogeneity component vector on number and cost distributions will be denoted, for the policyholder i

$$U_i = \begin{pmatrix} U_{ni} \\ U_{ci} \end{pmatrix},$$

where n stands for the numbers and c for the costs The link between heterogeneous and bayesian models is made clear in the example that follows

2.2 Examples of heterogeneous models

2.2.1 Number of claims

With the notations of 1 1, the distributions defined conditionally on u_{ni} are

$$N_{it} \sim P(\lambda_{it} u_{ni}), \text{ with } U_{ni} \sim \gamma(a, a)$$

in the heterogeneous model The expectation of U_{ni} is equal to one, and its variance is $1/a$ On a period, the number of claims distribution is negative binomial in the heterogeneous model

The negative binomial model can be considered as a Poisson model with a random component, if we write $\lambda_{it} U_{ni} = \tilde{\lambda}_{it}$ If the intercept is the first of k explanatory variables, and if e_1 is the first vector of the canonical base of $\mathbb{R}^k$, we have

$$\tilde{\lambda}_{it} = \exp(w_{it}\,\alpha + \log(U_{ni})) = \exp(w_{it}\,(\alpha + \log(U_{ni})e_1)) = \exp(w_{it}\,\tilde{\alpha}_i)$$

In the last expression of λ_{it}, the parameter $\tilde{\alpha}_i = \alpha + \log(U_{ni})e_1$ is random, and the formulation is bayesian But it is less tractable than that of the heterogeneous model, as well for bonus-malus computations as for statistical inference.

2.2.2 Gamma distributions for costs of claims

The heterogeneous models that follow, which allow us to design bonus-malus systems for average cost per claim, suppose the independence of heterogeneity components on the number and costs distributions The empirical results presented later will make this assumption plausible.

For the gamma model and with the notations of 1.2 1, the distributions conditional on u_{ci} are

$$C_{itj} \sim \gamma(d, b_{it} u_{ci}), \text{ with } U_{ci} \sim \gamma(\delta, \delta)$$

in the heterogeneous model The heterogeneity component is included, as the rating factors, in the scale parameter of the distribution

In the heterogeneous model, one can write: $C_{itj} = D_{itj}/(b_{it}U_{ci})$, with $D_{itj} \sim \gamma(d)$, $U_{ci} \sim \gamma(\delta,\delta)$, D_{itj} and U_{ci} being independent. The variable C_{itj} follows a GB2 distribution (see Cummins et al. (1990)), and D_{itj} represents the relative severity of the claim.

2.2.3 Log-normal distributions for costs of claims

With the notations of 1.2.2, the heterogeneous model is

$$\log C_{itj} = z_{it}\beta + \varepsilon_{itj} + U_{ci},\ U_{ci} \sim N(0,\sigma_U^2),$$

where the ε_{itj} and U_{ci} are independent. The variable ε_{itj} represents the relative severity of the claim.

The heterogeneous model used to design a bonus-malus system for pure premium will be presented after the empirical results related to the preceding models.

2.3 A sufficient condition for the existence of a bonus-malus system derived from a bayesian model

Experience rating with a bayesian model is possible only if there exists enough heterogeneity on the data. Considering for instance the negative binomial model without covariates, the estimated variance of the heterogeneity component is equal to zero if the variance of the number of claims is lower than their mean (see Pinquet et al. (1992)). In that case, a priori and a posteriori tariff structures do not differ, and the bayesian model fails.

A sufficient condition for the existence of a bonus-malus system derived from a bayesian model is provided here: it will be applied later on to the models for number and cost of claims.

Let us start from a heterogeneous model, as defined in 2.1. The heterogeneity component is supposed to be scalar, and its distribution is parameterized by the variance σ^2. The parameters of the model are $\theta = (\theta_1, \sigma^2)$ and we shall write $\hat{\theta}^0 = (\hat{\theta}_1^0, 0), \hat{\theta}_1^0$ being the maximum likelihood estimator of θ_1 in the a priori rating model.

If the right-derivative, with respect to σ^2, of the log-likelihood is positive in $\hat{\theta}^0, \hat{\sigma}^2$ will be positive in the heterogeneous model. The existence of a bonus-malus system is hence related to the sign of a lagrangian, which is part of the score test for nullity of σ^2 (see Rao (1948), Silvey (1959)). With the notations of 2.1, and denoting the lagrangian as $\mathcal{L}$, one can prove:

$$\sum_i \log f(y_i/\hat{\theta}_1^0, \sigma^2, x_i) - \sum_i \log f_0(y_i/\hat{\theta}_1^0, x_i) = \mathcal{L}\sigma^2 + o(\sigma^2), \text{ with}$$

$$\mathcal{L} = \frac{1}{2}\sum_i (res_i^2 - s_i);$$

$$res_i = \left(\frac{\partial}{\partial u}\log f_*(y_i/\hat{\theta}_1^0, x_i, u)\right)_{u=u^0};\ s_i = -\left(\frac{\partial^2}{\partial u^2}\log f_*(y_i/\hat{\theta}_1^0, x_i, u)\right)_{u=u^0}$$

See Pinquet (1996b) for a proof, and references to a recent literature. The term res_t is a residual, which is related to those encountered in the likelihood equations for numbers and costs. The condition for existence of a bonus-malus system is

$$\mathcal{L} > 0 \Leftrightarrow \sum_t res_t^2 > \sum_t s_t$$

It can be interpreted as an overdispersion condition on residuals.

2.4 Prediction with heterogeneous models and bonus-malus systems

Let us suppose a policyholder observed on T periods: $Y_T = (y_1, \ldots, y_T)$ is the sequence of severity variables, and $X_T = (x_1, \ldots, x_T)$ that of the covariates. The sequences X_T and Y_T take the place of x_t and y_t in the preceding sections. The date of forecast T must be explicited here, and the individual index can be suppressed, since the policyholder can be considered separately. Besides, belonging to the working sample is not mandatory for this policyholder.

We want to predict a risk for the period $T+1$, by means of a heterogeneous model. For the period t, this risk R_t is the expectation of a function of Y_t (y_t is the outcome of Y_t). For instance, Y_t is the sequence of both number and costs of claims in period t, and R_t, the pure premium, is the expectation of the total cost.

We now include a heterogeneity component u, as defined in 2.1. The distribution of Y_t conditional on u depends on θ_1, x_t and u. This applies to R_t, and we can write $R_t = h_{\theta_1}(x_t)\, g(u)$, for the three types of risk dealt with later (frequency of claims, average cost per claim, pure premium), g being a real-valued function.

A predictor for the risk in period $T+1$ can be written as $h_{\theta_1}(x_{T+1})\, \hat{g}^{T+1}(u)$, with $\hat{g}^{T+1}(u)$ a credibility estimator of $g(u)$, defined from:

$$\hat{g}^{T+1}(u) = \arg\min_a E_{\theta_2}\left[(g(U)-a)^2\, f_*(Y_T/\theta_1, X_T, U)\right],$$

$$f_*(Y_T\ \theta_1, X_T, U) = \prod_{t=1}^{T} f_*(y_t\ \theta_1, x_t, U).$$

The expectation is taken with respect to U, and one obtains

$$\hat{g}^{T+1}(u) = E_\theta[g(U)/X_T, Y_T] = \frac{E_{\theta_2}[g(U)\, f_*(Y_T/\theta_1, X_T, U)]}{E_{\theta_2}[\, f_*(Y_T/\theta_1, X_T, U)]},$$

the expectation of $g(U)$ for the posterior distribution of U. Replacing θ_1 and θ_2 by their estimations in the heterogeneous model, we obtain the a posteriori premium

$$\hat{R}_{T+1}^{T+1} = h_{\hat{\theta}_1}(x_{T+1}) E_{\hat{\theta}}[g(U)/X_T, Y_T],$$

computed for period $T+1$. It can be written as

$$\left(h_{\hat{\theta}_1}(x_{T+1}) E_{\hat{\theta}_2}[g(U)]\right) \times \frac{E_{\hat{\theta}}[g(U)/x_1, \ldots, x_T; y_1, \ldots, y_T]}{E_{\hat{\theta}_2}[g(U)]}$$

The first term is an a priori premium, based on the rating factors of the current period. The second one is a bonus-malus coefficient it appears as the ratio of two expectations of the same variable, computed for prior and posterior distributions Owing to the equality $E_\theta[E_\theta(g(U)/\mathcal{X}_T,\mathcal{Y}_T)]=E_\theta[g(U)]=E_{\theta_2}[g(U)]$, the rating is balanced.

2.5 Bonus-malus for frequency of claims

2.5.1 Theoretical results

With the notations of 2 2.1 and 2.4, we write: $y_t=n_t$, $x_t=w_t$, $\theta_1=\alpha$; $R_t=E(N_t)=\lambda_t u$, $h_{\theta_1}(x_t)=\lambda_t$, $g(u)=u$; $\mathcal{X}_T=(w_1,\ ,w_T)$, $\mathcal{Y}_T=(n_1,\ ,n_T)$. The posterior distribution of U is a $\gamma(a+\sum_t n_t, a+\sum_t \lambda_t)$ (see Dionne et al (1989, 1992)) Hence:

$$E_\theta[U/w_1,\ ,w_T,n_1,\ldots,n_T]=\hat{u}^{T+1}=\frac{a+\sum_{t=1}^{T} n_t}{a+\sum_{t=1}^{T}\lambda_t} \tag{1}$$

Replacing λ_t by $\hat{\lambda}_t=\exp(w_t\hat{\alpha})$ and a by $\hat{a}$ in equation (1) leads to the bonus-malus coefficient. There will be a frequency-bonus if the estimator of $\hat{u}^{T+1}-1$ is negative, or if the number-residual $\sum_t(n_t-\hat{\lambda}_t)$ is negative

Considering in equation (1) that N_t follows a Poisson distribution, with a parameter $\lambda_t u$, $\hat{u}^{T+1}$ converges towards u when T goes to $+\infty$ The heterogeneity on number distributions, which is not explained by the rating factors, is hence revealed completely with time. It may be interesting to investigate the distribution of bonus-malus coefficients on a portfolio of policyholders, as well as its time evolution (see section 2 5.2 for empirical results)

We explicit now the condition for existence of a bonus-malus system for frequencies On the working sample, and with the notations in 2 2.1, one can write

$$\log f_n(y_t/\hat{\theta}_1^0,x_t,u)=\sum_t\left[n_{it}(\log\hat{\lambda}_{it}+\log u)-\hat{\lambda}_{it}u-\log(n_{it}!)\right],$$

with $\hat{\lambda}_{it}=\exp(w_{it}\hat{\alpha}^0)$, $\hat{\alpha}^0$ being the estimator of α in the a priori rating model With the notations of 2 3, and with $u^0=1$, we obtain

$$res_i=\sum_t\left(n_{it}-\hat{\lambda}_{it}\right), s_i=\sum_t n_{it}, \mathcal{L}>0\Leftrightarrow\sum_i nres_i^2>\sum_i n_i,$$

where $nres_i=\sum_t(n_{it}-\hat{\lambda}_{it})$ is the number-residual for policyholder i, and $n_i=\sum_t n_{it}$ is the number of claims reported by this policyholder on all periods This condition means that, considering the total number of claims, its variance is superior to its mean, the variance being calculated conditionally on the explanatory variables. This empirical overdispersion condition can be related to the theoretical overdispersion of the

negative binomial model: if $N_i \sim P(\lambda_i U_i), U_i \sim \gamma(a,a)$ (with $a = 1/\sigma^2$), one gets: $V(N_i) = \lambda_i + \lambda_i^2\sigma^2 > \lambda_i = E(N_i)$

A score test for nullity of σ^2 can be performed from the Lagrange multiplier $L = (1/2)\sum_i (nres_i^2 - n_i)$ The previous remarks allow us to reject the nullity of σ^2 if L is large enough If the number of policyholders goes to infinity, $\xi^L = L/\sqrt{\hat{V}(L)}$ converges towards a $N(0,1)$ distribution. One can prove that $\hat{V}(L) = 1/2\sum_i \hat{\lambda}_i^2$, with $\hat{\lambda}_i = \sum_i \hat{\lambda}_{ir}$ If $u_{1-\varepsilon}$ is the quantile at the level $1-\varepsilon$ of a $N(0,1)$ distribution, the null hypothesis $\sigma^2 = 0$ will be rejected at the level ε if $\xi^L \geq u_{1-\varepsilon}$.

Besides, the lagrangian provides an estimator of the parameters. Starting from $\hat{\alpha}^0$ and $\widehat{\sigma^2}^0 = 0$ in the algorithm of the likelihood maximisation, one gets at the following step

$$\hat{\alpha}^1 = \hat{\alpha}^0; \widehat{\sigma^2}^1 = \frac{L}{\hat{V}(L)} = \frac{\sum_i nres_i^2 - n_i}{\sum_i \hat{\lambda}_i^2} = \frac{\sum_i \left[(n_i - \hat{\lambda}_i)^2 - n_i\right]}{\sum_i \hat{\lambda}_i^2} \qquad (2)$$

The estimators $\hat{\alpha}^1$ and $\widehat{\sigma^2}^1$ can be shown to be consistent for the negative binomial model (see Pinquet (1996b) for demonstrations)

2.5.2 Empirical results

From the sample described in 1.3, we obtain

$$\sum_i nres_i^2 = \sum_i (n_i - \hat{\lambda}_i)^2 = 3709.24; \sum_i n_i = n = 3493,$$

and experience rating is possible for frequencies Without explanatory variables (apart from total duration of observation for each policyholder), one obtains: $\sum_i nres_i^2 = 3746\ 25$ The sum of square of residuals decreases when explanatory variables are added, and the condition for existence of a bonus-malus system is more restrictive when they are present. This is logical because they are a cause of heterogeneity on a priori distributions

Besides, $\sum_i \hat{\lambda}_i^2 = 389\ 48$, and the estimator of σ^2 given in (2) is

$$\hat{\sigma}^2 = \frac{L}{\hat{V}(L)} = \frac{\sum_i nres_i^2 - \sum_i n_i}{\sum_i \hat{\lambda}_i^2} = \frac{216.24}{389\ 48} = 0.555.$$

As a comparison, the maximum likelihood estimation for the negative binomial model is $\hat{\sigma}^2 = 0\ 576$. The score test for nullity of σ^2 is based on the statistic

$$\xi^{L} = \frac{L}{\sqrt{\hat{V}(L)}} = \frac{\sum_i nres_i^2 - \sum_i n_i}{\sqrt{2\sum_i \hat{\lambda}_i^2}} = \frac{216.24}{\sqrt{778.96}} = 7\,75,$$

and the null hypothesis is rejected Examples of bonus-malus coefficients derived from the credibility formula are developped in actuarial and econometric literature (see Lemaire (1985), Dionne et al (1989,1992))

Evolution throughout time of bonus-malus coefficients, as well as a posteriori premiums related to them, will be investigated for the risks related to frequency and average cost per claim We consider here a simulated portfolio, derived from the working sample In this portfolio, the characteristics of each policyholder in the sample are those of the first period, and we suppose that they remain unchanged If this assumption does not hold individually, it is however plausible on the whole population Investigating the distribution of bonus-malus coefficients in the heterogeneous model, one can measure their dispersion on the portfolio by estimating their coefficient of variation after T years (see Pinquet (1996a)) Considering the frequencies, with the tariff structure obtained in 1.4 1 and $\hat{\sigma}^2 = 0\,576$, we obtain:

TABLE 1
REVELATION THROUGHOUT TIME OF HETEROGENEITY RELATED TO NUMBER DISTRIBUTIONS

	Coefficients of variation (frequency of claims) a priori premium 0 372				
	T=1	**T=5**	**T=10**	**T=20**	**T=+∞**
bonus-malus coefficient	0 144	0 300	0 392	0 494	0 759
a posteriori premium	0 411	0 515	0 590	0 673	0 891

The coefficient of variation is a measure of the relative dispersion of bonus-malus coefficients and premiums Apart from the a priori premium, the elements of the preceding table are an estimation of the expectation in the heterogeneous model. After nine years, the relative dispersion of the bonus-malus coefficients exceeds that of the a priori premium. This means that, after nine years, the heterogeneity revealed by the observation of policyholders becomes more important than that explained by the rating factors.

2.6 Bonus-malus for average cost per claim (gamma distributions)

2.6.1 Theoretical results

With the notations in 2.2 2 and 2.4, we can write: $y_t = (c_{ij})_{j=1,\ n_t}, x_t = z_t$; $R_t = E(C_{ij}) = d/(b_t u); \theta_1 = (\beta, d); h_{\theta_1}(x_t) = d/b_t$; $g(u) = 1/u$. The bonus-malus coefficient on average cost per claim for period $T+1$ is derived from the credibility estimator

of $1/u$. Since the a priori distribution of U is a $\gamma(\delta,\delta)$, with a density proportional to $f_\delta(u)=\exp(-\delta u)u^{\delta-1}$, one gets:

$$f_\delta(u)\times f_*(Y_T/\theta_1,X_T,u)=\exp((\delta+\sum_{t,j}b_t c_{tj})u)u^{d(\sum_t n_t)+\delta-1},$$

times a coefficient independent of u. The posterior distribution of U is therefore a $\gamma(\delta+d(\sum_t n_t),\ \delta+\sum_{t,j}b_t c_{tj})$, and:

$$\widehat{1/u}^{T+1}=E_\theta\left[\frac{1}{U}\middle/X_T,Y_T\right]=\frac{\delta+\sum_{t,j}b_t c_{tj}}{\delta-1+d(\sum_t n_t)}$$

We have $E_{\theta_2}(1/U)=\delta/(\delta-1)$ (we suppose $\delta>1$, a necessary condition for $1/U$ to have a finite expectation). Omitting the period index, and writing S_T for the set of claims reported by the policyholder during the first T periods, the bonus-malus coefficient is

$$\frac{E_{\hat\theta}\left[\frac{1}{U}\middle/X_1,Y_1\right]}{E_{\hat\theta_2}\left[\frac{1}{U}\right]}=\frac{\hat\eta+\sum_{j\in S_1}(c_j/E_{\hat\theta}(C_j))}{\hat\eta+|S_T|},\qquad(3)$$

where we wrote: $\eta=(\delta-1)/d$, $E_\theta(C_j)=E_{\theta_2}(d/(b_jU))=(d/b_j)(\delta/(\delta-1))$. The rating structure derived from (3) is obviously balanced. Writing $E_{\hat\theta}(C_j)=\hat c_j$, and $cres_T=\sum_{j\in S_T}(1-(c_j/\hat c_j))$ the cost-residual for the policyholder, there will be a cost-bonus if the cost-residual is positive. The bonus is then equal to

$$1-\frac{\hat\eta+\sum_{j\in S_T}c_j/\hat c_j}{\hat\eta+|S_T|}=\frac{cres_T}{\hat\eta+|S_T|}$$

The time evolution of the distribution of bonus-malus coefficients is investigated in 2.6.2. Considering the simulated portfolio defined in 2.5.2, the heterogeneity unexplained by the rating factors is revealed more slowly for cost than for number distributions. This is not surprising, as far as no claim means no information on the cost distribution — if there is no correlation between the two heterogeneity components — whereas no claim generates frequency-bonus.

Let us apply to this model the condition allowing experience rating. For the working sample, we denote S_i as the set of claims reported by the policyholder over the T_i periods. One can write

$$\log f_*(y_i/\hat\theta_1^0,x_i,u)=\sum_{j\in S_i}(\hat d^0\log u-\hat b_{ij}^0 c_{ij}u)+z_i,$$

where z_i does not depend on u. With the notations of 2.3 and with $u^0=1$, we obtain:

$$res_i = \sum_{j \in S_i} (\hat{d}^0 - \hat{b}^0_{ij} c_{ij}); s_i = n_i \hat{d}^0; \mathcal{L} > 0 \Leftrightarrow \frac{1}{n} \sum_i cres_i^2 > \frac{1}{\hat{d}^0}$$

The total number of claims over the sample is n, and $cres_i$ is the cost-residual for the policyholder i. This residual is equal to 0 without claims, and otherwise. $cres_i = \sum_{j \in S_i} (1 - (c_{ij} / \hat{c}^0_{ij})) = \sum_{j \in S_i} cres_{ij}$, where $\hat{c}^0_{ij} = \hat{d}^0 / \hat{b}^0_{ij}$ is the estimator for the expectation of C_{ij}. Now, we have: $E(1 - (C_{ij} / E(C_{ij})))^2 = V(C_{ij}) / E^2(C_{ij}) = CV^2(C_{ij}) = 1/d$, if $C_{ij} \sim \gamma(d, b_{ij})$. The condition for existence of a bonus-malus system is hence related to the square of coefficients of variation.

2.6.2 Empirical results

Considering the working sample, one obtains:

$$\frac{1}{n} \sum_i cres_i^2 = 1.092; \frac{1}{\hat{d}^0} = 0.821,$$

and experience rating for average cost of claims is possible. For the sample of policyholders that reported claims, the maximum likelihood estimators for the GB2 model are:

$$\hat{\delta} = 3.620, \hat{d} = 1.807, \hat{\eta} = (\hat{\delta} - 1) / \hat{d} = 1.45.$$

The bonus (negative in case of malus) related to average cost per claim is equal to $cres_i / (\hat{\eta} + |S_i|)$. It remains equal to zero as long as there are no claims. After the first claim, if we consider the cases where the ratio actual cost-predicted cost is equal, either to 0.5 or to 2, the related cost-residuals are equal to 0.5 and -1 respectively. The multiplicative coefficient $1/(1+\hat{\eta})$ being equal to 0.408, we obtain a cost-bonus of 20.4% in the first case, and a cost-malus of 40.8% in the second case. This coefficient is independent of the period during which the claim occurs.

The distributions of bonus-malus coefficients and a posteriori premiums can be investigated on the simulated portfolio defined in 2.5.2. With the tariff structures obtained in 1.4.1 and 1.4.2 and $\hat{\delta} = 3.62$, we obtain (see Pinquet (1996a)):

TABLE 2
REVELATION THROUGHOUT TIME OF HETEROGENEITY RELATED TO COST DISTRIBUTIONS

	Coefficients of variation (expected cost per claim) a priori premium 0.401				
	T=1	**T=5**	**T=10**	**T=20**	**T=+∞**
bonus-malus coefficient	0.128	0.268	0.356	0.453	0.786
a posteriori premium	0.427	0.504	0.568	0.648	0.937

The relative dispersion of the bonus-malus coefficients exceeds the dispersion of the a priori premium after fourteen years. Unexplained heterogeneity on cost distributions is revealed more slowly than it was for numbers.

2.7 Bonus-malus for average cost per claim (log-normal distributions)

2.7.1 Theoretical results

With the notations in 2.2.2 and 2.4, we write $y_t = (\log c_{tj})_{j=1,\ldots,n_t}$; $r_t = z_t$, $\log C_{tj} \sim N(z_t\beta + u, \sigma^2) \Rightarrow R_t = E(C_{tj}) = \exp(z_t\beta + u + (\sigma^2/2))$, $\theta_1 = (\beta, \sigma^2)$, $h_{\theta_1}(x_t) = \exp(z_t\beta + (\sigma^2/2))$; $g(u) = \exp(u)$. The bonus-malus coefficient is derived from the credibility estimator of $\exp(u)$. Now

$$f_{\sigma_U^2}(u) \times f_*(Y_T/\theta_1, X_T, u) = \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_U^2} + \frac{tn_T}{\sigma^2}\right)\left(u - \frac{tlc_T - E_{\theta_1}(TLC_T)}{tn_T + (\sigma^2/\sigma_U^2)}\right)^2\right]$$

times a coefficient independent from u. We wrote $tn_T = \sum_{t=1}^{T} n_t$, $tlc_T = \sum_{j \in S_T} \log c_j$, $E_{\theta_1}(TLC_T) = \sum_{j \in S_T} E_{\theta_1}(\log C_j)$; S_T is the set of claims reported by the policyholder during the T periods $\left(|S_T| = tn_T\right)$, and the period index is omitted. Hence, the posterior distribution of U is

$$U/(X_T, Y_T) \sim N\left(\frac{tlc_T - E_{\theta_1}(TLC_T)}{tn_T + (\sigma^2/\sigma_U^2)}, \frac{1}{(1/\sigma_U^2) + (tn_T/\sigma^2)}\right)$$

The bonus-malus coefficient for period $T+1$ is equal to

$$\frac{E_{\hat{\theta}}[\exp(U)/X_T, Y_T]}{E_{\hat{\theta}_2}[\exp(U)]} = \exp\left[\frac{lcres_T - (tn_T\hat{\sigma}_U^2/2)}{(\hat{\sigma}^2/\hat{\sigma}_U^2) + tn_T}\right],$$

writing $lcres_T = \sum_{j \in S_T} lcres_j$, $lcres_j = \log c_j - E_{\hat{\theta}_1}(\log C_j)$.

The condition for existence of a bonus-malus sytem is easily interpretable with the log-normal model. We have

$$\log f_*(y_t/\hat{\theta}_1^0, x_t, u) = -\sum_{j \in S_t} \frac{(lcres_{tj} - u)^2}{2\,\widehat{\sigma^2}^0}$$

plus terms that do not depend on u, with $lcres_{tj} = \log(c_{tj}) - z_{tj}\hat{\beta}^0$. With $u^0 = 0$ (see 2.3), the existence condition is:

$$\sum_t \frac{\left(\sum_{j \in S_t} lcres_{tj}\right)^2}{(\widehat{\sigma^2}^0)^2} - \frac{n}{\widehat{\sigma^2}^0} = \frac{1}{(\widehat{\sigma^2}^0)^2}\left[\sum_t \left(\sum_{j \in S_t} lcres_{tj}\right)^2 - n\,\widehat{\sigma^2}^0\right] > 0$$

Now, in the a priori rating model, $n\widehat{\sigma^2}^0 = \sum_{t,j} lcres_{tj}^2$, with $\widehat{\sigma^2}^0$ the maximum likelihood estimator of σ^2. Experience rating is possible if

$\sum_i \left(\sum_{j \in S_i} lcres_{ij} \right)^2 - \sum_{i,j} lcres_{ij}^2$ is positive, that is to say if

$$\sum_{i/n_i \geq 2} \sum_{j,k \in S_i, j \neq k} lcres_{ij}\, lcres_{ik} > 0$$

This condition means that, for claims related to policyholders having reported several of them, cost-residuals have rather the same sign. If the first claim has a cost greater than its prediction, it will be the same on average for the following ones.

One can prove that, if $\mathcal{L}$ is the lagrangian with respect to σ_U^2, we have

$$\hat{V}(\mathcal{L}) = \frac{\sum_i n_i(n_i - 1)}{2(\widehat{\sigma^2}^0)^2} \Rightarrow \widehat{\sigma_U^2}^1 = \frac{\mathcal{L}}{\hat{V}(\mathcal{L})} = \frac{\sum_{i/n_i \geq 2} \sum_{j,k \in S_i, j \neq k} lcres_{ij}\, lcres_{ik}}{\sum_i n_i(n_i - 1)},$$

and that $\widehat{\sigma_U^2}^1$ is an consistent estimator of σ_U^2 (see Pinquet (1996a)). It appears to be the average, for the policyholders having reported several claims, of the product of residuals associated to couples of different claims

2.7.2 Empirical results

From the working sample, we obtain $\sum_{i/n_i \geq 2} \sum_{j,k \in S_i, j \neq k} lcres_{ij}\, lcres_{ik} = 100\,80$, and experience rating is possible Hence

$$\widehat{\sigma_U^2}^1 = \frac{\sum_{i/n_i \geq 2} \sum_{j,k \in S_i, j \neq k} lcres_{ij}\, lcres_{ik}}{\sum_i n_i(n_i - 1)} = \frac{100\,80}{590} = 0.171.$$

The nullity of σ_U^2 is tested for with $\xi^L = \mathcal{L} / \sqrt{\hat{V}(\mathcal{L})} = 2.86$ The critical value for a one-sided test at a level of 5% is 1.645, and the null hypothesis is rejected The maximum likelihood estimators of σ_U^2 and σ^2 in the heterogeneous model are: $\hat{\sigma}_U^2 = 0\,172$, $\hat{\sigma}^2 = 0.855$.

Bonus-malus coefficients can be computed from the examples considered with the gamma distributions (one claim, and a ratio actual cost-expected cost equal to 0 5 or 2) The residual associated to a claim is the logarithm of the latter ratio In the first case, the bonus-malus coefficient is equal to

$$\exp\left[\frac{lcres_T - (tn_T \hat{\sigma}_U^2 / 2)}{\left(\hat{\sigma}^2 / \hat{\sigma}_U^2\right) + tn_T} \right] = \exp\left[\frac{-\log 2 - 0.086}{(0\,855 / 0.172) + 1} \right] = 0\,878,$$

and is associated to a cost-bonus of 12 2% In the second case, the bonus-malus coefficient is equal to 1 107, and implies a cost-malus of 10 7% These results can be compared with 20 4% and 40.8%, the boni and mali derived from the gamma distributions, although the ratios actual cost-expected cost are different in the two models. They

must be different, since the cost-residuals in the gamma and log-normal models are equal to $1-(c_{ij}/\hat{c}_{ij}^{\,gamma})$ and $\log(c_{ij}/\hat{c}_{ij}^{\,\log\text{-}normal})$ respectively, whereas they fulfill the same orthogonality relations with respect to the covariates.

Considering the simulated portfolio defined in 2.5.2, the heterogeneity on cost distributions that is unexplained by the a priori rating model is more important for gamma than for log-normal distributions This can be seen by comparing the limits of the coefficients of variation for the bonus-malus coefficients, as we did in sections 2 5 2 and 2 6 2 For the GB2 model, this limit is the coefficient of variation of $1/U, U \sim \gamma(\hat{\delta},\hat{\delta})$ (see Pinquet (1996a)) With $\hat{\delta}=3.62$, it is equal to $1/\sqrt{\hat{\delta}-2}=0\,786$ Considering the log-normal model, the limit is the coefficient of variation of $\exp(U)$, $U \sim N(0,\hat{\sigma}_U^2)$

With $\hat{\sigma}_U^2=0.172$, it is equal to $\sqrt{\exp(\hat{\sigma}_U^2)-1}=0\,433$.

This result can be related to a comparison between the two a priori rating models If F_{θ_1,x_j} is the continuous distribution function of Y_j (here equal to the cost of the claim j, or its logarithm) $\varepsilon_j=F_{\theta_1,x_j}(Y_j)$ is uniformly distributed on [0,1] Computing the residuals $e_j, e_j=F_{\hat{\theta}_1^0,x_j}(Y_j)$, and rearranging e_j in the increasing order, by $e_{(1)}\leq .\leq e_{(n)}$, we derive the Komolgorov-Smirnov statistic $KS=\sqrt{n}\max_{1\leq j\leq n}|(j/n)-e_{(j)}|$ We obtain KS=2 83 (resp KS=1.04) for the gamma (resp log-normal) distribution family. The latter family seems to fit the data better than the gamma family, and will be retained for the bonus-malus system on pure premium

The two last results can be related to each other. there is more unexplained heterogeneity for gamma than for log-normal distributions, and the latter provide a better fit to the data This fact raises a question: is apparent heterogeneity only explained by hidden information, or can it be also explained by the fact that the model does not make the best use of observable information?

3 BONUS-MALUS FOR PURE PREMIUM

3.1 The heterogeneous model

From the preceding results, we shall retain log-normal rather than gamma distributions for costs Besides, they are better integrated in a heterogeneous model with a joint distribution for the two heterogeneity components related to the number and cost distributions We retain here a bivariate normal distribution The parameters of the related heterogeneous model can be estimated consistently, although the likelihood is not analytically tractable

A way to derive consistent estimators for heterogeneous models is proposed in Pinquet (1996b) It is based on the properties of extremal estimators, the maximum likelihood estimator being of this type. The estimators of the parameters of the a priori

rating model have a limit if the actual distributions include heterogeneity, and this limit is tractable in the model investigated here Consistent estimators are then obtained from a method of moments using the scores with respect to the variances and the covariances of the heterogeneity components

The heterogeneous model is hence composed of Poisson distributions on numbers, log-normal distributions on costs, and of bivariate normal distributions for the two heterogeneity components. The notations are the following.

- The distributions conditional on u_{ni} and u_{ci}, the heterogeneity components for number and cost distributions of the policyholder i, are

$$N_{it} \sim P(\lambda_{it} \exp(u_{ni})), \log C_{itj} = z_{it}\beta + \varepsilon_{itj} + u_{ci}, \text{ with}$$

$$\lambda_{it} = \exp(w_{it}\alpha), \varepsilon_{itj} \sim N(0, \sigma^2), t = 1, \ldots, T_i; j = 1, \ldots, n_{it}$$

- In the heterogeneous model, U_{ni} and U_{ci} follow a bivariate normal distribution with a null expectation and a variance equal to

$$V = \begin{pmatrix} V_{nn} & V_{nc} \\ V_{cn} & V_{cc} \end{pmatrix}.$$

The parameters of the model are

$$\theta_1 = \begin{pmatrix} \alpha \\ \beta \\ \sigma^2 \end{pmatrix}, \theta_2 = \begin{pmatrix} V_{nn} \\ V_{cn} \\ V_{cc} \end{pmatrix}$$

Bonus-malus coefficients are computed in the heterogeneous model from the expression given in section 2.4

$$\frac{E_{\hat{\theta}}[g(U)/X_1, Y_T]}{E_{\hat{\theta}_2}[g(U)]} = \frac{E_{\hat{\theta}_2}\left[g(U) f(Y_1/\hat{\theta}_1, X_T, U)\right]}{E_{\hat{\theta}_2}[g(U)]\, E_{\hat{\theta}_2}\left[g(U) f(Y_1/\hat{\theta}_1, X_T, U)\right]} \tag{4}$$

We can write.

- $g(u_n, u_c) = \exp(u_n)$ for frequency
- $g(u_n, u_c) = \exp(u_c)$ for average cost per claim
- $g(u_n, u_c) = \exp(u_n + u_c)$ for pure premium,

because the expectations of N_i, C_{ij} and TC_i are respectively proportional to $\exp(u_n)$, $\exp(u_c)$ and $\exp(u_n + u_c)$, if computed conditionally on u_n and u_c The mathematical expectations that lead to the bonus-malus coefficients (see equation (4)) can be estimated if we can write $U = f_{\theta_2}(S)$, where the distribution of S is independent from θ_2 it is enough to simulate outcomes of S Such an expression can be obtained by writing the Choleski decomposition of the variances-covariances matrix, i e.

$$V = \begin{pmatrix} V_{nn} & V_{nc} \\ V_{cn} & V_{cc} \end{pmatrix} = T_\varphi T_\varphi'; \; T_\varphi = \begin{pmatrix} \varphi_{nn} & 0 \\ \varphi_{cn} & \varphi_{cc} \end{pmatrix} \Rightarrow V = \begin{pmatrix} \varphi_{nn}^2 & \varphi_{nn}\varphi_{cn} \\ \varphi_{nn}\varphi_{cn} & \varphi_{nn}^2 + \varphi_{cc}^2 \end{pmatrix}$$

One can write for the policyholder i

$$U_i = \begin{pmatrix} U_{ni} \\ U_{ci} \end{pmatrix} = T_\varphi S_i ; S_i = \begin{pmatrix} S_{ni} \\ S_{ci} \end{pmatrix}, S_i \sim N(0, I_2),$$

and we have $U_i = f_{\theta_2}(S_i)$, φ being related to V, hence to θ_2. The likelihood used in the bonus-malus expression (see equation (4)) is obtained as the product of the likelihoods related to numbers and costs. With the notations of 2.4, we have

$$\log f_*(Y_I / \theta_1, X_I, U) =$$

$$-\left(\sum_i \lambda_i\right)\exp(U_n) + \left(\sum_i n_i\right) U_n - \sum_{i,j} \frac{(\log c_{ij} - z_i\beta - U_c)^2}{2\sigma^2}, \text{ with}$$

$$X_I = (x_1, \ldots, x_T); x_i = (w_i, z_i), Y_T = (y_1, \ldots, y_T), y_i = (n_i, (c_{ij})_{j=1, \ldots, n_i}),$$

plus terms that do not depend on the heterogeneity components. Replacing θ_1 by $\hat{\theta}_1$, we obtain

$$f_*(Y_T / \hat{\theta}_1, X_I, U) = \exp(V_T) \times \text{terms independent from } U, \text{ with}$$

$$V_T = -\left(\sum_i \hat{\lambda}_i\right)\exp(U_n) + m_T U_n - \frac{m_T U_c^2 - 2U_c\, lcres_T}{2\hat{\sigma}^2} \tag{5}$$

A bonus-malus coefficient for a policyholder and for the period T+1 depends then on:

- $\sum_i \hat{\lambda}_i$, which is proportional to the frequency premium of the policyholder on all periods. This premium is equal to

$$\hat{E}(TN_T) = \sum_i \hat{\lambda}_i \, \hat{E}\left[\exp(U_n)\right] = \left(\sum_i \hat{\lambda}_i\right)\exp\frac{\hat{\varphi}_{nn}^2}{2} = \left(\sum_i \hat{\lambda}_i\right)\exp\frac{\hat{V}_{nn}}{2}.$$

- m_T, the number of claims reported by the policyholder during the T periods
- $lcres_I$, the sum of residuals on the logarithm of costs of claims reported by the policyholder: it represents their relative severity.

From equation (4), bonus-malus coefficients on frequency, expected cost per claim, and pure premium are respectively equal to

$$\frac{\hat{E}[\exp(U_n + V_I)]}{\hat{E}[\exp(U_n)]\,\hat{E}[\exp(V_I)]}, \frac{\hat{E}[\exp(U_c + V_I)]}{\hat{E}[\exp(U_c)]\,\hat{E}[\exp(V_I)]}, \frac{\hat{E}[\exp(U_n + U_c + V_T)]}{\hat{E}[\exp(U_n + U_c)]\,\hat{E}[\exp(V_T)]}.$$

The coefficients are estimated by simulations of outcomes of S_n and S_c. For instance, we infer that the estimated covariance

$$\widehat{Cov}\left(\frac{\exp(U_n)}{E[\exp(U_n)]}, \frac{\exp(V_I)}{E[\exp(V_I)]}\right)$$

is a frequency-malus The existence of boni and mali for the different risks can be interpreted through the sign of estimated covariances

The a posteriori premium is obtained by the expression given in section 2 4

$$\hat{R}_{I+1}^{T+1} = \left(h_{\hat{\theta}_1}(x_{T+1}) E_{\hat{\theta}_2}[g(U)] \right) \frac{E_{\hat{\theta}}[g(U)/X_T, Y_T]}{E_{\hat{\theta}_2}[g(U)]}$$

The first term is the a priori premium It is an estimation of

$$\lambda_{T+1} \exp(z_{T+1}\beta) E[\exp(U_n + U_c)] = \exp\left(w_{T+1}\alpha + z_{T+1}\beta + \frac{(\varphi_{nn} + \varphi_{cn})^2 + \varphi_{cc}^2}{2} \right),$$

because $U_n + U_c = (\varphi_{nn} + \varphi_{cn}) S_n + \varphi_{cc} S_c$.

Besides, $(\varphi_{nn} + \varphi_{cn})^2 + \varphi_{cc}^2 = V_{nn} + 2V_{cn} + V_{cc}$.

We should have consistent estimators for the parameters, in order to derive bonus-malus coefficients. A method to obtain such estimators was quoted in the introduction. When applied to the preceding model, it leads to the following results

We write $\hat{\alpha}^0, \hat{\beta}^0, \widehat{\sigma^2}^0$ the estimators of the parameters in the a priori rating model, and $\hat{\lambda}_i = \sum_t \exp(w_{it}\hat{\alpha}^0), tlc_i = \sum_{t,j} \log(c_{itj}), E_{\theta_1}(TLC_i) = \sum_t n_{it} z_{it}\beta, \hat{tlc}_i = E_{\hat{\theta}_1^0}(TLC_i) = \sum_t n_{it} z_{it}\hat{\beta}^0$

The variances and covariances of the two heterogeneity components are consistently estimated by:

$$\hat{V}_{nn} = \log(1 + \hat{V}_{nn}^1), \hat{V}_{nn}^1 = \frac{\sum_i (n_i - \hat{\lambda}_i)^2 - n_t}{\sum_i \hat{\lambda}_i^2}; \hat{V}_{cn} = \frac{\sum_i (n_i - \hat{\lambda}_i)(tlc_i - \hat{tlc}_i)}{\left(\sum_i \hat{\lambda}_i^2 \right)(1 + \hat{V}_{nn}^1)},$$

$$\hat{V}_{cc} = \frac{\sum_i \left[(tlc_i - \hat{tlc}_i)^2 - n_i \widehat{\sigma^2}^0 \right]}{\left(\sum_i \hat{\lambda}_i^2 \right)(1 + \hat{V}_{nn}^1)} - \hat{V}_{cn}^2 \tag{6}$$

Consistent estimators of φ_{nn}, φ_{cn} and φ_{cc} are given by the solutions of the equation

$$T_{\hat{\varphi}} T_{\hat{\varphi}}' = \hat{V}$$

The estimators of φ are used in the computation of bonus-malus coefficients. remember that $U_i = T_\varphi S_i$ ($S_i \sim N(0, I_2)$), and that the coefficients are estimated through simulations of outcomes of S_i. As for the parameters of the a priori rating model, they are consistently estimated by

$$\hat{\alpha} = \hat{\alpha}^0 - \frac{\hat{V}_{nn}}{2} e_{n,1}, \hat{\beta} = \hat{\beta}^0 - \hat{V}_{cn} e_{c,1}, \hat{\sigma}^2 = \widehat{\sigma^2}^0 - \hat{V}_{cc} \tag{7}$$

The intercepts are supposed to be the first of the k_n and k_c explanatory variables for the number and cost distributions, and $e_{n,1}$ (resp $e_{c,1}$) are the first vectors of the canonical base of $\mathbb{R}^{k_n}$ (resp $\mathbb{R}^{k_c}$)

3. 2 Empirical results

The numerical results $\sum_t (n_t - \hat{\lambda}_t)^2 - n_t = 216.24$; $\sum_t \hat{\lambda}_t^2 = 389\,48$, already used for bonus-malus on frequencies, lead to.

$$\hat{V}_{nn}^1 = \frac{\sum_t (n_t - \hat{\lambda}_t)^2 - n_t}{\sum_t \hat{\lambda}_t^2} = 0.555, \hat{V}_{nn} = \log(1 + \hat{V}_{nn}^1) = 0\,442 \Rightarrow \hat{\varphi}_{nn} = \sqrt{\hat{V}_{nn}} = 0\,665$$

In this paper, two distribution families are considered for the heterogeneity component related to numbers We first took into account the gamma, and now the log-normal family (writing the heterogeneity component in a multiplicative way)

Considering an insurance contract without claims, we can compare the boni derived from the two models The sum $\sum_t \hat{\lambda}_t$ being the cumulated frequency premium in the negative binomial model, the bonus for the policyholder is equal to

$$1 - \frac{\hat{a}}{\hat{a} + \sum_t \hat{\lambda}_t} = \frac{\sum_t \hat{\lambda}_t}{\hat{a} + \sum_t \hat{\lambda}_t} = \frac{\hat{V}_{nn}^1 \sum_t \hat{\lambda}_t}{1 + \left(\hat{V}_{nn}^1 \sum_t \hat{\lambda}_t\right)}, (\hat{a} = 1 / \hat{V}_{nn}^1).$$

For the log-normal family, the bonus can be written as

$$-\widehat{Cov}\left(\frac{\exp(U_n)}{E[\exp(U_n)]}, \frac{\exp(V_T)}{E[\exp(V_T)]}\right), U_n = \varphi_{nn} S_n, V_T = -\sum_t \hat{\lambda}_t \exp(U_n),$$

with $S_n \sim N(0,1)$ With the values of $\hat{V}_{nn}^1$ and $\hat{\varphi}_{nn}$ computed precendently, one obtains for example

TABLE 3

COMPARISON OF FREQUENCY-BONUS COEFFICIENTS FOR TWO DISTRIBUTIONS ON THE HETEROGENEITY COMPONENT (CONTRACTS WITHOUT CLAIMS REPORTED)

frequency premium	**0.05**	**0.1**	**0.2**	**0.5**	**1**	**2**
bonus (%, gamma distributions)	2 7	5 3	10	21 7	35 7	52 6
bonus (%, log-normal distributions)	2 6	5 1	9 4	19 3	30 3	43 6

The boni derived from log-normal distributions on the heterogeneity component are lower than those derived from the gamma distributions. The difference is all the more important since the frequency premium is high

Let us estimate the covariance between the two heterogeneity components:

$$\sum_i (n_i - \hat{\lambda}_i)(tlc_i - t\hat{l}c_i) = 7.96 \Rightarrow \hat{V}_{cn} = \frac{\sum_i (n_i - \hat{\lambda}_i)(tlc_i - t\hat{l}c_i)}{\left(\sum_i \hat{\lambda}_i^2\right)(1 + \hat{V}_{nn}^1)} = 0\,013.$$

One can think of relating a positive or negative sign of the covariance to the fact that the average cost per claim increases or decreases with the number of claims reported by the policyholder To see this, suppose that the duration of observation is the same for all the policyholders, and that the intercept is the only explanatory variable for number and cost distributions We would then have

$$\hat{\lambda}_i = \bar{n},\ t\hat{l}c_i = n_i \overline{\log c} \Rightarrow \sum_i (n_i - \hat{\lambda}_i)(tlc_i - t\hat{l}c_i) = \sum_i (n_i - \bar{n}) n_i (\overline{\log c}^i - \overline{\log c}) =$$

$$\sum_{i/n_i \geq 2} (n_i - 1) n_i (\overline{\log c}^i - \overline{\log c}),\ \text{because} \sum_i n_i (\overline{\log c}^i - \overline{\log c}) = 0.$$

We wrote $\overline{\log c}^i$ for the logarithms of costs of claims reported by the policyholder i, computed on average. The estimator of the covariance would be positive if the average of the logarithms of costs of claims related to the policyholders that reported several of them was superior to the global mean

On the working sample, the number of claims reported by the policyholder had little influence on the average cost

The preceding results justify the allowance for a non constant number of periods related to the observation of policyholders To see this, we remark that the more severe is a claim, the greater is the probability to change the vehicule afterwards. Hence, there is less severity on average for several claims reported on the same car If policyholders were not kept in the sample after changing cars, a negative bias would appear in the estimation of the correlation coefficient between the heterogeneity components. Now, keeping the policyholder in the sample as long as possible leads us to consider a non constant number of periods.

When computing bonus-malus coefficients for average cost per claim, we used (see 2 7 2)

$$\sum_i \left[(tlc_i - t\hat{l}c_i)^2 - n_i \widehat{\sigma^2}^0\right] = \sum_{i/n_i \geq 2} \sum_{j\ k \in S_i, j \neq k} lcres_{ij}\, lcres_{ik} = 100\,80$$

A bonus-malus system for average cost per claim can be considered if the observation of the ratio actual cost-expected cost for a claim brings information for the following claims. If the last expression is positive, the cost residuals of claims related to policyholders having reported several of them have rather the same sign The relative severity of a claim is associated to the sign of the residual, and it may be interesting to compare the sign of residuals for claims related to policyholders having reported two of them.

Considering the working sample, we obtain

number of policyholders having reported two claims	negative residual (second claim)	positive residual (second claim)
negative residual (first claim)	74	46
positive residual (first claim)	36	70

The sign of the residual does not change for 64% of policyholders having reported two claims

From equation (6), we infer

$$\hat{V}_{cc} = \frac{\sum_i (tlc_i - t\hat{l}c_i)^2 - n_i \widehat{\sigma^2}^0}{\left(\sum_i \hat{\lambda}_i^2\right)(1+\hat{V}_{nn}^1)} - \hat{V}_{cn}^2 = 0\ 166, \text{ and } \hat{r}_{cn} = \frac{\hat{V}_{cn}}{\sqrt{\hat{V}_{cc}\hat{V}_{nn}}} = 0\ 048$$

The correlation coefficient between the heterogeneity components is positive, but close to zero Hence

$$\hat{V}_{cn} = \hat{\varphi}_{nn}\hat{\varphi}_{cn} \Rightarrow \hat{\varphi}_{cn} = 0.020,\ \hat{V}_{cc} = \hat{\varphi}_{cn}^2 + \hat{\varphi}_{cc}^2 \Rightarrow \hat{\varphi}_{cc} = 0\ 407$$

The boni for average cost per claim and pure premium for the contracts without claims can be computed, and results can be compared to those obtained for frequency. From the expressions

$$-\widehat{Cov}\left(\frac{\exp(U_c)}{E[\exp(U_c)]}, \frac{\exp(V_T)}{E[\exp(V_T)]}\right), -\widehat{Cov}\left(\frac{\exp(U_n+U_c)}{E[\exp(U_n+U_c)]}, \frac{\exp(V_T)}{E[\exp(V_T)]}\right)$$

we obtain

TABLE 4

BONI FOR AVERAGE COST PER CLAIM AND PURE PREMIUM (CONTRACTS WITHOUT CLAIM REPORTED)

frequency premium	**0.05**	**0.1**	**0.2**	**0.5**	**1**	**2**
average cost per claim bonus (%)	0 1	0 1	0 2	0 5	0 9	1 5
pure premium bonus (%)	2 7	5 3	9 7	19 9	31 2	44 7

Because of the positive correlation between the two heterogeneity components, a cost-bonus appears in the absence of claims, but it is very low.

We now compute bonus-malus coefficients for policyholders that reported one claim They are a function of the cost-residual $lcres_1 = \log(c_1) - z_1\hat{\beta}$ (c_1 is the cost of the claim, and z_1 represents the policyholder's characteristics when the claim occured), and of the frequency premium From equations (5) and (7), we have

$$V_T = -\sum_t \hat{\lambda}_t \exp(U_n) + U_n - \frac{U_c^2 - 2U_c lcres_T}{2\hat{\sigma}^2},$$

$$\hat{\sigma}^2 = \widehat{\sigma^2}^0 - \hat{V}_{cc} = \frac{\sum_{i,j} lcres_{ij}^2}{n} - \hat{V}_{cc} = \frac{3588}{3493} - 0.166 = 0.861$$

We recall that the bonus-malus coefficients on frequency, expected cost per claim and pure premium are respectively equal to

$$\frac{\hat{E}[\exp(U_n + V_T)]}{\hat{E}[\exp(U_n)]\,\hat{E}[\exp(V_T)]}; \frac{\hat{E}[\exp(U_c + V_T)]}{\hat{E}[\exp(U_c)]\,\hat{E}[\exp(V_T)]}; \frac{\hat{E}[\exp(U_n + U_c + V_T)]}{\hat{E}[\exp(U_n + U_c)]\,\hat{E}[\exp(V_T)]}.$$

We obtain for example (the bonus-malus coefficients are given in percentage)

TABLE 5

BONUS-MALUS COEFFICIENTS (POLICYHOLDERS HAVING REPORTED ONE CLAIM)

frequency coefficient	frequency premium					
$lcres_T$	0.05	0.1	0.2	0.5	1	2
-1	147 4	142 1	133 1	113 9	94 5	73 4
-0 5	148 4	143	133 8	114 5	95	73 7
0	149 3	143 7	134 6	115	95 3	74
0 5	150 1	144 6	135 3	115 6	95 7	74 3
1	151	145 6	136	116 1	96 2	74 6

average cost per claim coefficient	frequency premium					
$lcres_T$	0.05	0.1	0.2	0.5	1	2
-1	84 8	84 7	84 6	84 3	84	83 5
-0 5	92	91 9	91 7	91 4	91	90 5
0	99 7	99 6	99 5	99 1	98 7	98 1
0 5	108 1	108	107 8	107 5	107	106 4
1	117 1	117	116 9	116 5	116	115 4

pure premium coefficient	frequency premium					
$lcres_T$	0.05	0.1	0.2	0.5	1	2
-1	124 6	120	112 2	95 6	78 9	60 9
-0 5	136 1	131	122 3	104 2	86	66 3
0	148 4	142 7	133 3	113 5	93 5	72 2
0 5	161 8	155 7	145 4	123 7	101 9	78 5
1	176 6	170	158 4	134 7	111	85 4

Because of the positive correlation between the two heterogeneity components, the frequency coefficients increase with the cost-residual, which is related to the severity of the claim In the same way, the coefficients related to average cost per claim decrease with the frequency premium, but these variations are very low Because of the correlation, the coefficients related to pure premium are not equal to the product of the

coefficients for frequency and expected cost per claim. Here also, differences are very low

4. CONCLUDING REMARKS

We recall the main results obtained in this paper

- The unexplained heterogeneity with respect to the cost distributions depends strongly on the choice of the distribution family.
- Besides, it is revealed more slowly throughout time than for number distributions
- On the working sample, the correlation between the heterogeneity components on the number and cost distributions is very low.

In the long run, it would be desirable to relax the assumption of invariance of the heterogeneity components with respect to time Because of this invariance, the age of claims has no influence on the bonus-malus coefficients Now, the fact that an ancient claim has the same influence on the coefficients that a recent one is questionable. The allowance for an innovation at each period for the heterogeneity components would raise new problems, and would make it necessary to observe policyholders on many periods.

REFERENCES

BUHLMANN, H (1967) Experience Rating and Credibility *ASTIN Bulletin* **4**, 199-207

CUMMINS, J D , DIONNE, G , MC DONALD, J B and PRITCHETT, B M (1990) Application of the GB2 Distribution in Modelling Insurance Loss Processes *Insurance Mathematics and Economics* **9**, 257-272

DIONNE, G and VANASSE, C (1989) A Generalization of Automobile Insurance Rating Models The Negative Binomial Distribution with a 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

LEMAIRE, J (1985) *Automobile Insurance Actuarial Models* Huebner International Series on Risk, Insurance and Economic Security

LEMAIRE, J (1995) *Bonus-Malus Systems in Automobile Insurance* Huebner International Series on Risk, Insurance and Economic Security

PINQUET, J , ROBERT, J C , PESTRE, G and MONTOCCHIO, L (1992) Tarification a Priori et a Posteriori des Risques en Assurance Automobile *Mémoire au Centre d'Etudes Actuarielles*

PINQUET, J (1996a) Allowance for Costs of Claims in Bonus-Malus Systems *Proceedings of the ASTIN colloquium, Copenhague 1996*

PINQUET, J (1996b) Hétérogénéité Inexpliquée *Document de travail THEMA n° 11*

RAO, C R (1948) Large Sample Tests of Statistical Hypothesis Concerning Several Parameters with Applications to Problems of Estimation *Proceedings of the Cambridge Philosophical Society* **44**, 50-57

RENSHAW, E A (1994) Modelling the Claims Process in the Presence of Covariates *ASTIN Bulletin* **24**, 265-285

SILVEY, S D , (1959) The Lagrange Multiplier Test *Annals of Mathematical Statistics* **30**, 389-407

JEAN PINQUET
Université de Paris X, U F R de Sciences Economiques
200, Avenue de la République 92001 NANTERRE CEDEX
Phone 33 1 49 81 72 45, Fax: 33 1 40 97 71 42,
E-mail. pinquet@u-paris10 fr

ALLOWANCE FOR THE AGE OF CLAIMS IN BONUS-MALUS SYSTEMS*

BY

JEAN PINQUET[1], MONTSERRAT GUILLÉN[2], CATALINA BOLANCÉ[2]

ABSTRACT

The purpose of the paper is to use the age of claims in the prediction of risks. A dynamic random effects model on longitudinal count data is presented, and estimated on the portfolio of a major Spanish insurance company. The estimated autocorrelation coefficients of stationary random effects are decreasing. A consequence is that the predictive ability of a claim decreases with the lag between the period of risk prediction and the period of occurrence. There is a wide gap between the long term properties of actuarial and real-world experience rating schemes. This gap can be partly filled if the age of claims is taken into account in the actuarial model.

KEYWORDS

Time-independent and dynamic random effects. Autocorrelation function for stationary random effects.

1. INTRODUCTION

The purpose of the paper is to use the age of claims in the prediction of risks. This issue has already been addressed in the actuarial literature. Solutions are obtained from credibility models which can be updated (Gerber, Jones (1975)), and from credibility estimators with geometric weights (Sundt (1988)).

The rating models presented in this paper are obtained after statistical inference on longitudinal count data. Let us first clarify the reasons which lead to question the assumption of time-independence for the random effects.

* Pinquet acknowledges financial support from the Fédération Française des Sociétés d'Assurance. Guillén and Bolancé thank the Spanish CICYT grant SEC99-0693. We thank a referee for his comments.

[1] (Corresponding author) University Paris X.

[2] University of Barcelona.

Hidden features of risk distributions vary with time, as do rating factors. The random effects added in an a priori rating model on longitudinal data should then be dynamic. Variations of rating factors between two dates should increase with the related lag, and the same result is expected for hidden features in risk distributions. Hence the predictive ability of a claim should decrease with the lag between the period of risk prediction and the period of occurrence. A stationary process for the random effects will relate the predictive power of claims to this lag. If the preceding intuition is verified, the estimated autocorrelation function of the random effects should be decreasing, a point already mentioned by Sundt (1988).

In Section 2, we present different Poisson models with random effects.

The variance of a time-independent random effect can be estimated from disaggregated data or from numbers of claims and frequency premiums which are summed across the periods. If the estimated variance obtained from disaggregated data is greater than the second one, the estimation of distributions for dynamic random effects can be considered. This condition is verified on our data set, which is drawn from the portfolio of a major Spanish insurance company.

An unconstrained autocorrelation function for dynamic random effects is then estimated from a Poisson model with regression components. For each lag, the corresponding autocorrelation is estimated from paired off products of lagged number-residuals and frequency-premiums. The autocorrelation function obtained in the empirical study decreases, but more slowly than a geometric one.

Optimal bonus-malus systems designed from a linear credibility approach are presented in Section 3 from the random effects models developed in Section 2.

In Section 4, we assess the consequences of a varying autocorrelation specification for the random effects on the dynamic of bonus-malus coefficients. An optimal bonus-malus system (later referred to as BMS) designed from a model with dynamic random effects and a decreasing autocorrelation function will behave in the following way. The no-claim discounts induced by a claimless year are lower than those obtained from the usual credibility model for a policyholder with a faultless history, but they are higher if claims were reported recently. The explanation is the same in both cases. The credibility granted to a given period of the past decreases rapidly as time goes by, due to the increase of risk exposure but mostly to the decrease in the autocorrelation coefficients.

Actuarial and real-world BMS differ with respect to the dynamic of bonus-malus coefficients. For example, the duration of a claimless history needed to offset the malus induced by a claim at fault is longer for actuarial than for real-life BMS. The aforementioned properties of the BMS with dynamic random effects allow us to reduce this difference.

Increases in premiums induced by claims from the different BMS are not very different in the empirical study. On the whole, an optimal BMS derived from a Poisson model with dynamic random effects seems acceptable to policyholders, if the estimated correlogram is decreasing. Besides, it would

entail strong incentives to careful driving for the drivers who reported a claim recently.

Another difference is that an optimal BMS designed from rating models with dynamic random effects reaches its limit faster than the usual ones. Besides, total credibility does not converge towards one, which entails a lower dispersion for the bonus-malus coefficients. In real-life situations, the dispersion of bonus-malus coefficients is much lower than what is obtained from actuarial models. Allowing for the age of claims in an optimal BMS reduces the dispersion of the bonus-malus coefficients.

Finally, the applicability of the results obtained in the paper is briefly discussed in Section 5. A possibly useful result for practitioners is the following. An optimal BMS estimated from short histories and applied to a longer duration will overestimate the individual credibilities. This result occurs if the autocorrelation function of the random effects decreases with the lag. Hence, attention should be paid to the link between the length of the histories used in risk assessment and the duration of application of the BMS.

2. Moment-based estimators for longitudinal count data

2.1. Two estimators for the variance of a time-independent random effect

Let us consider a portfolio composed of p policyholders. If we have an unbalanced panel data set, the policyholder i is observed during T_i periods, with $1 \leq T_i \leq T_{\max}$. If $n_{i,t}$ is the number of claims reported by the policyholder i in period t, the distributions retained for the frequency risk model are mixtures of Poisson distributions. Their likelihood is equal to

$$L(n_{i,t})_{1\leq t\leq T_i} = E\left[\prod_{t=1}^{T_i} P_{\lambda_{i,t}U_i}(n_{i,t})\right], \text{ with } P_\lambda(n) = \exp(-\lambda)\frac{\lambda^n}{n!}.$$

The expectation is taken with respect to the random effect U_i, and the coefficients $\lambda_{i,t}$ depend on regression components (represented by a line-vector $x_{i,t}$) and on the duration of the period $d_{i,t}$. We write

$$\lambda_{i,t} = d_{i,t}\exp(x_{i,t}a);\ a \in \mathbb{R}^k,$$

where a is a column-vector of parameters and where k is the number of regression components. The random effects $(U_i)_{i=1,\ldots p}$ are assumed i.i.d., and the two first moments are

$$E(U_i) = 1;\ V(U_i) = \sigma_U^2.$$

Within a semiparametric approach, we do not completely specify the distributions of the random effects. Estimators and predictors are obtained from second order moments of the random effects and from the maximum likelihood estimation of the Poisson model with regression components and

without random effects. Let us denote the cumulated number of claims and frequency risk for the policyholder i as $n_i = \sum_{t=1}^{T_i} n_{i,t}$ and $\lambda_i = \sum_{t=1}^{T_i} \lambda_{i,t}$. Two consistent estimators of σ_U^2 can be retained, which are

$$\widehat{\sigma_U^2}^1 = \frac{\sum_{i,t}\left[(n_{i,t} - \widehat{\lambda}_{i,t})^2 - n_{i,t}\right]}{\sum_{i,t} \widehat{\lambda}_{i,t}^2}; \quad \widehat{\sigma_U^2}^2 = \frac{\sum_i (n_i - \widehat{\lambda}_i)^2 - n_i}{\sum_i \widehat{\lambda}_i^2}, \tag{1}$$

where $\widehat{\lambda}_{i,t}$ and $\widehat{\lambda}_i$ are the frequency-premiums computed from likelihood maximization in the Poisson model without random effects. These two estimators are unbiased, and the second estimator should be preferred because its variance is lower. The intuition is that this estimator uses more information. Besides, $\widehat{\sigma_U^2}^1$ reflects a short term predictive ability of the claims. It should be greater than an estimator based on longer durations if the predictive ability of claims decreases with their age. The inequality

$$0 < \widehat{\sigma_U^2}^2 < \widehat{\sigma_U^2}^1$$

is fulfilled on our data set (see Section 4). It is a necessary condition for the estimation of simple dynamic random effects specifications (see Pinquet, Guillén, Bolancé (2000)).

2.2. A Poisson model with dynamic random effects

We now replace the time-independent random effect U_i by a dynamic random effect $U_{i,t}$. The likelihood of the random effects model is equal to

$$L(n_{i,t})_{1 \le t \le T_i} = E\left[\prod_{t=1}^{T_i} P_{\lambda_{i,t} U_{i,t}}(n_{i,t})\right].$$

As mentioned in the introduction, the time-dependence assumption for the random effects is natural. Two assumptions are retained on the random effects, which are the following.

- The distribution of $\underset{1 \le t \le T_i}{vec}(U_{i,t})$ depends only on T_i.
- If the distribution of $\underset{1 \le t \le T_i}{vec}(U_{i,t})$ is that of $\underset{1 \le t \le T_i}{vec}(U_t)$ the distribution of $\underset{1 \le t \le T_{\max}}{vec}(U_t)$ is supposed to be stationary. This invariance assumption with respect to time translations means that the predictive ability of a claim will depend on the lag between the period of risk prediction and the period of occurrence. We suppose that the squared random effects are integrable.

The covariances and autocorrelation coefficients for the random effects are denoted as

$$Cov(U_t, U_{t-h}) = \rho_U(h)\sigma_U^2;\ 0 \le h < t \le T_{\max},$$

with: $-1 \le \rho_U(h) \le 1; \rho_U(0) = 1$.

From the moment equations derived in the random effects model

$$\begin{aligned}
&E\left[\sum_{i,t}(N_{i,t}-\lambda_{i,t})^t x_{i,t}\right] = 0\left(0 \in \mathbb{R}^k\right);\\
&E\left[\sum_{i,t}(N_{i,t}-\lambda_{i,t})^2 - N_{i,t} - \lambda_{i,t}^2\sigma_U^2\right] = 0;\\
&E\left[\sum_{i|T_i>h}\ \sum_{T_i \ge t > h} (N_{i,t}-\lambda_{i,t})(N_{i,t-h}-\lambda_{i,t-h}) - \lambda_{i,t}\lambda_{i,t-h}\sigma_U^2\rho_U(h)\right] = 0,\\
&\qquad\qquad (0<h<T_{\max})
\end{aligned} \tag{2}$$

we obtain consistent moment-based estimators for a, σ_U^2 and $(\rho_U(h))_{0<h<T_{\max}}$.

The empirical counterpart of the moment equation related to a leads to

$$\sum_{i,t}(n_{i,t}-\hat{\lambda}_{i,t})^t x_{i,t} = 0, \hat{\lambda}_{i,t} = d_{i,t}\exp(x_{i,t}\hat{a}). \tag{3}$$

Hence, the maximum likelihood estimator of a in the Poisson model without random effects (i.e. the a priori rating model) is a consistent estimator of a in the model with random effects. From the second moment equation, a consistent estimator of σ_U^2 is $\widehat{\sigma_U^2}^1$ (see (1)). Finally, the estimated correlogram of U is obtained from

$$\widehat{\sigma_U^2}^1 \rho_U(h) = \frac{\sum_{i|T_i>h}\ \sum_{T_i \ge t > h} (n_{i,t}-\hat{\lambda}_{i,t})(n_{i,t-h}-\hat{\lambda}_{i,t-h})}{\sum_{i|T_i>h}\ \sum_{T_i \ge t > h} \hat{\lambda}_{i,t}\hat{\lambda}_{i,t-h}} \tag{4}$$

for $0 < h < T_{\max}$. All these estimators are consistent and asymptotically normal. They are given in Zeger (1988), along with modified estimators which use weights related to overdispersion and autocorrelation in the regression in order to reduce the asymptotic variance.

3. Linear credibility predictors derived from the preceding models

Let $(n_t)_{1 \le t \le T \le T_{\max}}$ be the history of claims recorded on an insurance contract (we suppress the individual index in order to simplify the notations). A linear credibility predictor (Bühlmann (1967)) for period $T+1$ is obtained from a

regression derived in the model with random effects. The predictor is equal to $\widehat{a}+\sum_{t=1}^{T}\widehat{b}_t n_t$, with

$$(\widehat{a},\widehat{b}_1,\ldots,\widehat{b}_T)=\arg\min_{a,(b_t)_{t=1,\ldots,T}}\widehat{E}\left[\left(U_{T+1}-a-\sum_{t=1}^{T}b_t N_t\right)^2\right],$$

where the expectation is estimated in the random effects model.

A linear predictor of the type $\widehat{a}+\widehat{b}\left(\sum_{t=1}^{T}n_t\right)$ is obtained from an expected value principle with the negative binomial model (see Lemaire (1985), and Dionne, Vanasse (1989)). The purpose of the paper is to obtain time-dependent coefficients in the linear combination. The intuition is that b_t should decrease with the age of the period t.

Since $E(U_{T+1})=1$, we have $\widehat{a}+\sum_{t=1}^{T}\widehat{b}_t\widehat{E}(N_t)=1$. Now $\widehat{\lambda}_t$, the frequency premium derived from likelihood maximization in the a priori rating model, converges towards the frequency risk $E(N_t)$ computed in the model with random effects (see the comments following equation (3)). Then we have

$$\widehat{a}+\sum_{t=1}^{T}\widehat{b}_t n_t=1+\sum_{t=1}^{T}\widehat{b}_t(n_t-\widehat{\lambda}_t),\text{ with}$$

$$(\widehat{b}_t)_{t=1,\ldots,T}=\arg\min_{(b_t)_{t=1,\ldots,T}}\widehat{V}\left[\left(U_{T+1}-\sum_{t=1}^{T}b_t N_t\right)^2\right]=[\widehat{V}(N)]^{-1}\widehat{Cov}(N,U_{T+1}).$$

We write $N=\underset{1\le t\le T}{vec}(N_t)$. From the consistent estimators given in Section 2.2, the estimators of the individual moments of interest are

$$\widehat{V}(N_t)=\widehat{\lambda}_t+\widehat{\sigma_U^2}^{1}\widehat{\lambda}_t^2;\ \widehat{Cov}(N_t,N_{t'})=\widehat{\lambda}_t\widehat{\lambda}_{t'},\widehat{\sigma_U^2}^{1}\widehat{\rho}_U(|t-t'|)\ (t\neq t');$$

$$\widehat{Cov}(N_t,U_{T+1})=\widehat{\lambda}_t\widehat{\sigma_U^2}^{1}\rho_U(T+1-t).$$

The bonus-malus coefficient can be written as

$$\left(1-\sum_{t=1}^{T}cred_t\right)+\sum_{t=1}^{T}cred_t\frac{n_t}{\widehat{\lambda}_t}$$

where $(cred_t=\widehat{b}_t\widehat{\lambda}_t)_{t=1,\ldots,T}$ are the credibility coefficients, which are the solutions of the linear system with $t=1,\ldots,T$ equations

$$\left(1+\widehat{\lambda}_t\widehat{\sigma_U^2}^{1}\right)cred_t+\widehat{\lambda}_t\sum_{t'\neq t}\widehat{\sigma_U^2}^{1}\widehat{\rho}_U(|t-t'|)\,cred_{t'}=\widehat{\lambda}_t\widehat{\sigma_U^2}^{1}\widehat{\rho}_U(T+1-t).\qquad(5)$$

This linear credibility system can be used with the unconstrained correlogram estimated in Section 2.2.

Properties of the credibility coefficients derived from equation (5) are not simple to obtain in a general setting. If the frequency premiums are negligible with respect to one, we infer from this equation

$$cred_t \sim \widehat{\lambda}_t \widehat{\sigma}_U^{2\,1} \widehat{\rho}_U(T+1-t)$$

A sequence of credibility coefficients should have the same shape as that of a correlogram with the same length and a reversed index.

Total credibility does not converge to one when T goes to infinity if the autocorrelation function decreases rapidly, and the limit can be very inferior to one. Let us consider for example a Gaussian $AR(1)$ process for the additive random effects $W_t = \log(U_t)$. With our data base, we obtain $\widehat{\rho}w(h) = 0.79^h$ (see Pinquet, Guillén, Bolancé (2000) for details related to the estimation). The total credibility for an average risk ($\widehat{\lambda}_t = 0.09\ \forall t$) converges towards 0.214 when T converges towards infinity. This limit is obtained in round figures after twenty years. It corresponds to the maximum bonus applied to the a priori frequency premium of a policyholder with a claimless history.

Simple updating formulas do not seem to be available for the credibility coefficients. Gerber and Jones (1975) prove that linear updating formulas exist under conditions which differ from the stationarity assumption retained in this paper for the random effects.

4. Empirical results

4.1. The data set

The working sample represents ten per cent of the portfolio of a major Spanish insurance company. We selected only policies covering cars for private use. The durations of individual histories range from one to seven years, hence $T_{max} = 7$ with the notations of the paper. Policyholders were observed between 1991 and 1997, and indicators of the calendar years are part of the regression components in order to allow for a trend in the past (see Besson, Partrat (1992) for optimal BMS with a trend). The other rating factors retained in the regression are the gender, the geographical area, the age of the driving licence, the seniority and age of the policyholder, the coverage level and the power of the vehicle.

In order to have similar rates of arrival and attrition in the working sample and in the portfolio, we selected the policyholders in the following way. Ten per cent of the policyholders present in 1991 were selected at random, and kept in the working sample as long as possible. Ten per cent of the newcomers in 1992 were included in the working sample, and so on. The size of the working sample increases from 120000 in 1991 to 200000 in 1997 (in round figures). The attrition rate varies between 8.5% and 10%. The working sample is an unbalanced panel data set which is composed of 269388 policyholders and of 1172701 periods. The average frequency of claims at fault per year is equal to 0.09. All the period durations are equal to one year, which means

that the characteristics of the policyholders are known only at each anniversary date. This inaccuracy in the observation of the regression components is of no consequence in our opinion.

4.2. Estimators for the correlogram of the random effects

The two consistent estimators quoted in Section 2.1 for the variance of a time-independent random effect are respectively

$$\widehat{\sigma_U^2}^1 = \frac{\sum_{i,t}(n_{i,t}-\widehat{\lambda}_{i,t})^2 - n_{i,t}}{\sum_{i,t}\widehat{\lambda}_{i,t}^2} = \frac{118554.78-105655}{10167.12} = 1.269.$$

$$\widehat{\sigma_U^2}^2 = \frac{\sum_{i}(n_{i}-\widehat{\lambda}_{i})^2 - n_{i}}{\sum_{i}\widehat{\lambda}_{i}^2} = \frac{144879.33-105655}{50359.14} = 0.779. \qquad (6)$$

As expected in Section 2.1, we have $\widehat{\sigma_U^2}^2 < \widehat{\sigma_U^2}^1$.

The correlogram of $(U_t)_{1\le t\le 6}$ is estimated from equations (1) and (4). We obtain

TABLE 1

AUTOCORRELATION COEFFICIENTS

h (lag)	1	2	3	4	5	6
$\widehat{\rho}_U(h)$	0.632	0.485	0.462	0.436	0.360	0.348

The difference between $\widehat{\rho}_U(1)$ and $\rho_U(0)=1$ reflects the loss of predictive ability related to a claim after one year. Owing to the shape of the correlogram, the predictive ability of claims decreases with their age. The consequences are assessed in the following section.

The moment-based estimators retained in this paper are unconstrained. For instance, the estimated autocorrelation coefficients are not bound to belong to [–1, 1]. This constraint is fulfilled by the preceding estimators, and a multivariate log-normal distribution for the $(U_t)_{1\le t\le 6}$ can be matched with the values of Table 1 (see Pinquet, Guillén, Bolancé (2000) for more details and further discussion in case of misspecification).

4.3. Experience rating from the different models

In the following tables, credibility coefficients for the different periods are computed for an insurance contract with an average frequency premium, which is equal to 0.09 per year. Within a linear credibility approach, we use the Poisson models with random effects presented in Section 2. The next table

provides credibility coefficients computed from time-independent and dynamic random effects. The coefficients are computed for histories ranging from one to six years. We used the usual credibility formula for time-independent random effects, and the linear credibility system given in Section 3 for the other model. We estimated the variance of the time-independent random effect from the number of claims and frequency-premiums summed across the periods. Hence, we retained $\widehat{\sigma_U^2}^2 = 0.779$ instead of $\widehat{\sigma_U^2}^1 = 1.269$ (see equation (6)).

Remember that a credibility coefficient is a bonus if no claim is reported.

TABLE 2

CREDIBILITY COEFFICIENTS FOR AN AVERAGE RISK (PERCENTAGE)
DYNAMIC VS. TIME-INDEPENDENT RANDOM EFFECTS

Duration of histories	Credibility per year (%) dynamic random effects						Total credibility (%)	
	t=1	t=2	t=3	t=4	t=5	t=6	dynamic random effects	time-independent random effects
1 year	6.47	0	0	0	0	0	6.47	6.55
2 years	4.57	6.17	0	0	0	0	10.74	12.29
3 years	4.15	4.32	5.98	0	0	0	14.45	17.37
4 years	3.74	3.94	4.14	5.83	0	0	17.65	21.89
5 years	2.83	3.57	3.82	4.03	5.72	0	19.97	25.95
6 years	2.66	2.68	3.46	3.71	3.94	5.65	22.10	29.60

For a given duration of the individual history, the credibility coefficients of the last table decrease with the lag between the prediction period and the current period, as do the autocorrelation coefficients. For example, the credibility given in Table 2 for the last year of a six years history outweighs the credibility of the two first years. Besides, total credibility (and hence the bonus applied to a claimless history) is lower if dynamic random effects are used in the rating model.

Let us now perform an impulse-response analysis of the evolution of the bonus-malus coefficient if one claim is reported during the first year, and none during the years that follow. We compare the two BMS for a car with the average frequency premium. Bonus-malus coefficients are expressed as percentages.

TABLE 3

IMPULSE-RESPONSE ANALYSIS OF BONUS-COEFFICIENTS AFTER ONE CLAIM

Years	1	2	3	4	5	6
time-independent random effects	166.2	156	147	139	131.7	125.2
dynamic random effects	165.5	140	131.7	123.8	111.4	107.5

The no-claim discounts are higher for the rating model with dynamic random effects. For instance, the discounts from period 1 to period 2 are equal to 6 and 15 per cent in the two models. This important difference is due to the fact that claimless periods increase risk exposure but also the age of claims reported in the past. This property of the rating model with dynamic random effects can be related to some clauses found in real-life BMS, which provide important discounts for bad drivers with recent good behaviour. In France for instance, a driver with a bonus-malus coefficient greater than the coefficient applied to beginners (less than three percent of the drivers are concerned) is rated according to this coefficient after two consecutive claimless years. This feature of real-life or optimal BMS entails strong incentives to drive carefully for policyholders with a bad accident record. Notice that economic analysis suggests that optimal insurance contracts with moral hazard should penalize recent claims more than older ones (Henriet, Rochet (1986)).

As shown in Table 3, the duration of a claimless history needed to offset the malus induced by a claim at fault is longer for an actuarial BMS designed from time-independent random effects than for the other BMS. If the frequency premium per year is equal to 0.09, ten claimless years are needed to offset the malus with the usual BMS. On the other hand, five claimless years are almost enough to obtain the same result if the other BMS is used. This duration is closer to those derived from real-life BMS, which range from three to five years in most cases (see Lemaire (1995) for a thorough description of compulsory BMS).

Let us compare total credibility for longer durations. We need to extend the correlogram for higher values of the lag. A possible extension is derived from the Yule-Walker equations applied to the additive random effects $W_t = \log(U_t)$. If we assume that the past is summarized by the last six years for the random effects process, we obtain an $AR(6)$ specification for the additive random effects. In that case, a parametric specification is needed in order to link the second order moments of the W_t with the corresponding moments of U_t. A multivariate Gaussian distribution for the additive random effects can be considered (see Pinquet, Guillén, Bolancé (2000) for more details). With an average risk and from the correlogram given in Table 1, we obtain

TABLE 4

LONG-TERM BEHAVIOUR OF TOTAL CREDIBILITY (PERCENTAGE)

Duration of histories	time-independent random effects	dynamic random effects
10 years	41.2	27.7
20 years	58.4	32.6
40 years	73.7	34.1

The result obtained in the last column is striking. After almost a full life as a car driver, a policyholder with a claimless history obtains a frequency-bonus of only 34 per cent, which is about five times the bonus after the first year.

In the last table, we compare the dispersion of the bonus-malus coefficients derived from the two models. We derive the standard deviation of bonus-malus coefficients for different durations of the histories. Computations are performed in the two different random effects models for an individual with an average frequency risk. The autocorrelation function of the dynamic random effects is extended as indicated before Table 4. The variance of the bonus-malus coefficients are those of the linear regression which defines the linear credibility predictor, that is to say

$$^t\widehat{Cov}(N, U_{T+1})[\widehat{V}(N)]^{-1}\widehat{Cov}(N, U_{T+1})$$

with the notations of Section 3. For different durations of the history, we obtain

TABLE 5

STANDARD DEVIATION OF BONUS-MALUS COEFFICIENTS

duration of the history (years)	1	5	10	20	40
time-independent random effects	0.226	0.450	0.567	0.674	0.758
dynamic random effects	0.228	0.355	0.389	0.398	0.399

In real-life situations, the dispersion of bonus-malus coefficients is much lower than what is obtained from actuarial models. For instance, the coefficient of variation of the bonus-malus coefficients in Belgium was equal to 0.154 in 1992 (see Lemaire (1995) for the distribution of bonus-malus coefficients at that time). The standard deviation of bonus-malus coefficients in an actuarial BMS is also the coefficient of variation, due to the fairness property of these systems. The differences between actuarial and real-life BMS are lower if the age of claims is allowed for in the statistical model.

5. Concluding remarks

The main features of an optimal BMS derived from stationary random effects with a decreasing correlogram seem acceptable to policyholders. The no-claim discounts are lower for claimless drivers than those derived from the usual optimal BMS. On the other hand, they can be much higher for policyholders who reported claims recently. Such systems would entail strong incentives for these drivers to drive carefully.

A useful result for the application of actuarial models is the following. If the autocorrelation between stationary random effects decreases with the lag, the variance of a time-independent random effect (estimated from aggregated numbers and frequency-premiums) will decrease with the average duration of the histories used in the estimation. For instance, the variance estimated from the observations of the first period is equal to 1.09, whereas the variance obtained from the full histories is equal to $\widehat{\sigma_U^2}^2 = 0.78$ (see Section 4.2). In this

framework, an optimal BMS estimated from short histories and applied to a longer duration will overestimate the individual credibilities. This result provides a supplementary reason to use the whole history of the policyholders in the statistical analysis.

REFERENCES

1. BESSON, J.L. and PARTRAT, C. (1992) Trend et systèmes de bonus-malus. *ASTIN Bulletin* **22**, 11-32.
2. BÜHLMANN, H. (1967) Experience rating and credibility. *ASTIN Bulletin* **4**, 199-207.
3. DIONNE, G. and VANASSE, C. (1989) A generalization of automobile insurance rating models: the negative binomial distribution with a regression component. *ASTIN Bulletin* **19**, 199-212.
4. GERBER, H. and JONES, D. (1975) Credibility formulas of the updating type. *Transactions of the Society of Actuaries* **27**, 31-52.
5. HENRIET, D. and ROCHET, J.C. (1986) La logique des systèmes bonus-malus en assurance automobile. *Annales d'Economie et de Statistiques*, 133-152.
6. LEMAIRE, J. (1985) *Automobile Insurance: Actuarial Models*. Kluwer Academic Publishers.
7. LEMAIRE, J. (1995) *Bonus-Malus Systems in Automobile Insurance*. Kluwer Academic Publishers.
8. PINQUET, J., GUILLEN, M. and BOLANCE, C. (2000) Long-range contagion in automobile insurance data: Estimation and implications for experience rating. Working paper 2000-43, http://thema.u-paris10.fr.
9. SUNDT, B. (1988) Credibility estimators with geometric weights. *Insurance: Mathematics and Economics* **7**, 113-122.
10. ZEGER, L.S. (1988) A regression model for time series of counts. *Biometrika* **74**, 721-729.

JEAN PINQUET
U.F.R. de Sciences Economiques
Université de Paris X
200, avenue de la République
92001 Nanterre Cedex
France
e-mail: pinquet@u-paris10.fr

MONTSERRAT GUILLÉN
Departament d'Econometria, Estadística i Economia Espanyola
Universitat de Barcelona
Diagonal, 690
08034 Barcelona
Spain
e-mail: guillen@eco.ub.es

CATALINA BOLANCÉ
Departament d'Econometria, Estadística i Economia Espanyola
Universitat de Barcelona
Diagonal, 690
08034 Barcelona
Spain
e-mail: bolance@eco.ub.es

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Insurance: Mathematics and Economics 33 (2003) 273–282
INSURANCE MATHEMATICS & ECONOMICS
www.elsevier.com/locate/econbase

Time-varying credibility for frequency risk models: estimation and tests for autoregressive specifications on the random effects

Catalina Bolancé [a], Montserrat Guillén [a], Jean Pinquet [b,*]

[a] *Department of Economics, University of Barcelona, Diagonal 690 08034 Barcelona, Spain*
[b] *Department of Economics, U.F.R. de Sciences Economiques, Université de Paris X, 200 Avenue de la République 92001 Nanterre Cedex, France*

Received 1 September 2002; received in revised form 1 March 2003; accepted 19 March 2003

Abstract

This paper estimates and tests autoregressive specifications for dynamic random effects in a frequency risk model. Linear credibility predictors are derived from the estimators. Examples are provided from the automobile portfolio of a Spanish insurance company.

JEL classification: C11; C13; C14

Subj. Class.: IM 31; IB 40

Keywords: Time-varying random effects; Autocorrelation function for stationary random effects; Generalized estimating equations

1. Introduction

Real-world experience rating schemes in property-liability insurance mostly depend on numbers of events, which are usually claims at fault. Only in a few publications (Gerber, 1975; Sundt, 1988; Pinquet et al., 2001) do frequency risk models take into account the age of events. These contributions use the intuition that the predictive ability for a period of the policyholder's history should decrease with the age. If a stationary specification is retained for time-varying random effects in a Poisson model, the estimated autocorrelation coefficients should be decreasing. This shape is indeed obtained from our data.

As durations of histories in insurance portfolios do not usually exceed a few years (7 at most in our working sample), only a short correlogram can be estimated from the data. Long-term computations of derived credibility models require correlation coefficients for higher lag values. The purpose of this paper is to provide extensions of the correlogram which are based on autoregressive specifications. Statistical inference is performed on the order p of an $AR(p)$ process applied to additive random effects. Estimators and tests are given in Section 2. Section 3 details advantages and possible limits of the linear credibility approach in a context where an intricate correlation structure is allowed for the random effects.

* Corresponding author. Tel.: +33-1-4097-4758; fax: +33-1-4097-5973.
E-mail address: pinquet@u-paris10.fr (J. Pinquet).

0167-6687/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0167-6687(03)00139-2

Empirical results are shown in Section 4. First, we investigate the link between the order of the autoregressive process and the credibility coefficients. Then, we derive bonus-malus coefficients from duration distributions on the policyholder–company relationship. This type of computation is motivated by the fact that real-life rating structures in non-life insurance implicitly take into account an expected loyalty of the policyholder to the company.[1] The risk period in actuarial models is usually the year to come, and ends with the renegotiation date of the contract. A time-varying credibility model used together with duration distributions on the policyholder–company relationship provides credibility coefficients which decrease with expected loyalty.

Lastly, concluding remarks are given in Section 5.

2. Poisson models with dynamic and stationary random effects

Let us consider a portfolio composed of p policyholders. If we have an unbalanced panel data set, the policyholder i is observed for T_i periods with $1 \leq T_i \leq T_{\max}$. If $n_{i,t}$ is the number of claims reported by the policyholder i in period t, the distributions retained for the frequency risk model are mixtures of Poisson distributions. Hidden features of risk distributions vary with time, as do rating factors. The random effects added to an a priori rating model with longitudinal data should then be time-dependent.

Let $U_{i,t}$ be the random effect related to the distribution of $N_{i,t}$. Conditionally on $U_{i,t} = u$, $N_{i,t}$ is supposed to follow a Poisson distribution with an expectation equal to $\lambda_{i,t}u$. The coefficients $\lambda_{i,t}$ depend on regression components (represented by a line-vector $x_{i,t}$) and on the duration of the period $d_{i,t}$. We write $\lambda_{i,t} = d_{i,t}\exp(x_{i,t}\beta)$; $\beta \in R^k$, where β is a column-vector of parameters and where k is the number of regression components. Besides, the independence of the $(N_{i,t})_{t=1,\ldots,T_i}$ is assumed if we condition by the random effects. In the model with random effects, the likelihood of the individual history of policyholder i within a parametric specification is equal to

$$L[(n_{i,t})_{1\leq t\leq T_i}] = E\left[\prod_{t=1}^{T_i} P_{\lambda_{i,t}U_{i,t}}(n_{i,t})\right], \quad P_\lambda(n) = \exp(-\lambda)\frac{\lambda^n}{n!}. \tag{1}$$

Expectation is taken with respect to the random effects.

The likelihood of a Poisson model with a complex correlation structure for random effects is not tractable, as is the case for linear models (see Frees et al. (1999) for a survey of linear models on panel data with applications to credibility predictors). However, second-order moments remain tractable and we will follow a semiparametric approach.

We assume equidistribution and stationarity at the second-order in the following way. The first- and second-order moments of $\mathrm{vec}_{1\leq t\leq T_i}(U_{i,t})$ are those of $\mathrm{vec}_{1\leq t\leq T_i}(U_t)$, and the moments of $\mathrm{vec}_{1\leq t\leq T_{\max}}(U_t)$ are supposed to be stationary at the second-order. This invariance assumption with respect to time translations means that the predictive ability of a claim will depend on the lag between the period of risk prediction and the period of occurrence, but not on calendar time. We suppose that the squared random effects are integrable, and that $E(U_t) = 1 \ \forall t = 1, \ldots, T_{\max}$. The second-order moments of the random effects are denoted as

$$\mathrm{Cov}(U_t, U_{t-h}) = \sigma_U^2\rho_U(h), \quad 0 \leq h < t \leq T_{\max} \tag{2}$$

with $-1 \leq \rho_U(h) \leq 1$, $\rho_U(0) = 1$.

The random effects model is parameterized by $\theta \in \mathbb{R}^{k+T_{\max}}$ with

$$\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad \alpha = \underset{0\leq h<T_{\max}}{\mathrm{vec}}(\sigma_U^2\rho_U(h)). \tag{3}$$

[1] For instance, insurance companies undercharge new cars and overcharge older ones. Since automobile risk decreases with the age of the car, commercial premiums implicitly use a more remote horizon than the year to come.

From the identity

$$E_\theta[M_{\theta,x_i}(N_i)] = 0, \quad 0 \in \mathbb{R}^{k+T_{\max}} \tag{4}$$

with

$$M_{\theta,x_i}(N_i) = \begin{pmatrix} \sum_{1\le t\le T_i}(N_{i,t} - \lambda_{i,t})\,{}^t x_{i,t} \\ \sum_{1\le t\le T_i}[(N_{i,t} - \lambda_{i,t})^2 - N_{i,t} - \lambda_{i,t}^2\sigma_U^2] \\ \mathrm{vec}_{1\le h<T_{\max}}\left(\sum_{h<t\le T_i}[(N_{i,t} - \lambda_{i,t})(N_{i,t-h} - \lambda_{i,t-h}) - \lambda_{i,t}\lambda_{i,t-h}\sigma_U^2\rho_U(h)]\right) \end{pmatrix}, \tag{5}$$

we obtain a moment-based estimator $\hat\theta^0$ which nullifies the empirical counterpart of Eq. (4). Hence

$$(\overline{M_\theta})_{\theta=\hat\theta^0} = 0, \quad \overline{M_\theta} = \frac{1}{p}\sum_{i=1}^{p} M_{\theta,x_i}(N_i). \tag{6}$$

Let us interpret the components of $\hat\theta^0$. The estimator $\hat\beta^0$ is the m.l.e. in the Poisson model without random effects. Then $\hat\alpha^0$ is derived from frequency premiums and number-residuals computed from $\hat\beta^0$ and the regression components. A frequency premium is the estimated expectation of a number of claims in the Poisson model, and a number-residual is the difference between a number of claims and the related frequency premium. The unconstrained estimator of the variance is positive if and only if there is overdispersion of number-residuals (i.e. the variance of the number-residuals is greater than the mean of the claim numbers). The autocovariance and autocorrelation coefficients are estimated for the lags $h = 1, \ldots, T_{\max} - 1$, the number of which depends on the maximal length of the individual histories. These coefficients are estimated from paired-off products of lagged number-residuals and frequency premiums.

Some points concerning these estimators are worth developing.

- The estimators for second-order moments of the random effects are unconstrained, which means that they are not bound to belong to the parameter space. For instance, the estimated variance is negative if there is underdispersion of number-residuals. Likewise, the estimated autocorrelation coefficients are not bound to belong to the interval $[-1, +1]$. Even if the preceding conditions are fulfilled, the estimated variance–covariance matrix of the random effects might not be positive semidefinite. In this case, the distribution family must be simplified in order to obtain estimators inside the parameter space.
- The unconstrained nature of the estimators cannot be seen as a drawback. The parameter space for a semiparametric or parametric specification of the random effects distribution admits a boundary.[2] If an unconstrained estimator is outside the parameter space, a restricted estimator derived from likelihood maximization belongs most often to the boundary. In this case, the distribution family of the random effects must be simplified, which leads to the conclusion reached with an unconstrained estimation strategy.
- Moment-based estimators can be improved in terms of asymptotic variance. The idea is to use the estimated moments of the random effects in the regression equation. This is the Generalized Estimating Equations approach, later referred to as GEE (see Liang and Zeger (1986), Zeger (1988), Brännäs and Johanson (1995) for an application to longitudinal count data).

Let us detail this estimation strategy. With the notations given above, we write

$$n_i = \underset{1\le t\le T_i}{\mathrm{vec}}(n_{i,t}), \qquad E(N_i) = \underset{1\le t\le T_i}{\mathrm{vec}}(\lambda_{i,t}) = E_i(\beta), \quad \lambda_{i,t} = d_{i,t}\exp(x_{i,t}\beta). \tag{7}$$

[2] Distributions for a time-independent random effect are parameterized by the variance, in which case the parameter space is $\mathbb{R}^+$ and the boundary is zero. Consider now dynamic random effects related to a 2-year history. The parameters of a stationary distribution family are a variance and a covariance (say, V_{11} and V_{12}). The parameter space is defined by the constraints $V_{11} \ge 0$ and $-V_{11} \le V_{12} \le V_{11}$. The boundary is composed of the two half-lines $V_{12} = \pm V_{11}$ with $V_{11} \ge 0$. For instance the random effects are time-independent if $V_{12} = V_{11}$, which entails a simplified correlation structure.

The m.l.e. of β in the Poisson model is the solution $\hat{\beta}^0$ of the equation:

$$\sum_{i,t}(n_{i,t}-\lambda_{i,t})\,{}^t x_{i,t} = \sum_i {}^t\left(\frac{\partial E_i}{\partial \beta}\right)[V_{0i}(\beta)]^{-1}(n_i - E_i(\beta)) = 0, \quad 0 \in \mathbb{R}^k, \tag{8}$$

where $V_{0i}(\beta) = V_0(N_i) = \mathrm{diag}_{1\le t\le T_i}(\lambda_{it})$ is the variances–covariances matrix of N_i in the Poisson model. On the other hand, the moment-based estimator of α is an explicit function of $\hat{\beta}^0$, and we write $\hat{\alpha}^0 = \alpha_*(\hat{\beta}^0)$.

Let us denote

$$V(N_i) = V_0(N_i) + V_0(N_i)V(U_i)V_0(N_i) = V_i(\alpha, \beta) \tag{9}$$

the variances–covariances matrix of N_i in the Poisson model with random effects. Note that the expectation of N_i in this model does not depend on α. Including this specification of the variance in the Poisson regression given in (8) and using the map α_*, we obtain an improved estimator of β which is the solution $\hat{\beta}$ to the equation:

$$\sum_i {}^t\left(\frac{\partial E_i}{\partial \beta}\right)[V_i(\alpha_*(\beta), \beta)]^{-1}(n_i - E_i(\beta)) = 0. \tag{10}$$

An estimation algorithm is detailed in Liang and Zeger (1986).

Eq. (10) expresses an orthogonality between number-residuals and the space spanned by the regression components. The metric retained in the equation (i.e. the inverse of the variance) is shown to be optimal in the aforementioned reference. Hence any estimator of this type, including the m.l.e. in the Poisson model, has a greater asymptotic variance matrix than that of $\hat{\beta}$ in the Poisson model with random effects. Finally, we obtain a new estimator $\hat{\alpha} = \alpha_*(\hat{\beta})$ for the second-order moments of the random effects.

Extending the estimated correlogram is of interest if we want to investigate the long-term properties of the derived bonus-malus systems. Autoregressive specifications for real-valued random effects provide such extensions.[3] Since the $(U_t)_{t\ge1}$ are greater than zero, an $AR(p)$ process should apply to additive random effects of the type $(W_t = \log(U_t))_{t\ge1}$. From the $T_{\max}-1$ autocorrelation coefficients estimated on the data, we can fit autoregressive processes of order p with p ranging from one to $T_{\max}-1$. A parametric specification is needed in order to link the second-order moments of the $(W_t)_{t\ge1}$ with the corresponding moments of $(U_t)_{t\ge1}$. We will use a multivariate Gaussian distribution for the $(W_t)_{t\ge1}$. Since we assumed $E(U_t) = 1$, we obtain the following link:

$$U_t = \frac{\exp(W_t)}{E[\exp(W_t)]}, \quad W_t \sim N(0, \sigma_W^2) \Rightarrow \sigma_W^2 = \log(1+\sigma_U^2). \tag{11}$$

The same identity holds for covariances. Hence we have

$$\gamma_h = \mathrm{Cov}(W_t, W_{t-h}) = \log(1 + \mathrm{Cov}(U_t, U_{t-h})) \quad \forall h,\ 0 \le h < t \le T_{\max}. \tag{12}$$

From the correlogram estimated in Section 4, we can fit a log-normal distribution for $(U_t)_{1\le t\le 6=T_{\max}-1}$. Poisson models with log-normal random effects are investigated by Aitchison and Ho (1989) and Purcaru and Denuit (2003).

A suitable value for the order p of the autoregressive process can be derived from the partial autocorrelation coefficients $(r_W(h))_{h=1,\dots,T_{\max}-1}$. For a given lag h, $r_W(h)$ is the coefficient of W_{t-h} in the probabilistic regression of W_t with respect to $W_{t-1}, \dots, W_{t-h}$.[4] If W follows an $AR(p)$, then $r_W(h) = 0$ if $h > p$. Rejecting $r_W(h) = 0$ implies that an $AR(h)$ specification for $(W_t)_{t\ge1}$ significantly outperforms an $AR(h-1)$ model on the data. A reference on stationary time series is Hamilton (1994).

The partial autocovariance coefficient $r_W(h)$ is the last component of a_h ($a_h \in \mathbb{R}^h$), the solution of the linear system: $V[\mathrm{vec}_{1\le k\le h}(W_{t-k})]a_h = \mathrm{vec}_{1\le k\le h}(\mathrm{Cov}(W_t, W_{t-k}))$. We have for instance

$$r_W(h) = 0 \Leftrightarrow g_h(\gamma) = 0, \tag{13}$$

[3] Moving average processes would not be realistic on our data, since the autocorrelation coefficients are all significantly greater than zero.

[4] This coefficient is the correlation between $W_t - W_t^*$ and $W_{t-h} - W_{t-h}^*$, where W_t^* and W_{t-h}^* are the probabilistic regressions of W_t and W_{t-h} with respect to $W_{t-1}, \dots, W_{t-h+1}$. Hence the terminology 'partial correlation coefficient', which supposes that $W_t \ne W_t^*$.

where $\gamma = \text{vec}_{0\le k<T_{\max}}(\gamma_k)$ (see (12)) is derived from the autocovariance function of W, and

$$g_1(\gamma) = \gamma_1, \qquad g_2(\gamma) = \gamma_0\gamma_2 - \gamma_1^2, \qquad g_3(\gamma) = \gamma_0^2\gamma_3 + \gamma_2^2\gamma_1 + \gamma_1^3 - \gamma_1^2\gamma_3 - 2\gamma_0\gamma_1\gamma_2. \tag{14}$$

Tests for the nullity of the partial autocorrelation coefficients are derived from the moment-based estimator $\hat{\theta}^0$ defined by Eq. (6). Its asymptotic variance is estimated by

$$\hat{V}(\hat{\theta}^0) = \frac{1}{p}({}^t\overline{J_\theta}(\overline{M_\theta\,{}^tM_\theta})^{-1}\overline{J_\theta})^{-1}_{\theta=\hat{\theta}^0}, \quad \overline{J_\theta} = \frac{\partial\overline{M_\theta}}{\partial\theta}. \tag{15}$$

Together with the asymptotic normality of the estimator, this equation means that the following convergence in distribution holds:

$$[\hat{V}(\hat{\theta}^0)]^{-1/2}(\hat{\theta}^0 - \theta) \underset{\mathcal{D}}{\to} N(0, I_{k+T_{\max}}). \tag{16}$$

A proof is given in Davidson and MacKinnon (1993). The asymptotic variance of the GEE estimator $\hat{\theta}$ could also be obtained from Eq. (15). However, the derivation of the variances and covariances involving the components of α would imply very intricate computations.

A test for the nullity of $r_W(h)$ is based on the statistic $g_h(\hat{\gamma})/\sqrt{\hat{V}(g_h(\hat{\gamma}))}$, which follows asymptotically a $N(0, 1)$ distribution. Since $\hat{\gamma} = f(\hat{\alpha}^0)$ (with f defined from (12)), the estimated variance of $g_h(\hat{\gamma})$ is equal to

$$\hat{V}(g_h(\hat{\gamma})) = J_{g_h}(\hat{\gamma})J_f(\hat{\alpha}^0)\hat{V}(\hat{\alpha}^0)\,{}^tJ_f(\hat{\alpha}^0)\,{}^tJ_{g_h}(\hat{\gamma}), \tag{17}$$

where J_f and J_{g_h} are the Jacobians of the maps f and g_h.

3. Linear credibility predictors for the frequency of claims

Let $(n_t)_{1\le t\le T}$ be the history of claims recorded on an insurance contract (we suppress the individual index in order to simplify the notations). A linear credibility predictor Bühlmann (1967) for period $T+1$ is obtained from an affine regression of U_{T+1} with respect to $N_1, \ldots, N_T$ in the model with random effects. This predictor is a bonus-malus coefficient for period $T+1$. Since $E(U_{T+1}) = 1$, this predictor is equal to $1 + {}^tb(n - \hat{\lambda})$ with

$$n = \underset{1\le t\le T}{\text{vec}}(n_t), \qquad \hat{\lambda} = \underset{1\le t\le T}{\text{vec}}(\hat{\lambda}_t), \qquad b = [\hat{V}(N)]^{-1}\widehat{\text{Cov}}(N, U_{T+1}). \tag{18}$$

All the estimations are taken in the random effects model, so we can use either the moment-based estimators or those derived from the GEE approach. The experience-rated frequency premium for period $T+1$ is the product of the bonus-malus coefficient and of the a priori frequency premium $\hat{\lambda}_{T+1}$. The former coefficient is equal to

$$1 + \sum_{t=1}^{T} \text{cred}_t\left(\frac{n_t}{\hat{\lambda}_t} - 1\right) \quad \text{with} \quad \text{cred}_t = \hat{\lambda}_t b_t. \tag{19}$$

Let us use the second-order moments computed in the preceding section. If we write $\widehat{\rho_U}(0) = 1$, the credibility coefficients are the solutions of the linear system

$$\text{cred}_t + \left(\hat{\lambda}_t\widehat{\sigma_U^2}\sum_{t'=1}^{T}\widehat{\rho_U}(|t'-t|)\,\text{cred}_{t'}\right) = \hat{\lambda}_t\widehat{\sigma_U^2}\widehat{\rho_U}(T+1-t) \quad \forall t = 1, \ldots, T. \tag{20}$$

If $T \geq T_{\max}$, this linear system uses an extension of the estimated correlogram, which is the main topic addressed by the paper.

If the frequency premiums are negligible with respect to one, we infer from the last equation: $\text{cred}_t \sim \hat{\lambda}_t \widehat{\sigma_U^2} \widehat{\rho_U}(T+1-t)$. A sequence of credibility coefficients should have the same shape as that of a correlogram with the same length and a reversed index.

The linear credibility approach outperforms the expected value principle in terms of computing time and readability of the experience rating schemes. There is a drawback however with this approach if we use intricate covariance structures for the random effects. The predictor can be negative even if the random effects model can be estimated on the data. Suppose for instance that the first estimated autocorrelation coefficient belongs to $]-1, 0[$. This unrealistic assumption would entail a financial penalty after a claim-free period. The bonus-malus coefficient and the experience-rated premium for the second period would then be negative if the number of claims reported in the first period was large enough.

On our data, correlation coefficients are greater than zero and decreasing. Let us give an upper bound for total credibility in this framework. We also suppose that credibility coefficients are positive, which is the case in our examples. We have $\widehat{\rho_U}(|t'-t|) \geq \widehat{\rho_U}(T+1-t)\ \forall t, t' = 1, \ldots, T$ since $\widehat{\rho_U}$ is decreasing. If we add the linear credibility equations (20) with respect to t, we obtain the following inequality:

$$\sum_{t=1}^{T} \text{cred}_t \leq \frac{\sum_{t=1}^{T} \hat{\lambda}_t \widehat{\sigma_U^2} \widehat{\rho_U}(T+1-t)}{1+\sum_{t=1}^{T} \hat{\lambda}_t \widehat{\sigma_U^2} \widehat{\rho_U}(T+1-t)}. \tag{21}$$

This upper bound is reached if the random effects are time-independent. If the frequency premiums per year are bounded and if $\sum_{h=1}^{+\infty} \widehat{\rho_U}(h) < +\infty$, total credibility will not converge towards one. Credibility is the discount granted to a claim-free history, and experience-rated premiums do not converge towards zero for such a history in real-world rating structures.

Prediction through an expected value principle is not subject to the criticism given above. With a parametric distribution family for the random effects (e.g. log-normal), prior and posterior expectations do not have a closed form, but can be approximated by numerical integration or by simulation. We did not retain this approach in the empirical study, since we would have to compute integrals of high dimension which are difficult to approximate. Moreover, there is no statistic of low dimension which summarizes the history in the expression of the bonus-malus coefficients, such as the sum of claims and the cumulative frequency premium in the model with constant random effects. The description of an experience rating scheme would then be very intricate.

4. Empirical results

4.1. The data set

The working sample of 80,994 policyholders represents 10% of the portfolio of a major Spanish insurance company. We selected only policies covering cars for private use and retained a balanced panel data set. Hence the individual histories have the same duration, which is equal to 7 years. We made this restriction because the GEE estimators are much easier to derive in this context.

Policyholders were observed between 1991 and 1997. Indicators of the calendar years are part of the regression components in order to allow for a trend in the past (see Besson and Partrat (1992) for optimal BMS with a trend). The other rating factors retained in the regression are gender, geographical area, the age of the driving license, the seniority and age of the policyholder, coverage level and the power of the vehicle. The average frequency of claims at fault per year is equal to 0.07. All the period durations are equal to 1 year, which means that the characteristics of the policyholders are known only at each anniversary date. In our view, this inaccuracy in the observation of regression components is of no consequence.

Table 1
Estimated autocorrelation coefficients

	h (lag)					
	1	2	3	4	5	6
$\hat{\rho}_U^0(h)$	0.733	0.524	0.504	0.483	0.425	0.401

4.2. Estimators for the correlogram of the random effects

The correlogram of $(U_t)_{1\leq t\leq 6}$ is estimated from the moment equations (5) and (6). The estimated variance of the process is equal to $\widehat{\sigma_U^2}^0 = 1.364$ and the autocorrelation coefficients are given in Table 1.

The estimated correlogram exhibits a decreasing shape, and there is less memory in the claim process if the panel data set under consideration is not balanced (see Pinquet et al., 2001).

In the Poisson model with random effects, the GEE estimator $\hat{\beta}$ defined in Eq. (10) has a lower variance–covariance matrix than $\hat{\beta}^0$, the m.l.e. in the Poisson model. The decrease in variance of an estimated linear form of the parameters ranges between 2 and 55% if GEE are used instead of the initial estimators.

The GEE estimators obtained from the 7-year histories are almost equal to those obtained from the simple moment equations. Slight differences are observed for the two following estimators:

$$\widehat{\sigma_U^2} = 1.366, \qquad \hat{\rho}_U(5) = 0.424. \tag{22}$$

The five other estimated autocorrelation coefficients have the same first three decimals. There is nothing odd about this result because the portfolio is very large. Even if the initial estimators are not optimal in terms of asymptotic variance, they are very accurate. The estimated coefficients of the regression are also very close, and the differences in premiums are always less than 2%.

From Table 1, we obtain the following values for the correlation and partial autocorrelation coefficients of the additive random effects (Table 2).

Tests for the nullity of partial autocorrelation coefficients enable the order of the autoregressive specification for random effects to be chosen. These tests were detailed at the end of Section 2 for $h = 1, 2, 3$. The estimated autocovariances $\gamma_h = \mathrm{Cov}(W_t, W_{t-h})$ of W are the following:

$$\hat{V}\begin{pmatrix}\widehat{\gamma_0}\\ \widehat{\gamma_1}\\ \widehat{\gamma_2}\\ \widehat{\gamma_3}\end{pmatrix} = 10^{-4}\times\begin{pmatrix}3.794 & 0.542 & 0.395 & 0.536\\ 0.542 & 2.964 & 0.784 & 0.683\\ 0.395 & 0.784 & 4.172 & 1.080\\ 0.536 & 0.683 & 1.080 & 4.929\end{pmatrix}.$$

Then the tests related to the null assumption $r_W(h) = 0 \Leftrightarrow g_h(\gamma) = 0$ are performed from the statistics

$$\frac{g_1(\hat{\gamma})}{\hat{\sigma}_{g_1(\hat{\gamma})}} = \frac{\widehat{\gamma_1}}{\hat{\sigma}_{\hat{\gamma}_1}} = 40.27, \qquad \frac{g_2(\hat{\gamma})}{\hat{\sigma}_{g_2(\hat{\gamma})}} = -0.604, \qquad \frac{g_3(\hat{\gamma})}{\hat{\sigma}_{g_3(\hat{\gamma})}} = 2.74. \tag{23}$$

Table 2
Estimated autocorrelation and partial autocorrelation coefficients

	h (lag)					
	1	2	3	4	5	6
$\hat{\rho}_W(h)$	0.806	0.627	0.608	0.588	0.531	0.507
$\hat{r}_W(h)$	0.806	−0.064	0.350	−0.002	0.053	0.089

On our data, the partial autocorrelation coefficients $r_W(1)$ and $r_W(3)$ are different from zero at the level $\alpha = 5\%$. We do not reject the nullity of $r_W(2)$. Hence an $AR(2)$ process for W does not significantly improve an $AR(1)$ specification, whereas an $AR(3)$ does.

4.3. Experience rating from the different models

In our preceding paper (Pinquet et al., 2001), we focused on the dynamic of bonus-malus coefficients. The main difference with credibility models based on time-independent random effects is the following. No-claim discounts caused by a claim-free year are lower than those obtained from the usual credibility model for a policyholder with a claim-free history, but they are higher if claims were reported recently. The explanation is the same in both cases. The credibility granted to a given period of the past decreases rapidly as time goes by, due partly to the increase of risk exposure but mostly to the decrease in the autocorrelation coefficients.

In this section, we first investigate the long-term behaviour of total credibility for different orders of the autoregressive process on the additive random effects. Let us consider a 40-year history and an average risk. Hence all the frequency premiums $(\widehat{\lambda_t})_{t=1,\ldots,40}$ are equal to 0.07 in the linear credibility system given in (20). We use the link given in (12) between the multiplicative and additive Gaussian random effects. We compare credibility models derived from time-invariant random effects and from $AR(p)$ specifications for additive Gaussian random effects with $p = 1, 3$, and 6. Autocorrelation coefficients for the additive Gaussian random effects are extended from the correlogram with the Yule–Walker equations, i.e.:

$$\hat{\rho}_W(h) = \sum_{k=1}^{p} \varphi_k \hat{\rho}_W(h-k) \quad \forall h \geq p. \tag{24}$$

In this linear recurrence equation on the autocorrelation coefficients, the $(\varphi_k)_{1\leq k\leq p}$ are the solutions of the linear system:

$$\hat{V}\left[\underset{1\leq k\leq p}{\text{vec}}\,(W_{t-k})\right] \underset{1\leq k\leq p}{\text{vec}}\,(\varphi_k) = \underset{1\leq k\leq p}{\text{vec}}\,(\widehat{\text{Cov}}(W_t, W_{t-k})) \tag{25}$$

which does not depend on t because of the stationarity assumption on $(W_t)_{t\geq 1}$. Then the correlogram of the multiplicative random effects results from the link given in Eq. (12).

The following table gives the credibilities granted to the first year, to the last year and to the whole history (Table 3).

Credibility coefficients decrease rapidly with the age of periods if random effects are dynamic. For instance, the last 2 (resp. 4, 5.5) years contribute to half the total credibility if an $AR(1)$ (resp. $AR(3)$, $AR(6)$) specification is used for the additive Gaussian random effects. Results are thus very different from the 20-year duration derived from time-independent random effects models.

We now derive bonus-malus coefficients from duration distributions on the policyholder–company relationship. Such distributions are assessed on an individual basis by Guillén et al. (2003). Actuarial models in non-life insurance usually estimate risks for the year to come, and the horizon is the renegotiation date of the contract. However, real-life rating structures create cross-subsidies between periods and implicitly refer to a more remote horizon.

Table 3
Long-term behaviour of credibility coefficients (40-year history)

Credibility coefficients (%)	$AR(1)$ specification for W_t	$AR(3)$ specification for W_t	$AR(6)$ specification for W_t
First year	10^{-4}	0.01	0.05
Last year	5.83	5.51	5.37
Total credibility	20.3	30.6	35.6

Table 4
Malus as a function of the attrition rate and of the order of the autoregressive process ($H = 30$ years)

	Malus (%) ($T = 1$, $H = 30$)	
	$r = 5\%$	$r = 20\%$
$AR(1)$	19.1	41.6
$AR(3)$	29.7	50.8
$AR(6)$	34.4	53.1

A rating model which allows for the age of periods can take into account an expected loyalty of the policyholder in the following way.

Let us denote a bonus-malus coefficient for period $T + 1$ as $\hat{E}(U_{T+1}|I_T)$, where U_{T+1} is the multiplicative random effect and where I_T is the information provided by periods $1, \ldots, T$. In the linear credibility approach, the expectation refers to a linear probabilistic regression and not to a conditional expectation. If $p_{T,h}$ is the probability that a policyholder in the portfolio during period T is still a customer of the insurance company at period $T + h$, we retain

$$\frac{\sum_{h=1}^{H} p_{T,h}\hat{E}(U_{T+h}|I_T)}{\sum_{h=1}^{H} p_{T,h}} \tag{26}$$

as a bonus-malus coefficient for period $T + 1$. In Eq. (26), H is the maximum number of years to consider in the future. This equation still provides a linear credibility predictor, which takes into account the duration distribution of the relationship between the policyholder and its insurance company.

If we suppose a constant attrition rate r (i.e. a constant probability of leaving the insurance company), we have $p_{T,h} = (1 - r)^h$. Bonus-malus scales for Markovian bonus-malus systems are computed with $H = 10$ years and at a null attrition rate by Brouhns et al. (2002).

For an average risk, the following table computes the increase in frequency premium after one claim reported during the first period for two attrition rates and different orders of the autoregressive process (Table 4).

A 5% attrition rate is roughly the average for a mutual insurance society in France; and 20% corresponds to a private insurance company. The results (which could be expressed in terms of credibility coefficients) are then strongly influenced by the expected loyalty of the policyholder to the insurance company.

5. Concluding remarks

Autoregressive specifications for time-dependent random effects provide a natural extension of the estimated correlation structure if the predictive ability of the periods decreases with their age. The purpose of the paper was to provide tools in order to determine the right order for the autoregressive process. The empirical results show that time-varying credibility models which use both retrospective and prospective points of view are strongly influenced by statistical modelling and by expectations concerning the loyalty of the policyholder.

References

Aitchison, J., Ho, C.H., 1989. The multivariate Poisson-log normal distribution. Biometrika 76, 643–653.
Besson, J.L., Partrat, C., 1992. Trend et systèmes de bonus-malus. ASTIN Bulletin 22, 11–32.
Brännäs, K., Johansson, P., 1995. Panel data regression for counts. Statistical Papers 37, 191–213.
Brouhns, N., Denuit, M., Guillén, M., Pinquet, J., 2002. Optimal Bonus-Malus scales in segmented tariffs, UCL working paper.
Bühlmann, H., 1967. Experience rating and credibility. ASTIN Bulletin 4, 199–207.
Davidson, R., MacKinnon, J.G., 1993. Stochastic Limit Theory. Oxford University Press, Oxford.

Frees, E.W., Young, V.R., Luo, Y., 1999. A longitudinal data analysis interpretation of credibility models. Insurance: Mathematics and Economics 24, 229–247.

Gerber, H., Jones, D., 1975. Credibility formulas of the updating type. Transactions of the Society of Actuaries 27, 31–52.

Guillén, M., Parner, J., Densgsoe, C., Perez-Marin, A.M., 2003. Using logistic regression models to predict and understand why customers leave an insurance company. In: Jain, L., Shapiro, A. (Eds.), Forthcoming in Intelligent Techniques in the Insurance Industry: Theory and Applications. World Scientific, Singapore.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.

Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.

Pinquet, J., Guillén, M., Bolancé, C., 2001. Allowance for the age of claims in bonus-malus systems. ASTIN Bulletin 31, 337–348.

Purcaru, O., Denuit, M., 2003. Dependence in dynamic claims frequency credibility models. ASTIN Bulletin, in press.

Sundt, B., 1988. Credibility estimators with geometric weights. Insurance: Mathematics and Economics 7, 113–122.

Zeger, L.S., 1988. A regression model for time series of counts. Biometrika 74, 721–729.

Insurance: Mathematics and Economics 43 (2008) 209–213

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

On the link between credibility and frequency premium

Catalina Bolancé [a], Montserrat Guillén [a], Jean Pinquet [b,*]

[a] Dept. d'Econometria, Estadistica i Economia Espanyola, Universitat de Barcelona, Diagonal 690, 08034 Barcelona, Spain

[b] Département d'Economie, Ecole Polytechnique, 91128 Palaiseau Cedex, France

ARTICLE INFO

Article history:
Received September 2007
Received in revised form
May 2008
Accepted 26 May 2008

JEL classification:
C14
C25

MSC:
IM31
IB40

Keywords:
Generalized linear models
Parametric and nonparametric links
Gamma process

ABSTRACT

This paper questions the equidistribution assumption for the random effects in a frequency risk model. Two models are presented, which use parametric and nonparametric links between the variance of the random effect and frequency risk. They are estimated on a Spanish automobile insurance portfolio, for which a decreasing link is obtained. Conclusions are drawn for credibility and bonus-malus coefficients.

1. Introduction

Standard actuarial models assume that the random effects used in the distributions of the number of claims are identically distributed. In this framework, the credibility granted to the history of the policyholder increases with the frequency premium. Credibility is the no claims discount for a claimless history and the increasing relationship between an actuarial bonus and an estimated risk level is quoted in various papers (e.g. Dionne and Vanasse (1989)).

There is however empirical evidence of a decreasing link between the variance of the random effect and the frequency risk for automobile insurance data (see Sections 3 and 4 with results obtained from a Spanish portfolio). In other words, the residual relative heterogeneity on claims number distributions is more important for low risks.

In this paper, the variance of the random effect applied to Poisson distributions is conditioned on frequency risk and hence on the rating factors. First, we retain a local estimation approach of a nonparametric link between the variance and the frequency.[1] Section 2 summarizes the main properties of kernel-based estimators in generalized linear models. This approach is used in Section 3 to estimate the nonparametric link. Second, a parametric power link is specified and estimated in Section 4 from the negative binomial model. Consequences of the credibility derivation are drawn in Section 5. Section 6 concludes and an appendix contains some mathematical details.

The main empirical finding is that the link between credibility (or no claims discount) and the frequency premium is lower when the equidistribution assumption of the random effect is relaxed than it is in the usual model. The opposite result is obtained for the increase in premium after a claim.

2. Kernel estimators in generalized linear models: The index model

Generalized linear models for a response variable Y and a vector of regressors $X (X \in \mathbb{R}^k)$ assume that

$$E(Y \mid X = x) = f(x'\beta), \quad \beta \in \mathbb{R}^k. \tag{1}$$

* Corresponding author. Tel.: +33 169333423; fax: +33 169333427.
E-mail addresses: bolance@ub.edu (C. Bolancé), mguillen@ub.edu (M. Guillén), jean.pinquet@polytechnique.edu (J. Pinquet).

0167-6687/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2008.05.015

[1] Local estimation techniques can also be used for prediction on time series (see Qian (2000) for applications to insurance.

The function f is a link between an index $x'\beta$ (i.e. a scalar product between regressors and parameters) and the expectation of the response variable. In the literature on generalized linear models, the link usually refers to the reciprocal of f, but here we retain the function which is estimated in the first place. For identifiability purposes detailed later, we suppose that the intercept is not included as a regressor in (1) and that the distribution of X is nondegenerate in $\mathbb{R}^k$. For basic generalized linear models, f is given and then an intercept must be included in the regression (see McCullagh and Nelder (1989)). For a count data model, f is usually the exponential function. If f is unknown in the specification so that an estimation is required, Eq. (1) is referred to as an index model (see Härdle et al. (1997)). In that case, there is an obvious identifiability conflict between β and f in Eq. (1). Only the line $\mathbb{R}\beta$ can be identified from the data. In other words, what is identified is the conditional expectation, assumed to be constant on affine hyperplanes of $\mathbb{R}^k$ which are orthogonal to a given vector. For a given value of β, a nonparametric estimation of $f(s)$ can be based on local weighted averages of the response variable, with weights which decrease with the distance between s and the individual values of the index.

A first estimation of index models can be obtained from a parametric specification of the distribution of Y defined conditionally on X. Let us assume that we have, in that case,

$$E(Y \mid X = x) = f_0(c + x'b); \quad c \in \mathbb{R},\ b \in \mathbb{R}^k, \tag{2}$$

with f_0 a given link function. Let $\widehat{b}$ be the maximum likelihood estimator of b. The conditional expectation defined in (1) can be estimated with a kernel estimator.

In a sampling model on (X, Y) with n observations $(x_i, y_i)_{i=1,\dots,n}$, an estimator of $E(Y \mid X)$ of the Nadaraya–Watson type is obtained from a kernel K (usually an even probability density function) and a bandwidth h in the following way:

$$\widehat{E}(Y \mid X) = \frac{\sum_{i=1}^{n} y_i K_h\left(X'\widehat{b} - x_i'\widehat{b}\right)}{\sum_{i=1}^{n} K_h\left(X'\widehat{b} - x_i'\widehat{b}\right)}, \qquad K_h(u) = \frac{K\left(\frac{u}{h}\right)}{h}. \tag{3}$$

The bandwidth is a smoothing parameter. The closer it is to zero, the more estimation is performed on a local basis. The estimation given in (3) exhibits an invariance property as it only depends on $\widehat{b}/h$.

A suitable value of h can be derived from a cross-validation method similar to that proposed in Härdle (1990) for the Nadaraya–Watson kernel estimator. It is equal to

$$\arg\min_h \ CV(\widehat{b}, h) = \sum_{i=1}^{n} (y_i - \hat{E}_{-i}(Y_i))^2, \tag{4}$$

where $\hat{E}_{-i}(Y_i)$ is the leave-one-out estimator[2] defined by

$$\hat{E}_{-i}(Y_i) = \frac{\sum_{j\neq i} y_j K_h\left(x_i'\widehat{b} - x_j'\widehat{b}\right)}{\sum_{j\neq i} K_h\left(x_i'\widehat{b} - x_j'\widehat{b}\right)} = \frac{\sum_{j\neq i} y_j K\left(\left(x_i - x_j\right)'\frac{\widehat{b}}{h}\right)}{\sum_{j\neq i} K\left(\left(x_i - x_j\right)'\frac{\widehat{b}}{h}\right)}. \tag{5}$$

This estimation of the conditional expectation $E(Y \mid X)$ is not necessarily consistent since it is derived from a wrong link function (f_0 instead of f). Two results on this issue are worth mentioning.

- On one hand, a consistent estimator of the conditional expectation can be obtained from the cross-validation criterion defined in (4) and (5). Replacing $\widehat{b}$ by b in (4) defines a function $CV(b, h)$ which can be minimized with respect to b and h. The optimal values of b and h are plugged into the expression for the estimated conditional expectation given in (3). Sufficient conditions for the consistency of the estimation are given in Härdle et al. (1993). However the minimization of the cross-validation criterion is cumbersome, since it necessitates a double sum on individuals, also, Eq. (5) shows that only b/h is identified. Hence an identifying constraint needs to be added in the estimation.
- On the other hand, the maximum likelihood estimation of a parametric model given in the first place can lead to consistent estimation of the conditional expectation under conditions which are first related to the distribution of the regressor X (see Li and Duan (1989), and the Appendix). Owing to the identification issue mentioned above, consistency means that $\widehat{b}$ converges towards a limit b_0 which belongs to the line $\mathbb{R}\beta$.

3. Kernel estimators for the variance of the random effect in a Poisson model

Let us consider cross-section data. The policyholders in the portfolio are indexed by $i = 1, \dots, p$. All the risk exposure durations are supposed equal (they are equal to one year in our empirical study). Frequency risk must be expressed for a time unit, otherwise the results on the link investigated in this paper would not be coherent with respect to period aggregation. We denote n_i as the number of claims reported by policyholder i, N_i the related random variable and x_i as the vector of regression components. If U_i is the random effect, the distribution of N_i in the Poisson model with random effects is obtained from an expectation taken with respect to the random effect U_i

$$P[N_i = n_i] = E\left[P_{\lambda_i U_i}(n_i)\right]; \quad \lambda_i = \exp(c + x_i'b);$$
$$P_\lambda(n) = \exp(-\lambda)\frac{\lambda^n}{n!}.$$

If the equidistribution assumption of the random effects is relaxed, we can link their variance and frequency risk and write for instance

$$E(U_i) = 1; \qquad V(U_i) = \sigma^2(\lambda_i), \qquad \lambda_i = E(N_i). \tag{6}$$

The function σ^2 must have nonnegative values. We will investigate a power link in Section 4 but we first let the data speak from a nonparametric estimation of σ^2. The starting point is the usual moment-based estimator

$$V(U_i) = \frac{V(N_i) - E(N_i)}{E^2(N_i)} = \frac{E(N_i^2) - E^2(N_i) - E(N_i)}{E^2(N_i)}. \tag{7}$$

The nonparametric link is then obtained with an index model strategy described in Section 2. First, an estimation is performed from the maximum likelihood estimation of the Poisson model with regression components

$$N_i \sim P(\lambda_i), \quad \lambda_i = \exp(c + x_i'b).$$

For each policyholder i, we obtain the index $s_i = \widehat{c} + x_i'\widehat{b}$ and the parametric frequency premium $\widehat{E}^0(N_i) = \exp(s_i)$. Then the variance of the random effect is estimated from Eq. (7) and from nonparametric estimators $\widehat{E}(N_i^m)$, $m = 1, 2$. These estimators are derived from Eq. (3), with $Y = N^m$, $m = 1, 2$. In what follows, we retain a Gaussian kernel. Hence K_h is the density of a $N(0, h^2)$ distribution. The bandwidth h_m retained for the estimation of $E(N_i^m)$ $(m = 1, 2)$ is obtained with the leave-one-out approach

[2] The individual for which the non parametric expectation is derived must be withdrawn from the computation, otherwise the bandwidth would converge towards zero in the minimization of the cross-validation criterion.

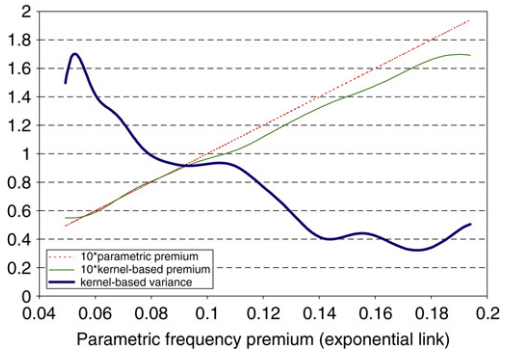


Fig. 1. Kernel-based estimation of the variance of the random effect and of the frequency premium as a function of the parametric frequency premium.

(see Eqs. (4) and (5)). From Eqs. (7) and (3), a nonparametric estimator of $V(U_i)$ is

$$\widehat{V}(U_i) = \frac{\widehat{E}(N_i^2) - \widehat{E}^2(N_i) - \widehat{E}(N_i)}{\widehat{E}^2(N_i)}.$$

The estimated variance of the random effect is then a nonparametric function of the frequency premium $\widehat{E}^0(N_i)$, since this property holds for the estimated moments $\widehat{E}(N_i^m)$, $m = 1, 2$.

Fig. 1 presents the non parametric estimations $\widehat{V}(U)$ and $\widehat{E}(N)$ as functions of $\widehat{E}^0(N)$ for a Spanish portfolio containing 80,994 policyholders observed during one year.[3] The working sample of 80,994 policyholders represents 10% of the portfolio of a major Spanish insurance company. The response variable is the number of claims at fault with respect to a third party. We selected only policies covering cars for private use in 1991. The average frequency of claims per year is equal to 0.07, and the average cost of claims is about 550 euros. More details on this data set can be found in Bolancé et al. (2003). The selected contracts do not include a deductible for third party liability claims. The rating factors are the gender, the geographical area, the age of the driving licence, the age of the policyholder and her seniority as a customer of the insurance company, the coverage level and the power of the vehicle. Fig. 1 is obtained with the two step estimation approach described after Eq. (7). A direct estimation of the regression parameters jointly with the bandwidth from the cross-validation criterion led to very similar results. For instance, the cosine between the two estimations linked to the regressors is equal to 0.9996, which indicates almost perfect colinearity. As only b/h is identified (see Eq. (5) and the following comments), this means that the estimations are almost equivalent. This result was not obvious *ex ante,* as consistency with link misspecification holds under assumptions on the distribution of regressors which are not necessarily fulfilled with the qualitative variables used in our regressions (see the Appendix). Fig. 1 exhibits a decreasing link between the local estimation for the variance of the random effect and the frequency premium. This result was confirmed on other data bases. The peak in the estimated variance between 0.04 and 0.06 can be explained by the effect of boundary bias that occurs in kernal estimations near the extremes of the domain interval (see Wand and Jones (1995)). We did not correct for this effect because the main focus is on the overall decreasing trend.

A power link between the variance of the random effect and frequency risk is estimated in the following section.

[3] The bandwidths retained for the first and second order moments of the response variable from the cross-validation criterion given in (4) are equal to $h_1 = h_2 = 0.009$. Notice that the two estimations of the frequency are close to each other around the average value (equal to 0.07), whereas the kernel-based estimation is lower than the fully parametric estimation for large values of the frequency.

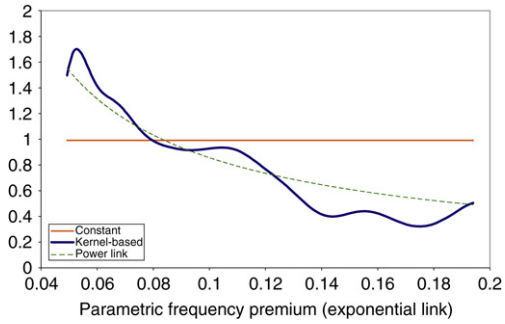


Fig. 2. Kernel-based vs. power link parametric estimation for the variance of the random effect as a function of the parametric frequency premium.

4. Negative binomial model with a power link between the variance of the random effect and frequency risk

In this section, we use a modified version of the negative binomial model with regression components, in which we include a power link between the variance of the random effect and frequency risk. Hence we have a parametric model which can reflect the decreasing link observed in Fig. 1. This model was proposed by Winkelmann and Zimmermann (1991).

We keep the notations of Section 3, and we suppose that the random effect follows a gamma distribution. We write

$$U_i \sim \gamma(a_i, a_i); \quad a_i = a\lambda_i^{-e}, \qquad \lambda_i = E(N_i) = \exp(c + x_i' b).$$

The gamma distribution is indexed by a shape and a scale parameter. We have

$$E(U_i) = 1; \qquad V(U_i) = \frac{1}{a_i} = \frac{\lambda_i^e}{a}, \tag{8}$$

and the parameter e is the elasticity of $V(U_i)$ with respect to frequency risk. The equidistribution assumption for the random effects means that this elasticity is null.

The variance of the random effect specified by (8) is an exponential function of the index $x_i' b$. Such a link was investigated for a linear model by Harvey (1976).

The likelihood is the usual one for the negative binomial model, hence

$$P[N_i = n_i] = E\left[P_{\lambda_i U_i}(n_i)\right] = \frac{a_i^{a_i}\lambda_i^{n_i}}{(\lambda_i + a_i)^{n_i + a_i}} \frac{\Gamma(n_i + a_i)}{\Gamma(a_i)\Gamma(n_i + 1)}.$$

The maximum likelihood estimators for the Spanish portfolio are the following.

$$\widehat{a} = 8.05; \qquad \widehat{e} = -0.839; \qquad \frac{\widehat{e}}{\widehat{\sigma_e}} = -2.141. \tag{9}$$

The estimated elasticity $\widehat{e}$ is negative, a result in accordance with the expected decreasing link between the estimated variance of the random effect and the frequency premium. The usual semiparametric estimator for a constant variance of the random effect is equal to

$$\widehat{\sigma^2} = \widehat{V}(U) = \frac{\sum_i (n_i - \widehat{\lambda}_i)^2 - n_i}{\sum_i \widehat{\lambda}_i^2} = 0.985.$$

Fig. 2 plots two estimated variances of the random effects, defined as a power or a kernel-based function of the parametric frequency premium, together with the constant estimation. The estimated variances of the random effect that are linked with the frequency premium (in a parametric or nonparametric way) are greater than the constant estimation if the frequency premium is

less than 0.08, and smaller afterwards. This threshold is close to the annual frequency premium, which is roughly equal to 0.07. We estimated this model on other data bases and always found this decreasing link, but no stable discrepancy between the parametric and nonparametric estimation.

5. Applications to credibility predictors

Linking the variance of the random effect and the frequency risk raises a difficulty in a longitudinal data analysis. Since frequency risk varies with time, the random effect cannot be supposed time-independent in the prediction. A solution is to derive the random effects from a stochastic process defined in continuous time, and to link the time index of the process with the frequency risk. Let us detail an example from a gamma process.

If the data are longitudinal, we simply replace i by the pair i, t (where t indexes the periods) in the preceding equations. For instance, the random effects are defined as follows

$$U_{i,t} \sim \gamma(a_{i,t}, a_{i,t}) \Leftrightarrow a_{i,t}U_{i,t} \sim \gamma(a_{i,t}). \tag{10}$$

We obtain the random effects $U_{i,t}$ from a gamma process $\left(G_{i,a}\right)_{a\geq 0}$, which is defined by three properties:

(1) $G_{i,0} \equiv 0$; (2) The process has independent increments; (3) The increments follow gamma distributions, with a parameter equal to the difference in dates (i.e. $G_{i,a_2} - G_{i,a_1} \sim \gamma(a_2 - a_1)\ \forall a_1, a_2,\ a_2 > a_1 \geq 0$). For instance, we have $G_{i,a} \sim \gamma(a)$.

We have a particular type of Levy process with indefinitely divisible distributions. This process exists from the well known property on gamma distributions

$$\gamma(a_1) * \gamma(a_2) = \gamma(a_1 + a_2) \quad (a_1,\ a_2 \geq 0)$$

and from Kolmogorov's theorem. The gamma process can also be seen as a limit of compound Poisson processes (see Dufresne et al. (1991), for definitions and applications to ruin theory).

Suppose that we have the link $V(U_{i,t}) = \sigma^2(\lambda_{i,t})$ given in Eq. (6) between the variance of the random effect and the frequency risk. For the distributions given in Eq. (10), we have $V(U_{i,t}) = 1/a_{i,t}$. This leads us to define the random effects as follows

$$U_{i,t} = \sigma^2(\lambda_{i,t}) \times G_{i,1/\sigma^2(\lambda_{i,t})}. \tag{11}$$

Then we have

$$\lambda_{i,t_1} = \lambda_{i,t_2} \Rightarrow U_{i,t_1} = U_{i,t_2}.$$

We can consider for instance the power link $\sigma^2(\lambda_{i,t}) = a \times \lambda_{i,t}^{-e}$ retained in Section 4.

Let us predict frequency risks with a linear credibility approach (Bühlmann, 1967). The bonus-malus coefficient for the second period of the policyholder i is derived from an affine probabilistic regression of U_{i2} with respect to N_{i1}. We have

$$\widehat{u}_{i2} = 1 + \frac{\widehat{\mathrm{Cov}}(U_{i2}, N_{i1})}{\widehat{V}(N_{i1})}(n_{i1} - \widehat{\lambda_{i1}}),$$

with

$$\widehat{\mathrm{Cov}}(U_{i2}, N_{i1}) = \widehat{\lambda_{i1}}\widehat{\mathrm{Cov}}(U_{i2}, U_{i1}); \qquad \widehat{V}(N_{i1}) = \widehat{\lambda_{i1}} + \widehat{\lambda_{i1}}^2\widehat{V}(U_{i1}).$$

Since the gamma process $G_{i,t}$ has independent increments, we have that

$$\mathrm{Cov}(G_{i,a_1}, G_{i,a_2}) = V(G_{i,\min(a_1,a_2)}) = \min(a_1, a_2).$$

From (11) and the last equation, we obtain

$$\widehat{\mathrm{Cov}}(U_{i2}, U_{i1}) = \frac{\widehat{\sigma^2}(\widehat{\lambda_{i1}}) \times \widehat{\sigma^2}(\widehat{\lambda_{i2}})}{\max\left(\widehat{\sigma^2}(\widehat{\lambda_{i1}}),\ \widehat{\sigma^2}(\widehat{\lambda_{i2}})\right)}.$$

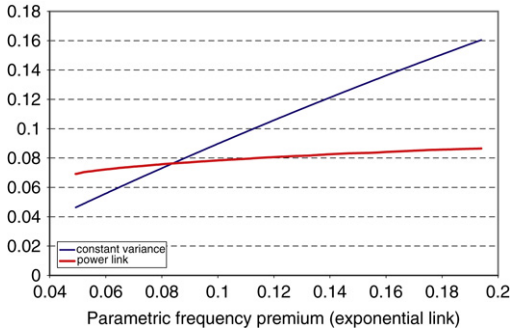


Fig. 3. Credibility as a function of the frequency premium (with a constant variance for the random effect and with a power link between frequency risk and the variance).

The linear credibility predictor is equal to

$$\widehat{u}_{i2} = 1 + \mathrm{cred}_i\left(\frac{n_{i1}}{\widehat{\lambda_{i1}}} - 1\right) = (1 - \mathrm{cred}_i) + \left(\mathrm{cred}_i \times \frac{n_{i1}}{\widehat{\lambda_{i1}}}\right).$$

The credibility cred_i granted to the first period is equal to

$$\begin{aligned}\mathrm{cred}_i &= \frac{\widehat{\lambda_{i1}} \times \widehat{\mathrm{Cov}}(U_{i2}, N_{i1})}{\widehat{V}(N_{i1})} \\ &= \frac{\widehat{\lambda_{i1}} \times \widehat{\sigma^2}(\widehat{\lambda_{i1}})}{1 + \left(\widehat{\lambda_{i1}} \times \widehat{\sigma^2}(\widehat{\lambda_{i1}})\right)} \times \frac{\widehat{\sigma^2}(\widehat{\lambda_{i2}})}{\max\left(\widehat{\sigma^2}(\widehat{\lambda_{i1}}),\ \widehat{\sigma^2}(\widehat{\lambda_{i2}})\right)}.\end{aligned} \tag{12}$$

Let us now consider Eq. (12). The first component of the product which defines the credibility is the usual formula, with a variance of the random effect $\widehat{\sigma^2}$ which depends on the frequency premium. As the function $\widehat{\sigma^2}$ decreases with the frequency premium on our data, this component of the credibility increases less than with the usual formula. It might even decrease if the estimated elasticity between the variance of the random effect and frequency risk (defined globally with a power link function or locally with a nonparametric link) was less than -1. The second component of the credibility in Eq. (12) is less than one. Frequency premiums do not vary much for a given policyholder from one period to the following,[4] and this component remains close to one. All the policyholders of our sample stay in the portfolio at the second period.[5] On our data, the average of the second component of (12) derived from the power link estimated in (9) is equal to 0.99. In Fig. 3, we plot two derivations of the credibility as a function of the frequency premium $\widehat{\lambda_{i1}}$. First, credibility is obtained from the usual formula with a constant variance for the random effect ($\widehat{\sigma^2} = 0.985$). The second derivation follows from Eq. (9) and (12). The credibility is not a deterministic function of $\widehat{\lambda_{i1}}$, due to the second component in Eq. (12). We averaged it with a Gaussian kernel. The optimal bandwidth is small ($h = 0.001$) because the variations around the average are very low.

As expected, the credibility is almost constant with the power link between the variance of the random effect and the frequency premium because the estimated elasticity is close to -1.

6. Conclusions

The relative variations after each year for coefficients of real-world bonus-malus scales do not depend on the frequency

[4] A thorough analysis of stochastic migration between risk levels during different periods is given in Brouhns et al. (2003).

[5] They are actually present during seven years. See Bolancé et al. (2003) for an investigation of the whole panel data set.

premium (however they do depend on their location in the scale: see (Lemaire, 1995), for a review). Relaxing the equidistribution assumption on the random effects may allow actuarial models to get closer to real-world rating structures concerning the bonus, if the estimated link between the variance of the random effect and frequency risk is decreasing. As an actuarial malus is close to the variance of the random effect if risk exposure is low, the malus also decreases with the frequency premium if the aforementioned link is decreasing. Hence a real-world bonus-malus scale is close to a usual actuarial model on the malus side, whereas on the bonus side it can be closer to the models developed in this paper.

Acknowledgements

The authors acknowledge financial support from FEDER and from the Spanish Ministry of Education and Science. Jean Pinquet acknowledges financial support from the AXA chair "Large Risks in Insurance".

Appendix. Consistency of the estimation under link violation

This appendix gives the conditions which provide a consistent estimation of the conditional expectation $E(Y \mid X)$ in a misspecification context on the link between the index and $E(Y)$. The reference is Li and Duan (1989).

Suppose that we have a sampling model on the pair (x, y), with the nonparametric link $E(Y \mid X = x) = f(x'\beta)$. Notations are those of Section 2. In the aforementioned reference, the distribution of Y given $X = x$ is supposed to be that of $g(x'\beta, \varepsilon)$, where ε is a given random variable and g a given function. This implicitly assumes that the distribution of Y is determined by the expectation $E(Y)$, which is the case for binary or Poisson distributions.

In addition, a parametric model with a scalar parameter m and a likelihood L_m is estimated on the data (observations are $(x_i, y_i)_{i=1,\dots,n}$). The scalar parameter is linked to the index $x'b$ by a given function f_0. We write

$$m_i = f_0(c + x_i'b); \quad -\log L_{m_i}(y_i) = h(c + x_i'b,\ y_i).$$

For instance, in a Poisson model the parameter is the expectation and the exponential link is usually retained for f_0. In that case we have: $h(s, y) = \exp(s) - ys + \log(y!)$.

Let $\widehat{c}, \widehat{b} = \arg\min_{c,b} \sum_i h(c + x_i'b,\ y_i)$ be the maximum likelihood estimators in the parametric model. If data are generated by the sampling model given in the first place, this estimator converges towards a limit c_0, b_0 usually called a pseudo-true value (Gouriéroux et al., 1984). We have that

$$b_0 \in \mathbb{R}\beta \quad (\text{i.e. } b_0 = \lambda\beta,\ \lambda \in \mathbb{R}) \tag{13}$$

under the following assumptions.

1. The maps $h(\bullet, y) : s \to h(s, y)$ are convex for every value of y. This assumption implies that the map $c, b \to E\left[h(c + X'b,\ Y)\right] = R(c, b)$ is convex.
2. The minimum of the map R defined in Assumption 1 is reached for only one pair c_0, b_0.
3. We have the following property

$$\forall b \in \mathbb{R}^k,\ \exists d, \lambda \in \mathbb{R}, \quad E(X'b \mid X'\beta) = d + (\lambda X'\beta). \tag{14}$$

Commenting now on the result and the assumptions: since we have the property $\widehat{b} \xrightarrow{a.e.} b_0 = \lambda\beta$, the estimation of $E(Y \mid X)$ obtained from (3) and (4) will be consistent. Indeed, it is the line $\mathbb{R}\beta$ which must be identified in an index model, as discussed in Section 2 and hence the multiplicative constant λ is not important in the estimation.

The first assumption relates to the concavity of the log-likelihood and is usually fulfilled. The consistency property of the maximum likelihood estimation and Assumption 2 lead to

$$\widehat{c}, \widehat{b} = \arg\min_{c,b} \frac{1}{n}\sum_{i=1}^{n} h(c + x_i'b,\ y_i) \xrightarrow{a.e.} c_0, b_0$$
$$= \arg\min_{c,b} R(c, b) = E\left[h(c + X'b,\ Y)\right].$$

From the strong law of large numbers, it is easily seen why Assumption 2 is necessary.

The consistency result given in (13) is obtained from (14) with Assumption 2 and Jensen's inequality applied on the functions $h(\bullet, y)$. Assumption 3 is generally not fulfilled for variables X with discrete values, as is the case in our empirical analysis. Eq. (14) is fulfilled if X follows a non degenerate Gaussian distribution. Indeed, it is well known that in that case the conditional expectation defined in (14) is obtained from the affine probabilistic regression of $X'b$ with respect to $X'\beta$. Property (14) is more generally fulfilled for elliptically symmetric distributions.

References

Bolancé, C., Guillén, M., Pinquet, J., 2003. Time-varying credibility for frequency risk models: Estimation and tests for autoregressive specifications on the random effects. Insurance: Mathematics and Economics 33, 273–282.
Brouhns, N., Guillén, M., Denuit, M., Pinquet, J., 2003. Bonus-malus scales in segmented tariffs with stochastic migration between segments. The Journal of Risk and Insurance 70, 577–599.
Bühlmann, H., 1967. Experience rating and credibility. ASTIN Bulletin 4, 199–207.
Dionne, G., Vanasse, C., 1989. A generalization of automobile insurance rating models: The negative binomial distribution with a regression component. ASTIN Bulletin 19, 199–212.
Dufresne, F., Gerber, H., Shiu, E., 1991. Risk theory with the gamma process. ASTIN Bulletin 21, 177–192.
Gouriéroux, C., Monfort, A., Trognon, A., 1984. Pseudo likelihood methods: Theory. Econometrica 52, 681–700.
Härdle, W., 1990. Applied nonparametric regression. In: Econometric Society Monographs. Cambridge University Press.
Härdle, W., Hall, P., Ichimura, H., 1993. Optimal smoothing in single index models. The Annals of Statistics 21, 157–178.
Härdle, W., Spokoiny, V., Sperlich, S., 1997. Semiparametric single index versus fixed link function modelling. The Annals of Statistics 25, 212–243.
Harvey, A.C., 1976. Estimating regression models with multiplicative heteroscedasticity. Econometrica 44, 461–465.
Lemaire, J., 1995. Bonus-Malus Systems in Automobile Insurance. In: Huebner International Series on Risk, Insurance and Economic Security.
Li, K.C., Duan, N., 1989. Regression analysis under link violation. The Annals of Statistics 21, 157–178.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, Second edition. Chapman and Hall, New York.
Qian, W., 2000. An application of nonparametric regression estimation in credibility theory. Insurance: Mathematics and Economics 27, 169–176.
Wand, M.P., Jones, M.C., 1995. Kernel Smoothing. Chapman & Hall, London.
Winkelmann, R., Zimmermann, K.F., 1991. A new approach for modelling economic count data. Economic Letters 37, 139–143.

Identification et évaluation des effets incitatifs à la prévention des risques créés par les historiques en assurance

- Synthèse sur l'identification entre antisélection et aléa moral à partir de données d'assurance

  J. Abbring, P.A. Chiappori, J. Heckman et J. Pinquet (2003) "Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish?", *Journal of the European Economic Association, Papers and Proceedings* **1**, 512-521.

- Modèles théoriques et économétriques relatifs à l'identification entre antisélection et aléa moral à partir de données d'assurance

  J. Abbring, P.A. Chiappori et J. Pinquet (2003) "Moral hazard and dynamic insurance data", *Journal of the European Economic Association* **1**, 767-820.

- Analyse théorique et économétrique des mécanismes incitatifs à une conduite prudente, et analyse de l'expérience québecoise

  G. Dionne, J. Pinquet, C. Vanasse et M. Maurice (2009) "Incentive mechanisms for safe driving: A comparative analysis with dynamic data", à paraître dans *The Review of Economics and Statistics*.

- Compléments à la publication précédente (analyse théorique des permis à points avec amnistie totale)

  Extraits du document de travail de G. Dionne, J. Pinquet, C. Vanasse et M. Maurice (2007): "Point-record incentives, asymmetric information and dynamic data", accessible à l'URL
  `http://hal.archives-ouvertes.fr/hal-00243056/fr/`

ADVERSE SELECTION AND MORAL HAZARD IN INSURANCE: CAN DYNAMIC DATA HELP TO DISTINGUISH?

Jaap H. Abbring
Free University

Pierre-André Chiappori
University of Chicago

James J. Heckman
University of Chicago

Jean Pinquet
Université Paris X-Nanterre

Abstract
A standard problem of applied contracts theory is to empirically distinguish between adverse selection and moral hazard. We show that dynamic insurance data allow to distinguish moral hazard from dynamic selection on unobservables. In the presence of moral hazard, experience rating implies negative occurrence dependence: individual claim intensities decrease with the number of past claims. We discuss econometric tests for the various types of data that are typically available. Finally, we argue that dynamic data also allow to test for adverse selection, even if it is based on asymmetric learning. (JEL: D82, G22, C41, C14)

1. Introduction

For two decades, contract theory has remained a predominantly theoretical field. However, a number of papers have recently been devoted to empirical applications of the theory.[1] It has been argued that insurance offers a particularly promising field for empirical work on contracts. Individual (automobile, housing, health, life, etc.) insurance contracts are largely standardized. Researchers have access to databases of insurance companies, which typically contain several millions of such contracts. The information in these databases can generally be summarized in a reasonably small number of quantitative and qualitative indicators. The 'outcome' of the contract—be it the occurrence of an accident, its cost, or some level of expenditure—is very precisely recorded in the firms' files, together with a detailed history of the contractual relationship (changes in coverage, etc.). Not surprisingly, several recent papers are aimed at

Acknowledgments: Support from the NSF (grant #0096516) is gratefully acknowledged. This paper was written while Jaap Abbring was visiting the Department of Economics of University College London. Abbring's research is supported by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

E-mail addresses: Abbring: jabbring@econ.vu.nl; Chiappori: pchiappo@midway.uchicago.edu; Heckman: jheckman@midway.uchicago.edu; Pinquet: pinquet@u-paris10.fr

1. See Chiappori and Salanié (2003) for a recent survey.

testing for the existence and estimating the magnitude of asymmetric information effects in competitive insurance markets.[2]

A popular strategy for studying asymmetric information is to test, conditional on observables, for a correlation between the choice of a contract and the occurrence or severity of an accident. Under adverse selection on risk, 'high-risk' agents are, everything else equal, both more likely to choose a contract with more complete coverage and more likely to have an accident. The basic moral hazard story is very close to the adverse selection one, except for an inverted causality. In a moral hazard context, agents first choose different contracts. Then, an agent facing better coverage and, therefore, weaker incentives will be less cautious and have more accidents. In both cases, the same pattern emerges: controlling for observables, more comprehensive coverage should be associated with higher realized risk—a property that can be tested using appropriate parametric or non-parametric techniques.

The conditional correlation approach has several advantages. It is simple and very robust, as argued by Chiappori et al. (2002). Furthermore, it can be used on static, cross-sectional data that are relatively easy to obtain. However, these qualities come at a cost. The past history of the relationship influences both the current contract (through experience rating) and the agent's behavior, and this effect is hard to take into account with cross-sectional data. More importantly, the correlation is not informative on the direction of the causality, which makes the two stories (moral hazard and adverse selection) very hard to distinguish. Still, such a distinction is crucial, if only because the optimal form of regulation of insurance markets varies considerably with the context.[3]

The research program summarized in the present paper relies on the insight that the dynamic aspects of the relationship can help distinguishing between adverse selection and moral hazard. Two approaches can be distinguished. First, the form of optimal dynamic contracts differs considerably between the two cases. Thus, the qualitative properties of observed contracts may provide useful insights into the type of problem they are designed to address. The research program described in this paper concentrates on a second approach, in which the (possibly suboptimal) contracts are taken as given and we concentrate on their implications for observed behavior. In particular, most 'real life' insurance contracts exhibit some form of experience rating. A typical property of experience rating schemes is that the occurrence of an accident shifts the entire incentive scheme the agent is facing. Under moral hazard, this results in a form

2. Pueltz and Snow (1994), Dionne and Vanasse (1992), Chiappori and Salanié (1997, 2000), Dionne, Gouriéroux, and Vanasse (1997), Richaudeau (1999) and Dionne et al. (2001), to name only a few, analyze automobile insurance contracts, while Holly et al. (1998), Chiappori, Durand, and Geoffard (1998), Chiappori, Geoffard, and Kyriadizou (1998), Cardon and Hendel (1998) and Hendel and Lizzeri (1999) use health or life insurance data, and Poterba and Finkelstein (2003) consider annuity contracts.

3. For a review of various attempts to distinguish between moral hazard and adverse selection, see Chiappori (2000).

of autocorrelation in the accident process. Thus, an empirical analysis of this process can be informative on the presence of moral hazard.

In addition, dynamic data allow to address the problem of asymmetric learning. Conventional wisdom suggests that, in many cases, asymmetric information may not be present at the beginning of the relationship (e.g., the relative quality of a young driver is unknown to her and her insurer). Rather, it emerges gradually as a consequence of different learning processes (say, the young driver learns from near misses that are not even observed by the insurer). Then the contractual changes that take place during the relationship may be informative about the agent's riskiness, even if the initial choice of a contract is uncorrelated with residual risk (as found by most studies).

2. Dynamic Moral Hazard Under Experience Rating: Theory

The model is directly borrowed from Chiappori and Heckman (2000) and Abbring et al. (forthcoming). We consider a dynamic version of an insurance model à la Mossin (1968). Time is discrete. In each period t, the agent receives an income normalized to one and may with some probability $(1 - p_t)$ incur a fixed monetary loss L. She is covered by an insurance contract involving a fixed deductible D and a premium Q_t that depends on past experience. Specifically, the evolution of Q_t is governed by the following 'bonus-malus' system:

$$Q_{t+1} = \delta Q_t \text{ if no accident occurred in period } t$$

$$= \gamma Q_t \text{ if an accident occurred in period } t$$

where $\delta < 1 < \gamma$.[4]

The no-accident probability p_t is subject to moral hazard. Specifically, in each period t the agent chooses an effort level $e_t \geq 0$, resulting in a no accident probability $p_t = p(e_t)$ for some increasing, concave function p. The cost of effort is assumed separable, i.e., the agent attaches utility

$$u(x) - e$$

to income x if he exerts effort e, where u is increasing and strictly concave. The horizon is infinite and agents maximize expected discounted utility.

According to the bonus-malus scheme, each accident shifts the incentive scheme faced by the agent upward, thus modifying her incentives. It follows that the 'cost' of an accident, in terms of higher future premia, depends on random events (the sequence of future accidents) and endogenous decisions (the se-

4. Proportional bonus-malus schemes of this type are empirically frequent. The French system, which is relevant for our empirical application, corresponds to $\delta = .95$ and $\gamma = 1.25$. In addition, the French system imposes a floor and a ceiling on θ_t, respectively equal to .5 and 3.5. In our discussion of the French system, we ignore the fact that accidents occur continuously but premiums are only updated annually. See Abbring et al. (forthcoming) for a formal discussion.

quence of future efforts). Technically, the agent must solve a stochastic control problem. Here, we simply summarize the main properties of the solution; the reader is referred to Abbring et al. (forthcoming) for a precise analysis. A first result is that past experience matters for the current decision only through the current level of the premium; i.e., Q_t is the only state variable of the control problem. Secondly, the optimal effort is increasing in the premium, at least when both the premium and the deductible are small relative to the agent's income. It follows that the accident probability process of any given agent will exhibit a *negative occurrence-dependence property.* In the absence of an accident, the premium—hence, by our result, the agent's incentives—decreases. Effort is optimally reduced, resulting in a steady increase of the accident probability. However, the occurrence of an accident generates a discrete jump in the premium, which boosts incentives and ultimately results in a drop in the accident probability. The main testable implication of the model is thus the following:[5]

> *The accident process exhibits* negative *occurrence dependence, in the sense that individual claim intensities decrease with the number of past claims.*

This suggests that we can test for moral hazard by simply testing for negative occurrence dependence in the raw data. One should however be careful at this point. While moral hazard implies occurrence dependence effects at the individual level, individual claim intensities also vary with observed characteristics (such as age, driving experience, region, etc.) and, more importantly, with unobserved individual heterogeneity factors. In automobile insurance, for example, existing work strongly suggests that unobserved heterogeneity is paramount. It is well known that unobserved heterogeneity results in (spurious) *positive* occurrence dependence in the data. The intuition is that those individuals whose risk is persistently high for unobserved external reasons will be more likely to have had accidents in the past *and* to have accidents in the future (in other words, to the extent that 'bad' drivers remain bad for at least a while, we should expect to find a positive correlation between past and future accident rates). Of course, this effect, which is overwhelmingly confirmed by the data, does not contradict the theoretical analysis sketched above: whatever the distribution of unobserved heterogeneity, it is still true that under moral hazard, the accident probability *of each individual* decreases with *the person's* number of past claims. But any empirical investigation of this property must address the problem of disentangling the 'true,' negative dependence induced by the dynamics of incentives from the 'spurious,' positive contagion generated by unobserved heterogeneity.

5. An additional (and standard) difficulty comes from the fact that we observe claims, not accidents, and that the decision to file a claim is endogenous. See Chiappori and Salanié (1997) for a precise discussion of this problem.

This problem is a manifestation of a general question, namely distinguishing heterogeneity and state dependence. This issue has been abundantly discussed in the labor literature since the seminal contribution of Heckman and Borjas (1980). An interesting side aspect of our research, thus, is that it establishes a link between an existing literature in labor economics and questions that arose recently in applications of contract theory.

3. Testing for Moral Hazard

In most empirical studies in insurance, data are drawn from the files of one (or more) insurance companies. Many relevant characteristics of the driver (age, gender, place of residence, seniority, type of job) and the car (brand, model, vintage, power) are used by companies for pricing purposes. All these are available to the econometrician as well. The same is true for the characteristics of the contract (type of coverage, premium, deductible, . . .). Finally, each accident—or more precisely each claim—is recorded with all the relevant background information.

The main differences between data sets can be traced back to the way past history is recorded. Existing situations can be gathered in three broad cases:

- In the most favorable situation, the exact date of each accident is recorded. Then the occurrence of an accident can be modelled in continuous time, using event-history models.
- Many experience-rating schemes can be implemented with information on the number of accidents in each contract year only. In such cases, insurance companies will often only provide researchers with individual counts of claims over the years. In some cases, information on whether at least one accident has occurred or not in any year (rather than the exact number of accidents in each year) is sufficient. Then, for each agent we only observe a sequence of 0s (for years without accidents) and 1s (years with accidents).
- Finally, the minimum information that is needed to implement a bonus-malus scheme may be even poorer. If all past accidents are treated symmetrically whatever their exact timing (as in our theoretical model), the computation of a bonus-malus coefficient only requires information on the total number of past accidents. In our model, an agent who has been driving for t periods and has had n accidents will be charged a premium of $\gamma^n\delta^{t-n}$ times her initial premium, whatever the exact timing of each of the accidents. In this case, a single draw from an insurance company's files may only give a cross-section of total counts of accidents for a group of clients that has been driving for periods of varying length. Dynamics can only be studied by comparing across individuals of different (driving) seniority.

In each of these three cases, the dynamics of accidents can be used to test for the presence of moral hazard, against the null that the accident probability does not depend on the agent's incentives and only evolves according to some predetermined law (possibly depending on observables, such as age of the driver, age of the car, and others).

3.1 Heterogeneity Versus Moral Hazard in Continuous-Time Event-History Data

In the first case, the essence of the test is clear. Under moral hazard, the hazard rate of an accident, conditional on observable and unobservable heterogeneity, should be steadily increasing throughout any period without an accident and drop discontinuously whenever an accident occurs. Under the null, however, the hazard rate should not change after an accident. This can either be tested parametrically or non-parametrically. Denote the number of claims up to and including time t by $N(t)$ and let $X(t)$ be some vector of observable covariates (age, gender, etc.) at time t. Abbring et al. (forthcoming) assume that the intensity θ of claims, conditional of the claim history $\{N(u);\, 0 \leq u < t\}$ and the covariate history $\{X(u);\, 0 \leq u \leq t\}$ up to time t takes the form

$$\theta\left(t|\lambda, \{N(u);\, 0 \leq u < t\}, \{X(u);\, 0 \leq u \leq t\}\right) = \lambda \beta^{N(t-)} \psi(t) e^{X(t)'\gamma},$$

where ψ is a fully nonparametric baseline hazard function, $\beta > 0$ a scalar parameter, λ a nonnegative unobservable covariate reflecting unobserved heterogeneity, and γ a vector of parameters. Note that $N(t-)$ is the number of claims up to, but not including, time t. Thus, the parameter $\beta > 0$ captures true occurrence dependence effects. In the bonus-malus system described above, moral hazard leads to a decline in the intensity of claims with the number of previous claims ($\beta < 1$). Without moral hazard, we expect $\beta = 1$. Distinguishing these cases (testing), and estimating β, is the focus of the empirical analysis. Statistical tests are developed and applied to a French sample of 79,684 contracts, of which 4,831 have one claim in the contract year and 287 have two claims or more. The null ($\beta = 1$) cannot be rejected at any conventional level, suggesting that moral hazard is not a major problem in the data under consideration.

3.2 Testing for Moral Hazard from Sequences of Accident Counts

When only the total numbers of accidents by year are known, we can develop and apply similar methods for testing occurrence dependence in panel count data. Here, we focus on the more challenging case in which we only observe an annual sequence of 0s and 1s, corresponding to respectively years without and

years with at least one accident, for each agent. Econometric procedures for testing for occurrence dependence on such data have been developed by Heckman (1978, 1981a, 1981b), Honoré (1993), Kyriazidou (1997), and Honoré and Kyriazidou (2000). They rely on the assumption that each agent's accident probability remains constant throughout the observation period (stationarity).[6] To get the intuition in a simple way, assume the system is malus only (i.e., the premium increases after each accident, but does not decrease subsequently), and consider two sequences of 4 years, $A = (1, 0, 0, 0)$ and $B = (0, 0, 0, 1)$, where a 1 (resp. 0) denotes the occurrence of an accident (resp. no accident) during the corresponding year. In the absence of moral hazard, and assuming away learning phenomena, the probabilities of observing either of the two sequences should be exactly the same; in both cases, the observed accident frequency is 25 percent. Under moral hazard, however, the first sequence is more probable than the second: in A, the sequence of three years without accidents happens after an accident, hence when the premium, and consequently the marginal cost of future accidents and the incentives to take care are maximum. In other words, for a given average frequency of accidents, the timing of the occurrences can provide valuable information on the importance of incentives.

The test described here assumes stationarity. The analogy with the methods for continuous-time data of Abbring et al. (forthcoming) discussed earlier suggests that tests can be developed that are informative on moral hazard even if individual accident probabilities may change over time for external reasons. Richer panel-count data, that do not only record whether an accident has occurred at all but also how many accidents have occurred in any year, may be helpful here. This is on our research agenda.

3.3 Testing for Moral Hazard from Total Number of Accidents Only

Even in the case in which information is minimal—i.e., in which only the total number of past accidents is known for each agent in the insurer's database—it is still possible to test for moral hazard. In this case, we essentially have a cross-section of total accident counts over the periods that agents have been driving (seniority). Under additional stationarity assumptions, one can exploit the variation in seniority in the data set to test for moral hazard. Specifically, assume that (a) individual accident probabilities are constant over time, and (b) the distribution of unobserved heterogeneity in the population is identical across cohorts (seniority levels). That is, conditionally on observables, the no accident probability p is distributed among the drivers of any given seniority according to some distribution μ that is identical across seniorities. The idea of the test, as developed by Chiappori and Heckman (2000), is the following. Under the null of no moral hazard, for any driver with seniority t and no accident probability

6. In the duration model, this would correspond to a constant ψ.

p, the probability of having no accident throughout the observation period is p^t. Hence the proportion of drivers with no accident throughout the period is, under the null, equal to

$$m_t = \int p^t d\mu(p),$$

i.e., to the t-th moment of the distribution. It follows that the numbers m_1, $m_2, \cdots$ must, under the null, be the successive moments of the same distribution, which generates a first set of restrictions (see Heckman 1978 and in a different context Chiappori 1985). In addition, one can see that again under the null, the proportion, within the subpopulation of seniority t, of agents with exactly one accident is

$$m_t^1 = \int tp^{t-1}(1-p)d\mu(p)$$

$$= t(m_{t-1} - m_t).$$

This provides a set of simple, linear restrictions involving three statistics, namely m_{t-1}, m_t and m_t^1. Additional restrictions can be derived involving higher numbers of accidents. An analogous analysis for the moral hazard case is required to judge the power of tests based on these restrictions. These are topics for future research. Note that, in any case, these tests involve a comparison of *disjoint* subpopulations and heavily exploit stationarity assumptions.

3.4 Testing for Adverse Selection

In the three cases considered, the null (no moral hazard) is consistent with the presence of unobserved heterogeneity, whatever its type. Such heterogeneity reflects the impact of any variable that is *not* observed by the insurance company (and therefore the researcher), whether it is known by the insuree or not. In other words, one does not, under the null, distinguish between *adverse selection* and symmetrically imperfect information. Testing for adverse selection (and particularly asymmetric learning) requires analyzing the joint process followed by accidents and contractual choices.

In a dynamic setting, adverse selection can be modelled in various ways. One way is to assume that each agent is characterized by some constant parameter reflecting her 'quality' as a driver, which is known by the agent but not by the insurer at the beginning of the relationship. In this setting, adverse selection can be tested using the simple, cross-sectional approach described in the introduction. In the case of automobile insurance, most existing analyses fail to find positive conditional correlation, at least on populations of young drivers.

This suggests that adverse selection, if any, is not adequately described by the 'fixed quality parameter' story.

A more complex but also more convincing version relies on the asymmetric learning argument sketched in the introduction. There, adverse selection gradually emerges during the relationship. A natural empirical strategy is to study the causal relationship between the sequences of accidents and contract choices (or amendments). In particular, agents who learn their risk is above average are more likely to switch to a contract entailing a more comprehensive coverage. The previous (heterogeneity versus occurrence-dependence) perspective must then be extended to a two-dimensional process. Again, this will be the topic of future work.

References

Abbring, J. H., P. A. Chiappori, and J. Pinquet (forthcoming). "Moral Hazard and Dynamic Insurance Data," *Journal of the European Economic Association.*

Cardon, J. and I. Hendel (1998). "Asymmetric Information in Health Insurance: Evidence From the National Health Expenditure Survey," mimeo. Princeton, New Jersey: Princeton University.

Chiappori, P. A. (1985). "Distribution of Income and the Law of Demand," *Econometrica,* 53, pp. 109–127.

Chiappori, P. A. (2000). "Econometric Models of Insurance under Asymmetric Information," In *Handbook of Insurance,* edited by G. Dionne. Amsterdam: North Holland.

Chiappori, P. A., F. Durand, and P. Y. Geoffard (1998). "Moral Hazard and the Demand for Physician Services: First Lessons from a French Natural Experiment." *European Economic Review,* 42, pp. 499–511.

Chiappori, P. A., P. Y. Geoffard, and E. Kyriadizou (1998). "Cost of Time, Moral Hazard, and the Demand for Physician Services," mimeo. University of Chicago.

Chiappori, P. A. and J. Heckman (2000). "Testing for Moral Hazard on Dynamic Insurance Data: Theory and Econometric Tests," mimeo. University of Chicago.

Chiappori, P. A., B. Jullien, B. Salanié, and F. Salanié (2002). "Asymmetric Information in Insurance: Some Testable Implications," mimeo. University of Chicago.

Chiappori, P. A. and B. Salanié (1997). "Empirical Contract Theory: The Case of Insurance Data," *European Economic Review,* 41, pp. 943–951.

Chiappori, P. A. and B. Salanié (2000). "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy,* 108, pp. 56–78.

Chiappori, P. A. and B. Salanié (2003). "Testing Contract Theory: A Survey of Some Recent Work," in *Advances in Economics and Econometrics: Theory and Application, Eighth World Congress,* M. Dewatripont, L. Hansen, and P. Turnovsky, eds. *Econometric Society Monographs,* Cambridge University Press, Cambridge, pp. 115–149.

Dionne, G., C. Gouriéroux, and C. Vanasse (1997). "The Informational Content of Household Decisions, With an Application to Insurance under Adverse Selection." Working Paper, HEC, Montreal.

Dionne, G., M. Maurice, J. Pinquet, and C. Vanasse (2001). "The Role of Memory in Long-Term Contracting with Moral Hazard: Empirical Evidence in Automobile Insurance," mimeo., University of Montreal.

Dionne, G. and C. Vanasse (1992). "Automobile Insurance Ratemaking in the Presence of Asymmetrical Information." *Journal of Applied Econometrics,* 7, pp. 149–165.

Heckman, J. J. (1978). "Simple Statistical Models for Discrete Panel Data Developed and

Applied to Test the Hypothesis of True State Dependence Against the Hypothesis of Spurious State Dependence." *Annales de l'INSEE,* 30–31, pp. 227–269.

Heckman, J. J. (1981a). "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Panel Data with Econometric Applications,* edited by C. Manski and D. McFadden. Cambridge: MIT Press.

Heckman, J. J. (1981b). "Heterogeneity and State Dependence," in *Studies of Labor Markets,* edited by S. Rosen. Chicago: University of Chicago Press.

Heckman, J. J. and G. J. Borjas (1980). "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence." *Economica,* 47, pp. 247–283.

Hendel, I. and A. Lizzeri (1999). "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance," working paper, Princeton University.

Holly, A., L. Gardiol, G. Domenighetti, and B. Bisig (1998). "An Econometric Model of Health Care Utilization and Health Insurance in Switzerland." *European Economic Review,* 42, pp. 513–522.

Honoré, B. (1993). "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependant Variables," *Journal of Econometrics,* 59, pp. 35–61.

Honoré, B. and E. Kyriazidou (2000). "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica,* 68, pp. 839–874.

Kyriazidou, E. (1997). "Estimation of Panel Data Sample Selection Models," *Econometrica,* 65, pp. 1335–1364.

Mossin, J. (1968). "Aspects of Rational Insurance Purchasing," *Journal of Political Economy,* 76, pp. 553–568.

Poterba, J. and A. Finkelstein (2003). "Adverse Selection in Insurance Markets: Policyholder Evidence From the U.K. Annuity Market," mimeo., MIT.

Puelz, R. and A. Snow (1994). "Evidence on Adverse Selection: Equilibrium Signalling and Cross-Subsidization in the Insurance Market," *Journal of Political Economy,* 102, pp. 236–257.

Richaudeau, D. (1999). "Automobile Insurance Contracts and Risk of Accident: An Empirical Test Using French Individual Data." *The Geneva Papers on Risk and Insurance Theory,* 24, pp. 97–114.

MORAL HAZARD AND DYNAMIC INSURANCE DATA

Jaap H. Abbring
Free University Amsterdam

Jean Pinquet
Université Paris X-Nanterre

Pierre-André Chiappori
University of Chicago

Abstract
This paper exploits dynamic features of insurance contracts in the empirical analysis of moral hazard. We first show that experience rating implies negative occurrence dependence under moral hazard: individual claim intensities decrease with the number of past claims. We then show that dynamic insurance data allow to distinguish this moral-hazard effect from dynamic selection on unobservables. We develop nonparametric tests and estimate a flexible parametric model. We find no evidence of moral hazard in French car insurance. Our analysis contributes to a recent literature based on static data that has problems distinguishing between moral hazard and selection and dealing with dynamic features of actual insurance contracts. Methodologically, this paper builds on and extends the literature on state dependence and heterogeneity in event-history data. (JEL: D82, G22, C41, C14)

1. Introduction

Empirical tests of contract theory using insurance data have recently attracted much attention. Several papers test for the existence and estimate the magnitude of asymmetric-information effects in competitive insurance markets. Puelz and Snow (1994), Dionne and Vanasse (1992), Chiappori and Salanié (1997, 2000), Dionne, Gouriéroux, and Vanasse (1999, 2001), and Richaudeau (1999), to name only a few, analyze car-insurance contracts, whereas Holly, Gardiol, Domenighetti, and Bisig (1998), Chiappori, Durand, and Geoffard (1998),

Acknowledgments: We are indebted to Jim Heckman for numerous illuminating discussions. We also thank Ivar Ekeland, Bo Honoré, Bentley McLeod, Geert Ridder, Aad van der Vaart, Quang Vong, the editor (Richard Blundell), three anonymous referees, seminar participants at UCLA, Yale, Chicago, Northwestern, UCL, INSEE-CREST, Tinbergen Institute, CentER Tilburg, Toulouse, USC, Bristol and LSE, and participants in the SITE 2001 Summer Workshop and the Seventeenth Annual Congress of the European Economic Association for helpful comments and suggestions. Errors are ours. Jaap Abbring worked on this paper while visiting the University of Chicago, University College London and the Institute for Fiscal Studies. Abbring's research is supported by a fellowship of the Royal Netherlands Academy of Arts and Sciences. Support from the NSF (grant # 0096516) is gratefully acknowledged.

E-mail addresses: Abbring: jabbring@econ.vu.nl; Chiappori: pchiappo@midway.uchicago.edu; Pinquet: pinquet@u-paris10.fr

Chiappori, Geoffard, and Kyriadizou (1998), Cardon and Hendel (1998), and Hendel and Lizzeri (1999) use health or life insurance data, and Finkelstein and Poterba (2002) concentrate on annuities.

1.1 The "Conditional-Correlation" Approach

One popular empirical strategy focuses on the correlation between the contract choice and the occurrence of an accident, conditional on observables. This correlation is informative on asymmetric-information effects. Under adverse selection, for instance, agents know whether their accident probability exceeds the average in their risk class (as defined by the insurer in terms of the available information). If it does, they are both more likely to choose a contract with more complete coverage and more likely to have an accident, everything else equal. It follows that, conditional on observables, the choice of full insurance should coincide with a higher accident rate. This is a property that can be tested using parametric or nonparametric techniques.

The conditional-correlation approach has several advantages. It is simple and very robust, as argued by Chiappori and Salanié (1997) and Chiappori, Jullien, Salanié, and Salanié (2001). Furthermore, it can be used on static, cross-sectional data that are relatively easy to obtain. However, these qualities come at a cost. First, the effect of the past history of the relationship on the current contract is both difficult to model and difficult to estimate. Nevertheless, it can be of crucial importance, especially if experience rating plays a role. Second, the conditional-correlation approach may not allow to identify the type of information asymmetry involved (if any). Under adverse selection, accident-prone agents choose to buy more insurance. Moral hazard, on the other hand, suggests the opposite causality: agents who, for any reason, buy more insurance become more risky because the extensive coverage has a negative effect on incentives and discourages cautious behavior. To the extent that static data only allow to identify correlations, these two mechanisms cannot be distinguished.[1]

1.2 Adverse Selection, Moral Hazard, and the Dynamics of Asymmetric Information

The approach used in this paper relies on the idea that adverse selection and moral hazard can be distinguished by analyzing the dynamic aspects of the relationship (Chiappori 2000). This can be done in two different ways. One possible strategy compares the features of existing *contracts* to theoretical

1. Several articles try to empirically disentangle adverse selection and moral hazard. For instance, Holly et al. (1998) and Cardon and Hendel (1998) estimate structural models of health insurance, while Chiappori, Durand, and Geoffard (1998) and Dionne, Maurice, Pinquet, and Vanasse (2001) exploit "natural experiments" in which a new regulation exogenously changes incentives.

predictions about the form of optimal contracts under adverse selection and moral hazard. This approach exploits the fact that, in a dynamic setting, optimality has different implications in each case. Hence, a careful empirical investigation of the dynamic features of observed *contracts* may provide useful insights in the type of problem they are designed to address.[2] An alternative strategy, which we adopt throughout the present paper, does not assume optimality of existing contracts. Instead, it takes existing (and possibly suboptimal) contracts as given and contrasts the *behavior* implied by theory under adverse selection and moral hazard to observed behavior. The idea is that particular features of existing contracts, whether optimal or not, have different theoretical implications for observed behavior under adverse selection and moral hazard. Thus, the two can be distinguished by a careful analysis of observed behavior.

The two strategies just described have their own advantages and disadvantages. The first approach should in principle be very robust, to the extent that it relies on simple, qualitative characteristics of optimal contracts. Another virtue is that its implementation requires only data on contracts, which are in general much easier to obtain than data on induced behavior.[3] However, these qualities come at a cost. First, the qualitative characteristics of optimal dynamic contracts under asymmetric information may be very difficult to derive, except for very specific cases. They may moreover involve complex schemes, such as randomized contracts or sophisticated revelation mechanisms, which are hardly ever observed in real life.[4] A second concern is that optimal contracts are typically derived within a simplified framework (assuming for instance linear technologies, no loading, no transaction costs, etc.). The robustness of the corresponding conclusions in a more realistic and therefore more complex setting is not guaranteed. A particularly important issue is the presence of unobserved heterogeneity in individual characteristics, notably risk aversion. Arguably, any "realistic" model of optimal insurance contracts should not only consider the particular feature under study (moral hazard or adverse selection on risk), but should also take the paramount presence of adverse selection on preferences into account. This in general requires a characterization of optimal contracts under either adverse selection and moral hazard or multidimensional adverse selection. These are very difficult problems, especially in a dynamic context, and little is known about their solutions.[5] Finally, even casual empiricism indicates that actual insurance contracts are not always optimal (to say the least). For instance,

2. See Dionne and Doherty (1994) for an early example.

3. For instance, Hendel and Lizzeri (2001) find evidence of symmetric learning by comparing the actuarial value of life insurance contracts in which future premia may or may not vary with the agent's future health status. Interestingly, their analysis does *not* require data on actual mortality.

4. For example, Chiappori, Macho, Rey, and Salanié (1994) show that, under moral hazard, renegotiation-proof implementation of any effort level above the minimum is impossible without randomized contracts if savings are not observable. This is true except for the special case of monetary cost of effort and CARA preferences.

5. The reader is referred to Rochet and Stole (2000) for a survey of multidimensional adverse selection, and to Chassagnon and Chiappori (2000), Jullien, Salanié, and Salanié (1999), and

theory suggests that the characteristics of an optimal experience-rating scheme should be specific to each class of risk; that individuals, at least in the presence of adverse selection, should be offered a menu of various experience-rating schemes; and that not only the premium, but also the deductible (and more generally the whole nonlinear reimbursement profile) should depend on past experience. These features, however, are rarely observed in real life.[6]

For these reasons, we choose to adopt the second approach: we study the behavior induced by existing contracts without necessarily assuming that those contracts are optimal in any sense. Specifically, we exploit the fact that most real-life insurance contracts exhibit some form of experience rating (although not necessarily the optimal form predicted by theory). Under moral hazard, experience rating has very interesting implications. The occurrence of an accident affects the whole schedule of future premia. This changes not only the expected wealth of the agent and the expected average cost of insurance, but also, more importantly, the (expected) discounted marginal cost of future accidents. The cost of the next accident (in terms of, say, expected future premia or the corresponding certainty-equivalent) thus depends on the current premium and hence on the past accident history. It follows that the occurrence of an accident changes the incentives faced by a driver and therefore, under moral hazard, the future accident probability. This suggests that we can test for moral hazard by testing for such dynamics in the agent's accident process. A contribution of this paper is to point out the close link between this idea and a problem that has been studied at length in econometrics, the distinction between pure heterogeneity and state dependence.

1.3 Heterogeneity Versus State Dependence

The problem of distinguishing heterogeneity and state dependence originally appeared in economics in relation to unemployment and labor-supply issues (see Heckman and Borjas 1980, and Heckman 1981).[7] For example, it is well-known that individuals who are unemployed now are more likely to be unemployed in the future. There are two explanations for this empirical finding. One is that past unemployment has a direct, negative impact on the worker's future employment prospects (because of e.g. stigma effects, decreased investments in human capital, etc.). This is the state-dependence explanation, whereby unemployment

Araujo and Moreira (2002) for examples of models involving moral hazard and adverse selection. These papers, however, only consider a static setting.

6. From a more technical point of view, the optimality assumption also leads to difficult endogeneity issues. For instance, it is not possible, in general, to compare the performances of the different schemes that coexist on the market without taking into account the inherent selection bias: since each schedule is assumed optimal, the coexistence of different schemes must reflect differences in the corresponding populations. Such bias can be very difficult to correct for.

7. The statistical problem of distinguishing between spurious and true state dependence has a long history, with seminal contributions by Feller (1943) and Bates and Neyman (1952).

spells have a genuine effect on the agent's behavior in the sense that "an otherwise identical individual who did not experience unemployment would behave differently in the future than an individual who experienced unemployment" (Heckman and Borjas 1980, p. 247). Alternatively, the observed pattern may simply reflect the dynamic selection effects of unobserved heterogeneity. Assume that workers differ by some unobserved characteristics that affect their employment probability. Suppose that these characteristics are positively related over time and not affected by labor-market outcomes. Then, "less employable" workers are more likely to be unemployed both now and in the future. This results in a positive association between past and future unemployment, which is spurious in the sense that past unemployment matters only as a proxy for their unobservable employability (which, by assumption, is not affected by labor-market outcomes). The literature has clarified this distinction between state dependence and pure heterogeneity and has produced various methods for their empirical analysis. In particular, panel-data methods have been developed that exploit the fact that the two explanations have different implications for the dynamics of employment.

In this paper, we will more formally develop the idea that moral hazard leads to a particular form of state dependence in the accident process under experience rating.[8] We will then argue that testing for moral hazard boils down to testing for this true state dependence in the presence of unobserved heterogeneity. We will follow the labor literature and exploit that true state dependence, and therewith moral hazard, can be detected by analyzing the dynamics of, in our case, accidents.[9]

1.4 The French "Bonus-Malus" Scheme and State Dependence

We consider a scheme used by French insurance companies, the so-called "bonus-malus" mechanism. This mechanism is both simple and explicit, which considerably simplifies the empirical investigation. Contracts are renewed and premiums are revised annually. The premium is the product of two factors, the "base premium" and the "bonus-malus coefficient" at the time of contract renewal. The base premium is computed at the beginning of the relationship. It can be defined freely, but can only depend on observables and must be uniform over agents with identical characteristics. It cannot be modified during the relationship unless some observable characteristic changes, and only in a predefined way. Experience rating operates through the second component, the

8. Note that in the contract theory framework, the theoretical structure may moreover provide specific predictions on the direction of state dependence effects. In our context, for instance, the characteristics of the experience-rating scheme at stake imply that the occurrence an accident can only increase prevention efforts.

9. For a similar approach on labor data in a learning framework see Chiappori, Salanié, and Valentin (1999).

bonus-malus coefficient, on which we shall particularly concentrate in the paper.[10] An important feature of the French system is that the bonus-malus coefficient "sticks" to the agent, in the sense that an agent switching insurers will bring her old coefficient into the new contract. In particular, attrition cannot be explained by an attempt to "escape" the current coefficient. This will be important for the empirical analysis.

The evolution of the bonus-malus coefficient only depends on accidents for which the insuree is fully or partly responsible (accidents entirely caused by a third party are fully covered, in general by the third party's insurance and at no cost for the victim). Each year without such an accident (or, more appropriately, claim) decreases the coefficient by some fixed factor $\delta < 1$ (currently 0.95). Each accident caused by the insuree—there can be more than one in a contract year—increases the coefficient by a factor $\gamma > 1$ (currently 1.25).[11] It follows that any accident shifts the whole distribution of future (contingent) premia upwards, by a factor γ. Roughly, the "cost" of the $(n + 1)$-th accident is γ times larger than that of the n-th.

These properties will, in turn, affect the optimal effort profile. A natural conjecture is that the increased marginal cost results in more cautious behavior and smaller accident probabilities. This intuition, however, deserves more careful scrutiny, because of the complex nature of the problem. Several effects should be considered. For instance, the upward shift in the premium schedule decreases the agent's expected wealth and the resulting wealth effect can modify risk aversion in a way that may confound our results. Also, the "future cost" alluded to above is in fact a random variable. Its distribution depends not only on the risk characteristics of the agent, but also on the future effort profile. Conversely, the latter will depend on (the consequences of) current behavior. In other words, the determination of the optimal effort level in each period requires the solution of an optimal control problem. In Section 2, we carefully investigate this problem by developing a theoretical model of dynamic moral hazard under experience rating. We show that, under standard convexity assumptions on preferences and the prevention technology, the intuition above is correct if the cost of insurance (premium and deductible) is a small fraction of income. Everything else equal (i.e., controlling for heterogeneity), the optimal effort level should increase with the premium. This implies that, conditionally on the driver's characteristics, the dynamics of accidents should exhibit negative "occurrence dependence": the occurrence of an accident decreases the individ-

10. In the period covered by our data, the base premium could actually depend on the claim history as well. If anything, however, the base premium varied like the bonus-malus coefficient and amplified the experience-rating scheme.

11. In addition, there exists a cap and a floor of the bonus-malus coefficient (currently, 3.5 and 0.5). Also, insurees who have had the maximum bonus without a claim for at least three years receive a "malus-deductible": their next claim at fault does not trigger a malus but only loss of the malus-deductible. It is easy to see that this does not qualitatively affect our conclusions on moral hazard and occurrence dependence.

ual's probability of future accidents.[12] This is the true state-dependence mechanism in our insurance context. Note that this prediction relies on the presence of moral hazard.

As in the labor example above, this conclusion only holds conditionally. Unobserved heterogeneity introduces an opposite association in the raw data: good drivers, who pay lower premia, tend to have both a smaller number of past accidents and a smaller probability of future accidents. Therefore, as in the labor literature alluded to above, our main empirical task is to disentangle the effects of pure heterogeneity from those of the particular type of state dependence that is induced by the presence of moral hazard. Section 3 exploits these ideas in the empirical analysis of moral hazard using longitudinal insurance data. We specify a nonparametric econometric model of claim times that allows for both occurrence dependence and dynamic selection on unobservables. The model is specified in terms of the individual's accident (or, better, claim) rate, which is assumed to be proportional in occurrence-dependence and heterogeneity effects on the one hand and the effects of time on the other hand. Our analysis extends existing results of the state-dependence literature in various directions. We develop a model that allows for general nonstationarity (through proportional pure time effects) in the claim intensity. We propose new tests that correct for such nonstationarity, and discuss their relation to the existing literature. Finally, we analyze the identifiability of a special case of the model and present some estimation results. Section 4 concludes. Details are relegated to the Appendix.

2. Dynamic Moral Hazard under Experience Rating: Theory

2.1 The Model

We consider a dynamic version of an insurance model along the lines of Mossin (1968). Time is continuous on $[0, \tilde{T}]$, for some finite $\tilde{T} > 0$. The wealth of agent i at time t is denoted by $W_i(t)$ and accumulates as follows. At time 0, agent i is endowed with some initial wealth $W_i(0)$. Then, between t and $t + dt$ agent i receives some income flow $w_i(t)dt$ and chooses a consumption flow $C_i(t)dt$ (where $0 \leq t < \tilde{T}$).[13] In addition, the agent causes an accident with some probability $p_i(t)dt$.[14] If so, she incurs some monetary loss, which is covered by an insurance contract involving a fixed deductible D_i and a premium $q_i(t)dt$ that is paid continuously.

The premium depends on past experience. In particular, it satisfies the

12. This type of state dependence is labelled "structural occurrence dependence" by Heckman and Borjas (1980).

13. We assume that the income path is integrable on $[0, \tilde{T}]$.

14. Accidents that are not caused by the agent are fully covered and have no impact on future premiums. Such accidents can be and are disregarded in our analysis.

following "bonus-malus" system. If agent i does not cause an accident between t and $t + dt$, the premium is decreased by an amount $dq_i = \delta q_i(t)dt$ (the "bonus"). If she causes an accident, on the other hand, the premium jumps discontinuously to $\gamma q_i(t)$. Here, $\gamma - 1 > 0$ is the proportional "malus."

Accidents caused by the agent are subject to moral hazard. That is, at each time t, the agent chooses the intensity $p_i(t)$ of having an accident from some bounded interval $[0, \bar{p}_i]$, at a utility cost $\Gamma_i(p_i(t))$. We assume that Γ_i is twice differentiable on $(0, \bar{p}_i)$, with $\Gamma_i' < 0$ and $\Gamma_i'' > 0$. In words, reducing accident rates is costly and returns to prevention are decreasing. For definiteness, we also assume that $\lim_{p \uparrow \bar{p}_i} \Gamma_i'(p) = 0$.

The agent's instantaneous utility from consuming $C_i(t)$ and driving with accident intensity $p_i(t)$ at time t is $u_i(C_i(t)) - \Gamma_i(p_i(t))$. We assume that u_i is increasing and strictly concave, so that the agent is risk-averse. The agent chooses consumption and accident-intensity plans that maximize total expected utility

$$\mathbb{E}\left[\int_0^{\tilde{T}} (u_i(C_i(t)) - \Gamma_i(p_i(t)))dt\right]$$

subject to some final wealth constraint, given the wealth and premium dynamics described above. For simplicity, we assume that the agent perfectly foresees her income path $\{w_i(t);\ 0 \leq t \leq \tilde{T}\}$. Thus, she only has to form expectations on future accidents and their implications.

2.2 Results

For notational convenience, we now drop the index i. It should be clear, however, that all results are valid at the individual level, irrespective of the distribution of preferences and technologies across agents. In particular, the results hold for any type of unobserved heterogeneity in these primitives of the model.

A first result is

LEMMA 1. *At each time t, the optimal consumption and accident intensities only depend on the past history through the agent's wealth* $W(t)$ *and the premium* $q(t)$.

Lemma 1 states that the only channel through which past accidents influence current behavior is their impact on the incentives faced by the agent, for which the current premium is a sufficient statistic, and on wealth. We have disregarded alternative channels, such as learning, fear, or cautionary reaction to an accident by assuming that the prevention technology, as represented by the cost function

Γ, does not depend on the past experience of accidents directly. The empirical relevance of this assumption will be discussed in Section 3.

The agent faces an optimal control problem. The value function V for this problem satisfies the Bellman equation

$$V(t, W, q) = \max_{C,p}\{u(C)dt - \Gamma(p)dt + pdtV(t + dt, W + w(t)dt - D - Cdt - qdt, \gamma q) + (1 - pdt)V(t + dt, W + w(t)dt - Cdt - qdt, q - \delta qdt)\}.$$

In words, between t and $t + dt$ the agent derives utility from her consumption and disutility from her prevention effort. If no accident occurs (with probability $1 - pdt$), her wealth is increased by the income flow minus consumption and the premium, and the premium is continuously reduced. If the agent causes an accident (with probability pdt) then she must in addition pay the deductible, and the premium jumps to γq. The Bellman equation can be rewritten as

$$0 = \max_{C,p}\{u(C) - \Gamma(p) - p[V(t, W, q) - V(t, W - D, \gamma q)] + V_t(t, W, q) + V_W(t, W, q)[w(t) - C - q] - V_q(t, W, q)\delta q\}, \quad \text{(B)}$$

with V_t, V_W, and V_q the partial derivatives of V with respect to t, W, and q, respectively.

We are interested in the qualitative properties of the value function V and the optimal accident intensity. First note that agents dislike high premiums. Take some premiums q and $q' < q$. At any moment t, an agent who faces q' could derive higher utility than under the higher premium q by simply using the strategy that would be optimal under q. Using the optimal strategy at q' instead can only further improve the gain. Hence,

LEMMA 2. *The value function V is decreasing in the premium q.*

We now concentrate on the impact of the current premium on the optimal accident intensity. First note that the agent's behavior for "large" premium levels may be atypical. Given the dynamics described previously, the premium could exceed the agent's current income, current wealth, or even lifetime wealth if the number of accidents is large enough. This situation, however, will not arise because the agent can always choose a zero accident probability (although presumably at a high cost), say by giving up driving. Note that in that case, the accident probability becomes totally inelastic to the premium.

In most cases, however, both the premium and the deductible are "small" relative to income.[15] We now show that for such "small" values of the premium and the deductible, the prevention effort is increasing (the accident intensity is

15. Both premiums and deductibles are a few hundred dollars in our sample, which is one or two percent of the median household income in the population under consideration.

decreasing) in the premium, as intuition suggests. To this end, consider the first order conditions of the program (B),

$$u'(C^*(t, W, q)) = V_W(t, W, q)$$
$$-\Gamma'(p^*(t, W, q)) = V(t, W, q) - V(t, W - D, \gamma q),$$

where $p^*(t, W, q)$ and $C^*(t, W, q)$ are, respectively, the optimal accident and consumption intensities at time t, wealth W, and premium q. The first equation is the standard Euler condition for intertemporal optimality. The second condition implies that

$$\Gamma''(p^*(t, W, q))\, p_q^*(t, W, q) = \gamma V_q(t, W - D, \gamma q) - V_q(t, W, q),$$

where p_q^* is the partial derivative of p^* with respect to q. For "small" values of q and D, this condition becomes

$$p_q^*(t, W, q) \approx \frac{\gamma - 1}{\Gamma''(p^*(t, W, q))} V_q(t, W, 0) < 0$$

and the accident intensity decreases with the premium.

As a consequence, the intensity of the accident process will drop discontinuously at the time of an accident, in response to a discontinuous jump in the premium. We formally state this conclusion as

PROPOSITION 1. *For small enough values of the premium and the deductible, the optimal accident intensity drops discontinuously at the time of an accident.*

Proposition 1 provides a simple testable implication of moral hazard under experience rating: under moral hazard, the occurrence of an accident results in a discontinuous drop in the accident intensity. In other words, the accident process should exhibit *negative occurrence dependence.* The rest of the paper is devoted to empirical tests of this prediction.

It is important to stress again that the theoretical analysis in this section operates at the individual level. For any given agent i the dynamics of accidents should exhibit the type of state dependence just described, irrespective of the distribution of preferences and technologies across agents. However, empirical tests of occurrence dependence have to rely on interindividual comparisons, if only to control for time-effects that are common across individuals. Unobserved heterogeneity thus becomes a critical issue in the empirical analysis.

The theoretical model provides a simplified representation of actual experience-rating schemes and the agent's behavior under these schemes. We already mentioned the fact that we ignore the nonmonetary consequences of an accident. A more technical issue is that the theory assumes that the premium is adjusted continuously, whereas premiums are typically only adjusted at discrete dates (e.g., annually) in actual experience-rating systems. Finally, real-life

schemes often entail ceilings and floors on the premium.[16] Our view is that the previous model nevertheless provides a useful approximation of actual behavior.

3. Econometric Specification and Empirical Analysis

3.1 Introduction

We now turn to the main problem of this paper, the empirical distinction of moral hazard on the one hand and adverse selection or, more generally, selection induced by unobserved heterogeneity on the other hand. As argued in the introduction, we propose to exploit dynamic data on claims under experience-rated contracts. Proposition 1 suggests a simple, direct test on negative occurrence dependence in observed claim rates. However, we can only directly control for observed individual characteristics and the occurrence-dependence effects due to moral hazard are likely to be confounded with the effects of dynamic selection on unobservable characteristics. In particular, insurees with a history of many accidents are likely to be more accident-prone for unobserved reasons. This leads to a positive effect of past claims on current claim probabilities that counters the negative effect of any moral hazard. In other words, the problem of distinguishing between moral hazard and selection is similar to a standard problem of distinguishing state (occurrence) dependence and unobserved heterogeneity.

The solution to this problem depends primarily on the type of data available. In many empirical studies in insurance (including ours), data are derived from the insurance companies' administrative files. Many relevant characteristics of the driver (age, gender, place of residence, type of job, etcetera) and the car (brand, model, vintage, power, etcetera) are used by companies for pricing purposes. These data are typically available to the econometrician as well. The same is true for the characteristics of the contract (type of coverage, premium, deductible, etc.). Finally, each accident—or more precisely each claim—is recorded with all the relevant background information.

Data sets mainly differ in the way the past claim history of insurees is recorded and made available to researchers. Many experience-rating schemes (including the French) can be implemented with data on the number of claims in each contract year only. Often this results in only (panel) counts of claims being provided to econometricians, without any information on the exact dates

16. In France, for instance, the bonus-malus coefficient cannot fall below 0.5 or increase above 3.5. Given the actual values of γ (1.25) and δ (0.95 annually), however, the 0.5 level cannot be reached within thirteen years. The 3.5 coefficient can be reached more quickly, but requires at least six accidents (and more if the agent receives bonuses for claim-free years before having six accidents). Therefore, given that drivers have on average one accident every seven years, reaching the ceiling quickly is a very rare event. Indeed, we do not find any driver in our data who is at the ceiling.

of the accidents. Some schemes can even be implemented if only the total number of claims in a given period is known. This gives rise to (cross-sectional) claim-count data. The empirical distinction of state-dependence and heterogeneity with such (panel or cross-sectional) count data is a difficult problem, but one that has been studied in the literature. However, we leave application and extension of this literature to our insurance problem for future work.

In this paper, we study the more favorable situation in which the exact date of each claim is provided. This suggests specifying a continuous-time event-history model of claims that allows for occurrence dependence and unobserved heterogeneity. The theory of Section 2 shows that moral hazard leads to negative dependence of individual claim intensities on the occurrence of previous claims. Unobserved heterogeneity in claim intensities captures any dynamic selection effects. In general, the occurrence-dependence effects of moral hazard will be heterogeneous in the population. To accommodate this, we allow for general interactions between occurrence dependence and individual-specific effects in most of our analyses. We will derive some extra, stronger results for an econometric model in which occurrence dependence acts proportionally on the claim intensity and is homogeneous across agents.

Three caveats should be mentioned. First, premiums are only updated annually and not continuously as in the theoretical model. Under discounting, this may introduce nonstationarity in the agent's decision problem beyond that implied by the finite horizon in the theoretical model: when contract renewal is near, the "cost" of an accident is higher than when premiums have just been updated. This effect seems to be of only minor importance. It does however suggest that we should allow for nonstationarity in "contract time" (i.e., time since the last premium update).[17] This is particularly true because contract-time effects may bias our assessment of occurrence dependence in arbitrary ways.

Second, we observe only claims, not accidents, and the decision to file a claim is endogenous. It is well known that this introduces a second type of moral hazard, which is often referred to as "ex post" in the insurance literature. For instance, experience rating results in more cautious driving ex ante and increased reluctance to file a claim for a minor accident ex post.[18] Neither the

17. In the French system, claims enter the bonus-malus coefficient with a delay of two months. More precisely, the history considered in determining the new bonus-malus coefficient at any particular contract renewal date consists of all claims corresponding to losses incurred during the year that has ended two months before the renewal date. For example, a new premium that is issued on January 1, 1989 is based on the history used in writing the old (January 1, 1988) contract and all claims in the period November 1, 1987–October 31, 1988. One could say that contract time is lagging claim-history time by two months. Clearly, theory predicts nonstationarity in claim-history time rather than in contract time. However, because the difference between the two is common across all agents, we can simply control for claim-history time by flexibly controlling for contract time.

18. A possible solution, used by Chiappori and Salanié (2000) in their cross-sectional analysis, is to consider exclusively accidents in which a second party was involved (in which case a claim is almost automatically filed). However, this solution has two drawbacks in the present context: it would decrease the number of accidents (and especially the number of cases in which two accidents

theory nor the empirical analysis distinguishes between these two types of moral hazard (in a sense, the prevention technology in the theoretical model can be seen as a reduced form for both). It is not clear, actually, that ex-ante and ex-post moral hazard should be distinguished at all (typically they should not be from the insurer's perspective). Anyhow, since both effects go in the same direction, the presence of ex-post moral hazard (if any) can only bias our estimate of ex-ante moral hazard upwards. Since we fail to find any significant effect at all, we suspect that both effects are negligible.

Third, there may be learning effects. In the theory section, we assume that past accidents only affect current behavior through a monetary channel (i.e., increased monetary incentives to exercise caution). In reality, there may be other channels as well. In particular, a young driver is presumably not perfectly informed about her driving ability and may learn about this ability from accidents. In the presence of moral hazard, such learning may in turn affect the probability of future accidents. For example, an upward reassessment of the accident probability in the absence of effort may enhance the perceived benefits of cautious driving. We address this problem in two different ways. First, even though accidents in which the driver is at fault typically have both incentive and learning effects, accidents that are entirely caused by a third party can only have learning effects and have no monetary-incentive effects. We exploit this fact by repeating our tests on a sample including all accidents, rather than just accidents at fault. Second, learning should be more important for young drivers than for experienced drivers. Thus, we can test for the relevance of learning by repeating our analysis on subsamples of inexperienced and experienced drivers and contrasting the results. These additional analyses, which are discussed in detail in Subsection 3.6, suggest that our conclusions are robust to learning effects.

The remainder of the paper will be concerned with the specification, identification and empirical analysis of an appropriate continuous-time model of insurance claims. We have seen that, within such a framework, a test on moral hazard boils down to a test with a null of no (genuine) occurrence dependence against the alternative of (genuine, negative) occurrence dependence. Our analysis will build on and extend the existing literature on state dependence and heterogeneity in continuous-time event-history models.

It should be stressed from the outset that the null of no moral hazard is consistent with the presence of unobserved heterogeneity, whatever its type. Such heterogeneity may reflect the impact of any information that is not available to the insurance company, but may or may not be known by the insuree himself. In other words, we do not, under the null, distinguish between adverse selection and symmetrically imperfect information.[19] It is important to

---

are observed) and it would lead to ignoring events (i.e., one-person accidents) that nevertheless influence incentives through their impact on the premium.

19. Moreover, in most of our analysis we do not control for observed heterogeneity (see the next subsection). This observed heterogeneity will be absorbed by the individual-specific effect.

note, however, that testing for adverse selection is certainly possible in this context, and can provide interesting insights in the nature of learning processes. The idea would be to analyze the changes in insurance contracts initiated by the drivers, and the subsequent impact on the accident hazards.[20] This is left for future work.

3.2 The Econometric Model

Our analysis focuses on the occurrence of car insurance claims in a single insurance contract year, i.e., the period bounded by two consecutive contract renewal dates. We first present a model for the population of claim histories in the contract year.

Let time have its origin at the start of the contract year. Then, if the contract year is of length T, it can be represented by the interval $[0, T]$. Let T_k be the time of the k-th claim in the contract year. Denote the corresponding counting process by $N[0, T] := \{N(t); 0 \leq t \leq T\}$, where $N(t) := \#\{k : T_k \leq t\}$ counts the number of claims in the contract year up to time t. $N[0, T]$ is the focus of our model and empirical analysis.

The intensity θ of claims at time t, conditional on the claim history $N[0, t) := \{N(u); 0 \leq u < t\}$ up to time t and a nonnegative individual-specific effect λ is

$$\theta(t|\lambda, N[0, t)) = \lambda\beta(\lambda)^{N(t-)}\psi(t), \tag{M}$$

with $\beta : [0, \infty) \to (0, \infty)$ a bounded measurable function and $\psi : [0, T] \to (0, \infty)$ a continuous function that captures contract-time effects.[21] We frequently use the notation $\Psi(t) := \int_0^t \psi(u)du$. We normalize $\Psi(T) = 1$, so that λ captures the scale of θ. We assume that λ has marginal distribution G. Together with equation (M), this fully specifies the distribution of $N[0, T]$.

Equation (M) gives the *individual* claim intensity at time t of an insuree with characteristics λ and claim history $N[0, t)$. Any nontrivial dependence of

Therefore, any "unobserved" heterogeneity found may also reflect symmetrically *observed* information.

20. Chiappori and Salanié (2000) find no evidence of adverse selection on a sample of "young" (i.e., recent) drivers. They note, however, that adverse selection may also arise during the relationship, due to asymmetries in learning between the firm and the client (e.g., such an informative event as a near-miss is typically observed by the driver only). For a theoretical investigation, see de Garidel (1997).

21. The model only recognizes contract time and does not explicitly consider the effects of calendar time (or duration since last event for that matter). In a typical sample, different contracts have different renewal dates, so that contract time and calendar time do not coincide (see the next section). Nonstationarity in contract time arises in theory because of discounting and the discrete-time nature of contract renewal, and possibly because of learning. Also, of all time effects, contract-time effects are most likely to confound our analysis of occurrence dependence, which concerns previous occurrence of claims in the contract year. We therefore want to deal with contract-time effects in a flexible manner.

this claim intensity on the insuree's claim history $N[0, t)$ can be interpreted as true state dependence. In our model (M), this state dependence takes the form of occurrence dependence: individual claim intensities at time t only depend on the claim history $N[0, t)$ through the number of past claims $N(t-)$.[22] The function β captures these occurrence-dependence effects. For an individual with characteristics λ, the claim intensity is multiplied by a factor $\beta(\lambda)$ each time a claim occurs. We allow this effect to be different across insurees with different characteristics by allowing β to be a nontrivial function of λ.

The claim intensity in (M) is multiplicative in individual heterogeneity and occurrence-dependence effects on the one hand and time-effects on the other hand. In this sense it is a proportional-hazards model. It should however be stressed that our model does not involve *observed* individual characteristics and that our main analyses do not use covariate information. Any covariate effects are subsumed in the (unobserved) individual effect λ. This individual effect can then also capture the cumulated effects of claims that have occurred before the sampled contract year started. In Subsection 3.6, we repeat our analyses on samples that are stratified on covariates. Through stratification we allow for general interactions of the covariates on the one hand and λ and the other model components on the other hand. Either way, we avoid initial-conditions problems.[23]

As explained in the Section 2, in the French car-insurance system moral hazard leads to a decline in the claim intensity with the number of previous claims: $\beta(\lambda) < 1$. Without moral hazard, we expect that $\beta(\lambda) = 1$ for all λ. Our empirical analysis focuses on distinguishing these two cases, in two stages.

First, in Subsection 3.4 we focus on testing the prediction expressed by Proposition 1 without further assumptions on preferences and technologies. In terms of the econometric model above, this amounts to testing the null hypothesis that $\beta(\lambda) = 1$ for all λ against the general moral-hazard alternative that $\beta(\lambda) < 1$ for all λ. We first state a basic result that is useful in the development of such tests: Ψ and G are identified under the null of no moral hazard. We then provide two nonparametric tests. The first of these tests, developed in Subsection 3.4.2, is rooted in the work of Bates and Neyman (1952) and Heckman and Borjas (1980, Section II.a) for exponential models with general unobserved heterogeneity. It exploits that, under the null, the total number of claims in a given (data) period is a sufficient statistic for the unobserved heterogeneity in the claim intensities. Our contribution is to provide a closely related test that allows for general nonstationarity in the claim intensities (by not imposing any restrictions on Ψ). In Subsection 3.4.3 we develop an alternative test that is inspired by the regression approach to event-history analysis (see, e.g., Heck-

22. Heckman and Borjas (1980) call this "structural occurrence dependence."

23. Rather than stratifying on covariates, we could extend (M) to include proportional covariate effects. In a random-effects setting, we would typically assume that λ is independent of these covariates. However, such independence will not hold if λ captures the cumulated effects of claims that have occurred before the sampled contract year.

man and Borjas 1980, Section II.b) and the methods for paired duration data developed by Holt and Prentice (1974) and Chamberlain (1985). This test uses within-individual variation in durations between claims to control for unobserved heterogeneity. The main distinguishing feature of our test is again that it allows for general nonstationarity.

Second, in Subsection 3.5 we concentrate on a particular version of the model in which the occurrence-dependence effects $\beta(\lambda)$ are the same across insurees. More formally, β is a trivial function of the individual-specific effect λ. We provide some new identification results for this model, and point out a relation to the literature on the identifiability of the two-sample mixed proportional hazard (MPH) model (Elbers and Ridder 1982, and Kortram et al. 1995). Finally, we provide some parametric estimates of β, Ψ, and G.

3.3 Sampling and Data

We observe the claim histories for all insurance contracts at a French insurance company in a given and common calendar time period of two years, October 1, 1987–September 30, 1989. Different contracts have different renewal dates, so that contract time and calendar time do not coincide.

The length of the contract year is, appropriately, one year (i.e., $T = 1$, with time measured in years).[24] Ideally, contracts cannot be terminated within a contract year, except in special circumstances. An example is death of the agent. Contracts can however be changed during the year. For example, the deductible can be altered or the contract can be transferred to a new car, which may be in a different risk class. In our data set, each change of contract triggers creation of a new record. Records have to be consolidated back into single contracts. Fortunately, high quality information is available to facilitate such linking. However, some linking problems may remain and generate some spurious attrition. Overall, 9.4 percent of the contracts written or renewed in the first sample year fail to survive a full contract year after that.[25] In this paper, we ignore both true and spurious attrition.

With that qualification, we observe at least one full contract year for each contract that is not terminated during the first sample year or written anew during the second. For contracts with renewal dates coinciding with the start date of the sample, we observe two full contract years if the contract is renewed after one year. Our observations are the flow of contracts that are written or renewed in the first sample year, and each observation provides information on

24. More precisely, given the presence of a leap year in the sample period, we take it to be 365 days.

25. More precisely, we compute the rate of attrition as a fraction of all the contracts that are active at their renewal date in the first sample year, and survive for at least thirty-one days after that. The latter condition excludes contracts that are terminated at their renewal date in the first sample year, with a thirty-one-day grace period.

TABLE 1. MAXIMUM-LIKELIHOOD ESTIMATES (DISCRETE HETEROGENEITY; TWELVE TIME INTERVALS ψ)

Occurrence dependence	
β	0.974 (0.677)
Unobserved heterogeneity	
λ^a	0.052 (0.006)
λ^b	0.310 (0.427)
Pr $(\lambda = \lambda^a)$	0.939 (0.104)
Pr $(\lambda = \lambda^b)$	0.061 (0.104)
Contract time (piecewise-constant ψ)	
Wald statistic $\psi \equiv 1$	15.617
Degrees of freedom	11
p-value	0.156
Number of observations by number of claims	
$M_{0,n}$ (no claims)	74,566
$M_{1,n}$ (1 claim)	4,831
$M_{2,n}$ (2 claims)	270
$M_{3,n}$ (3 claims)	15
$M_{4,n}$ (4 claims)	2
Log-likelihood	−3,536.16

$N[0, T]$ for the contract year following the renewal or writing of the contract.[26] Consistently with the prior discussion, contracts suffering from attrition during the contract year are discarded. For contracts with two contract years in the sampling period, the second year is discarded.

Let the resulting random sample contain n contracts, labelled $1, \ldots, n$. The i-th observation in the sample is denoted by $N_i[0, T]$, $i = 1, \ldots, n$. Each claim history $N_i[0, T]$ in the sample can alternatively be characterized as the number of claims $N_i(T)$ with, if $N_i(T) > 0$, a vector $(T_{1,i}, \ldots, T_{N_i(T),i})$ of claim times. Each observation $N_i(0, T)$ is complemented with an unobserved effect λ_i that has distribution G. The claim intensity for observation i is assumed to be given by (M) evaluated at sample variables.

The bottom panel of Table 1 provides some information on the sample. Accidents are fairly rare, but the number of contracts n for which we have at least one full contract year is large, 79,684. Of these contracts, 4,831 have one claim in the contract year, 270 have two claims, 15 have three claims, and 2 have four claims. No contracts have more than four claims.

One additional subtlety should be discussed here. The data distinguish between various types of claims. Two types of claims are labeled to be at fault, either full (inducing a 25 percent premium increase) or partial (12.5 percent),

26. Recall from Footnote 17 that claims in the last two months of the contract year $[0, T]$ do not affect the new premium that will be issued at T, but only the premium that may be issued one year later, at $2T$. We sample contract years to avoid attrition problems.

and are directly relevant to our analysis of moral hazard. We will not distinguish between these two types of claims and treat each claim, either partially or fully at fault, to be an event counted by $N(t)$. The sensitivity of the results with respect to these and other choices is investigated by repeating the analyses on other samples in Subsection 3.6.

More generally, note that we only model part of the relevant events at this point. We do not deal with changes of contract as described above, and we ignore claims that are not at fault (and, sometimes, claims at partial fault). To some extent, we could deal with these events by including them as time-varying covariates. However, because of the endogenous nature of, for example, changes in contract, we plan to pursue a more structural approach, leading to a richer event-history specification. This is left for future research.

3.4 Testing for Moral Hazard

3.4.1 Identification and Estimation of Ψ under the Null of No Moral Hazard. We will first show that the contract-time function Ψ can be identified and estimated under the null hypothesis that $\beta(\lambda) = 1$ for all λ (which, in the sequel, we simply denote by $\beta = 1$). This result will be useful later. As a by-product, we will be able to discuss and apply a well-known test for occurrence dependence due to Bates and Neyman (1952) and Heckman and Borjas (1980). Our first moral-hazard test, which will be discussed in next subsection, is based on this test. Appendix A.1 provides details that are omitted from this subsection.

Let H_1 be the distribution of the first claim time T_1 in the subpopulation with exactly one claim in the contract year ($N(T) = 1$):

$$H_1(t) = \Pr(T_1 \leq t | N(T) = 1).$$

Clearly, H_1 is identified from the distribution of the claim history $N[0, T]$ and can be estimated consistently by the empirical analog of H_1,

$$\hat{H}_{1,n}(t) = \frac{1}{M_{1,n}} \sum_{i=1}^{n} I(T_{1,i} \leq t,\ N_i(T) = 1).$$

Here, $M_{k,n} := \sum_{i=1}^{n} I(N_i(T) = k)$ more in general denotes the number of contracts in the sample with exactly k claims. It is easy to show that

$$H_1(t) = \frac{\Psi(t)}{\Psi(T)} = \Psi(t) \qquad \text{(H1)}$$

under the null hypothesis that $\beta = 1$. Note in particular that (H1) does not involve the distribution G of the individual-specific effect λ. This reflects the fact noted earlier that $N(T)$ is sufficient for λ under the null. It follows that Ψ

is identified directly from H_1 and can be consistently estimated by $\hat{H}_{1,n}$ under the null.

Note that identification of G under the null easily follows. The probability $q_0(t) := \Pr(N(t) = 0)$ of observing no claims up to time t is given by

$$q_0(t) = \mathcal{L}(\Psi(t)),$$

where $\mathcal{L}(s) := \int \exp(-\lambda s) dG(\lambda)$ is the Laplace transform of G. We have just seen that Ψ is identified from H_1 under the null. This implies that $\mathcal{L}$ is identified on [0, 1] from q_0 and H_1 under the null. As a consequence, $\mathcal{L}$ and G are identified under the null.[27]

So far, we have focused on the identification of Ψ from H_1 under the null hypothesis that $\beta = 1$. If we would know Ψ, we could turn the argument around and test the null hypothesis that $\beta = 1$ by testing the equality $H_1 = \Psi$. It is easy to show that

$$H_1(t|\lambda) := \Pr(T_1 \leq t|\lambda, N(T) = 1) = \begin{cases} > \Psi(t) & \text{if } \beta(\lambda) < 1 \text{ and} \\ < \Psi(t) & \text{if } \beta(\lambda) > 1 \end{cases} \qquad (\text{H1}^+)$$

for all $\lambda > 0$ and $t \in (0, T)$. Thus, a test based on the difference between $\hat{H}_{1,n}$ and Ψ could be designed to have power against the alternatives of moral hazard ($\beta(\lambda) < 1$ for all λ, or simply $\beta < 1$) and, more generally, occurrence dependence ($\beta(\lambda) \neq 1$ for some λ, or just $\beta \neq 1$).

Obviously, such a test would not be feasible because we do not know Ψ. In the literature, feasible tests have been developed under the additional assumption of stationarity (Bates and Neyman 1952). We say that the model in (M) is "stationary" if $\psi(t)$ is constant over time t ($\psi = T^{-1}$). The assumption of stationarity simply pins down $\Psi(t)$ to be t/T, so that we can test the null hypothesis of no moral hazard by testing for uniformity of H_1. In this paper, we do not want to impose stationarity to facilitate a test on moral hazard. We will however present a uniformity test of H_1 and follow Heckman and Borjas (1980) by interpreting this as a test of the joint null hypothesis of stationarity and no moral hazard.

Standard distributional test statistics can be computed from $\hat{H}_{1,n}$. We first investigate $\hat{H}_{1,n}$ graphically. The top panel of Figure 1 plots $\hat{H}_{1,n}$ and the uniform distribution function, together with some other functions that are only of later concern. The bottom panel graphs a kernel estimate of the density of H_1, using an Epanechnikov kernel with bandwidth 0.05. All analyses are based on the data described in the bottom panel of Table 1 and include both claims at full

27. This follows successively from the real analyticity and the uniqueness of the Laplace transform (see, e.g., Widder 1946). Nonparametric estimation of G under the null is relatively hard and will not be pursued here. The Laplace transform $\mathcal{L}$ on [0, 1] can be directly estimated from $\hat{H}_{1,n}$ and the empirical analog of q_0 on [0, T]. However, nonparametric estimation of $\mathcal{L}$ on $(1, \infty)$ and of G would somehow involve analytic extension and deconvolution, respectively, both of which are relatively hard to implement empirically. A computationally feasible estimator can possibly be developed along the lines of the method-of-moments estimator discussed in Heckman, Robb, and Walker (1990) and in papers referenced therein.

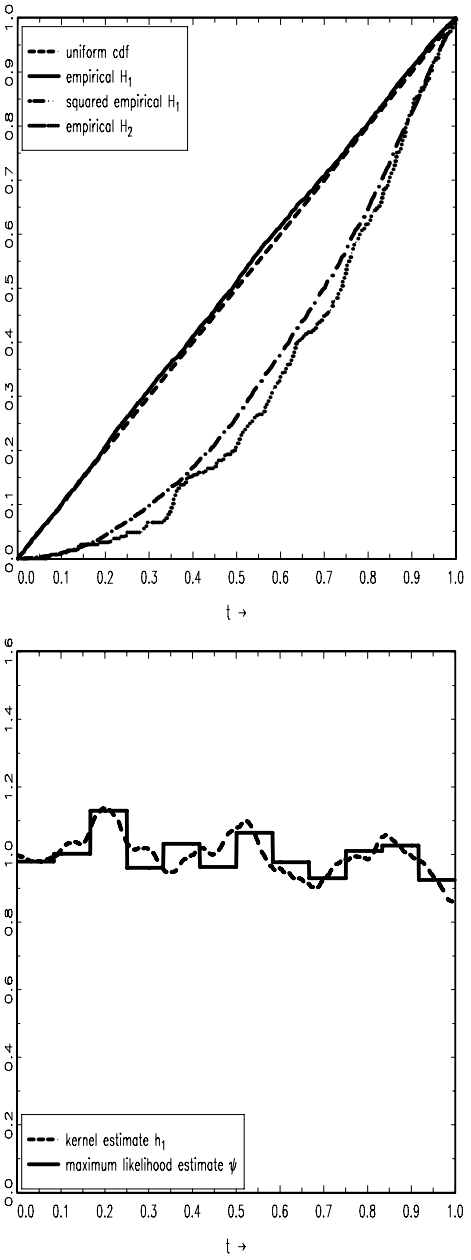


FIGURE 1. Empirical Distribution Function $\hat{H}_{1,n}$ of $T_1|(N(1) = 1)$, $\hat{H}^2_{1,n}$, and Empirical Distribution Function $\hat{H}_{2,n}$ of $T_2|(N(1) = 2)$ (Top), and Kernel Estimate of the Density of $T_1|N(1) = 1$ (h_1) and Maximum-Likelihood Estimate of ψ (Bottom)

Note: Based on sample and estimates from Table 1. For the bottom graph, an Epanechnikov kernel with bandwidth 0.05 is used (see Appendix A.1).

fault and claims at partial fault. At first glance, we find that $\hat{H}_{1,n}(t) > t/T$. If we would maintain the stationarity assumption, we could take this as evidence of moral hazard ($\beta < 1$). However, we will later conclude that the deviation of $\hat{H}_{1,n}$ from a uniform distribution should be explained by nonstationarity rather than moral hazard.

We have computed Pearson's χ^2-tests and a two-sided Kolmogorov-Smirnov (KS) test. A natural grouping for the χ^2-statistics is in 365 daily intervals. With 365 intervals, we have roughly thirteen observations per interval. The χ^2-statistic for uniformity is 408.4 and has 364 degrees of freedom. The asymptotic p-value is 0.054. The p-value increases to 0.484 if time is grouped in seventy-three intervals of five days. The (two-sided) KS-test is in line with the first, finer χ^2-test, with $\sup_{t\in[0,T]}|\hat{H}_{1,n}(t) - t/T| = 0.019$ and a corresponding p-value equal to 0.058. The p-value is based on the finite-sample distribution conditional on $M_{1,n}$.

We conclude that a stationary model without occurrence dependence is only (marginally) accepted at a size of 5 percent.

3.4.2 Comparison of the Distributions of the First and Second Claim Times. In this subsection and Subsection 3.4.3, we concentrate on testing the null hypothesis that $\beta = 1$ against moral hazard ($\beta < 1$) or occurrence dependence ($\beta \neq 1$) without further assumptions on Ψ and G.

Analogously to H_1, define H_2 to be the distribution of the second claim time T_2 in the subpopulation with exactly two claims in the contract year ($N(T) = 2$):

$$H_2(t) = \Pr(T_2 \leq t | N(T) = 2).$$

Recall that $H_1 = \Psi$ under the null that $\beta = 1$. Now, it is easy to derive that $H_2(t) = \Psi(t)^2$ and therefore that

$$H_1(t)^2 = H_2(t)$$

for all $t \in [0, T]$ under the null. The latter equality holds for all Ψ and G. So, a feasible test that allows for general nonstationarity (Ψ) and heterogeneity (G) can be based on the difference between the empirical counterparts $\hat{H}^2_{1,n}$ and $\hat{H}_{2,n}$ of H_1^2 and H_2, where

$$\hat{H}_{2,n}(t) := \frac{1}{M_{2,n}} \sum_{i=1}^{n} I(T_{2,i} \leq t, N_i(T) = 2)$$

is defined analogously to $\hat{H}_{1,n}$. To our knowledge, such tests have not been used before.

For these tests to have power, the equality $H_1^2 = H_2$ has to break down under the alternatives of moral hazard ($\beta < 1$) and occurrence dependence ($\beta \neq 1$). A formal power analysis is beyond the scope of this paper. Instead, we will show that $H_1^2 - H_2$ is different from 0 under local alternatives to the null $\beta = 1$. Recall that β is a function, giving the occurrence-dependence effect $\beta(\lambda)$ for

each individual-specific effect λ. Thus, one-sided local alternatives to the null $\beta = 1$ can be expressed as $\beta = 1 + qu$ for some bounded measurable positive function u and some small $q \in \mathbb{R}$.[28] Now consider $H_1(t)^2 - H_2(t)$ as a function of β for given t. Add the argument β and write $H_1(t; \beta)^2 - H_2(t; \beta)$ to make this explicit. We can get some idea of how $H_1(t; \beta)^2 - H_2(t; \beta)$ would change if we would move from the null $\beta = 1$ to some local alternative $\beta = 1 + qu$ by computing the "directional (Gateaux) derivative" of $H_1(t; \beta)^2 - H_2(t; \beta)$ at $\beta = 1$ in the direction u. In this case, this directional derivative is simply the ordinary derivative of $H_1(t; 1 + qu)^2 - H_2(t; 1 + qu)$ with respect to q at $q = 0$. It equals (see Appendix A.2.1)

$$\frac{d}{dq}\left[H_1(t; 1 + qu)^2 - H_2(t; 1 + qu)\right]_{q=0} = \Psi(t)^2(1 - \Psi(t))\left[\frac{-\mathscr{L}_u'''(1)}{\mathscr{L}''(1)} - \frac{\mathscr{L}_u''(1)}{-\mathscr{L}'(1)}\right], \quad \text{(G1)}$$

where $\mathscr{L}_u(s) := \int u(\lambda)\exp(-\lambda s)dG(\lambda)$.

First, consider the special case that $u(\lambda) = 1$ for all λ. Then, the derivative in (G1) represents the change in $H_1(t; \beta)^2 - H_2(t; \beta)$ in response to a small homogeneous (across λ) change in β at $\beta = 1$. In this case, $\mathscr{L}_u = \mathscr{L}$ and the factor in brackets on the right-hand side of (G1) reduces to[29]

$$\frac{-\mathscr{L}'''(1)}{\mathscr{L}''(1)} - \frac{\mathscr{L}''(1)}{-\mathscr{L}'(1)} \geq 0.$$

Thus, the right-hand side of (G1) is nonnegative in this case. This suggests that moral-hazard alternatives ($\beta < 1$) near $\beta = 1$ with homogeneous β correspond to $H_1^2 < H_2$. Under homogeneous $\beta > 1$, on the other hand, we should expect that $H_1^2 > H_2$.

For general directions u this result may not hold. However, if u is positive and increasing, the derivative in (G1) is generally positive. Thus, the results for the homogeneous-β case still hold if occurrence dependence is stronger ($|\beta(\lambda) - 1|$ is larger) for high-λ agents. In particular, in the case that $\beta < 1$ and decreasing we should expect that $H_1^2 < H_2$. In words, if there is moral hazard for all agents and if high-λ (high-risk) agents are more responsive to incentives, then the results for the homogeneous moral-hazard case still apply.

The top panel of Figure 1 plots $\hat{H}_{1,n}^2$ and $\hat{H}_{2,n}$. Apart from some minor reversals at the tails, we find that $\hat{H}_{1,n}^2 > \hat{H}_{2,n}$. The analysis above suggests that this is evidence of $\beta > 1$. A typical one-sided test against the moral-hazard

28. This corresponds to local alternatives such that either $\beta(\lambda) < 1$ (if $q < 0$) or $\beta(\lambda) > 1$ (if $q > 0$) for all λ and includes homogeneous-β alternatives.

29. The inequality follows from the facts that $-\mathscr{L}'$ is completely monotonic and that completely monotonic functions are log-convex (Widder 1946). It holds strictly unless λ is degenerate or has two points of support of which one is 0, in which case it is binding. See Lemma 4 in Appendix A.2.1.

alternative $\beta < 1$ would accept the null that $\beta = 1$ at all sizes. In retrospect, the fact that $\hat{H}_{1,n}(t) > t/T$ in Subsection 3.4.1 should be read as evidence of nonstationarity rather than moral hazard. Once we correct for both heterogeneity and nonstationarity, we do not find evidence of moral hazard.

One final concern is that we may even have evidence in favor of $\beta > 1$, for which we have not put forward any economic theory. To assess the statistical significance of the discrepancy between $\hat{H}^2_{1,n}$ and $\hat{H}_{2,n}$, we compute a two-sided KS-statistic,

$$K_n = \sup_{t\in[0,T]} |\hat{H}_{1,n}(t)^2 - \hat{H}_{2,n}(t)|.$$

This test is distribution-free and finite-sample p-values (conditional of $(M_{1,n}, M_{2,n})$) are easily simulated (see Appendix A.2.2). We find $K_n = 0.067$ (p-value of 0.423, conditional on $(M_{1,n}, M_{2,n}) = (4828, 272)$). So, we do not reject the null that there is no occurrence dependence at any reasonable test size.

3.4.3 Direct Comparison of the First and Second Claim Durations. The test in the previous subsection is based on a comparison of first and second claim times across contracts. Here, we develop a test based on a more direct comparison of the first and second claim times of each contract with two claims (or more).

Main Intuition: The Stationary Case Without Censoring. To develop the main intuition, first consider the stationary model, i.e., let $\psi = T^{-1}$. Then, for given λ, T_1/T and $(T_2 - T_1)/T$ are independent exponential durations with parameters λ and $\beta(\lambda)\lambda$, respectively. So, we can write

$$\ln(T_1) = \ln(T) + \ln(E_1) - \ln(\lambda)$$

and

$$\ln(T_2 - T_1) = \ln(T) - \ln(\beta(\lambda)) + \ln(E_2) - \ln(\lambda)$$

for some unit exponential random variables E_1 and E_2 that are mutually independent and independent of λ. It follows that

$$\ln(T_1) - \ln(T_2 - T_1) = \ln(\beta(\lambda)) + \ln(E_1) - \ln(E_2),$$

with $\ln(E_1) - \ln(E_2)$ independent of λ and symmetrically distributed around 0. This suggests that we use within-individual variation in claim durations to learn about $\ln(\beta(\lambda))$.

Suppose we have an uncensored sample $((T_{1,1}, T_{2,1}), \ldots, (T_{1,n}, T_{2,n}))$ from the joint distribution of (T_1, T_2). Then, we can estimate $\mathbb{E}[\ln(\beta(\lambda))]$ by

$$\widehat{\ln \beta^*_n} = \frac{1}{n}\sum_{i=1}^{n} [\ln(T_{1,i}) - \ln(T_{2,i} - T_{1,i})],$$

in analogy to within-estimation with standard linear panel data, and simply base a test on $\widehat{\ln \beta_n^*}$. Such a test would be a special case of the regression test for "mean" occurrence dependence proposed by Heckman and Borjas (1980, end of Section II.b).

Alternatively, note that

$$\Pr(T_1 \geq T_2 - T_1 | \lambda) = \frac{\beta(\lambda)}{1 + \beta(\lambda)}.$$

If $\beta(\lambda) = 1$, either duration is equally likely to be the largest. If $\beta(\lambda) < 1$, then $\Pr(T_1 \geq T_2 - T_1 | \lambda) < 1/2$, as expected. This suggests that we can base a robust test of the null $\beta = 1$ against the alternative of moral hazard on the share of contracts for which the time up to the first claim is at least as large as the duration between the first and the second claims,

$$\hat{\pi}_n^* = \frac{1}{n} \sum_{i=1}^{n} I(T_{1,i} \geq T_{2,i} - T_{1,i}).$$

This approach is somewhat reminiscent of the methods for paired duration data that have been developed by Holt and Prentice (1974), Chamberlain (1985), and Ridder and Tunalı (1999).[30] In our case, it is hard to apply these methods directly, for two reasons. First, we face a censoring problem: we only observe (at least) two spells for contracts with $T_2 \leq T$. Second, we allow for nonstationarity. The standard methods can deal with duration dependence, i.e., a common dependence of the hazards of T_1 and $T_2 - T_1$ on the duration since the contract renewal date and the first claim, respectively. This generality carries over to our robust statistic above. However, as the durations T_1 and $T_2 - T_1$ are consecutive, nonstationarity can bias our comparison in arbitrary ways. We will now investigate how we can deal with these two problems.

The Censoring Problem under Stationarity. First, focus on the censoring problem and maintain the stationarity assumption $\psi = T^{-1}$ (Appendix A.3.1 provides details). Here, we explicitly analyze the case in which we select only contracts with exactly two claims. This has some analytical advantages, notably that the distribution of $(T_1, T_2) | (\lambda, N(T) = 2)$ does not depend on λ under the null $\beta = 1$ (again, because of the sufficiency of $N(T)$ for λ under the null). This is particularly convenient when we can adapt $\widehat{\ln \beta_n^*}$ into

$$\widehat{\ln \beta_n} = \frac{1}{M_{2,n}} \sum_{i=1}^{n} \ln\left(\frac{T_{1,i}}{T_{2,i} - T_{1,i}}\right) I(N_i(T) = 2).$$

30. See also Van den Berg (2001) for an overview.

It is easy to show that $\widehat{\ln \beta_n}$ is asymptotically normal under the null $\beta = 1$, with expectation 0 and variance $\pi^2/(3np_2)$. Here, p_k is more generally the probability that a contract has k claims in the contract year. Note that, given p_2, the asymptotic standard error does not involve (properties of) the distribution G of λ and can be consistently estimated by $\pi/\sqrt{3M_{2,n}}$.

Next, note that under the null

$$\Pr(T_1 \geq T_2 - T_1 | \lambda, N(T) = 2) = \frac{1}{2}$$

is known and independent of λ as before. Thus, it makes sense to adapt the second test $\hat{\pi}_n^*$ into

$$\hat{\pi}_n = \frac{1}{M_{2,n}} \sum_{i=1}^{n} I(T_{1,i} \geq T_{2,i} - T_{1,i},\ N_i(T) = 2).$$

Under the null $\beta = 1$, $\hat{\pi}_n$ is asymptotically normal with mean 1/2 and variance $1/(4np_2)$. The variance can be estimated consistently by $1/(4M_{2,n})$. Note that

$$\Pr(T_1 \geq T_2 - T_1 | \lambda, N(T) = 2) = \begin{cases} < \frac{1}{2} & \text{if } \beta(\lambda) < 1 \text{ and} \\ > \frac{1}{2} & \text{if } \beta(\lambda) > 1 \end{cases}$$

for all $\lambda > 0$, so that we can construct the test to have power against moral hazard ($\beta < 1$) in particular.

Appendix A.3.2 discusses the alternative case in which we select all contracts with at least two claims. The analysis of the second, robust statistic $\hat{\pi}_n$ directly extends to this case. Extending $\widehat{\ln \beta_n}$ to all contracts with at least two claims is somewhat more cumbersome because, even given $p_2, p_3, \ldots$, the asymptotic distribution of this statistic depends on (properties of) the distribution G of λ under the null. We will not pursue this here.

In our data set, we observe 270 contracts with exactly two claims. We find that $\widehat{\ln \beta_n} = -0.043$. This seems consistent with moral hazard ($\beta < 1$), but the estimated asymptotic standard error of $\widehat{\ln \beta_n}$ under the null is relatively large, 0.110. Thus, the null that $\beta = 1$ is accepted at conventional test sizes. We also find that the duration up to the first claim is larger than the duration between the first and the second claims for $\hat{\pi}_n = 50.4$ percent of the 270 contracts. If we test against the alternative of moral hazard, we do not reject $\beta = 1$ at any size. The estimated standard error of $\hat{\pi}_n$ is 3.0 percent, so that we do not reject against the two-sided alternative of occurrence dependence at conventional test sizes either. We can also compute the equivalent of the second statistic for all 287 contracts with at least two claims. We find that the duration up to the first claim is larger

than the duration between the first and the second claims for 50.2 percent of these contracts, with an estimated asymptotic standard error of 3.0 percent. This confirms our conclusion based on $\widehat{\ln \beta_n}$ and $\hat{\pi}_n$.

A General Test under Nonstationarity. These results depend on the stationarity assumption $\psi = T^{-1}$. Because we have found some circumstantial evidence of nonstationarity in Subsections 3.4.1 and 3.4.2, we would like to explore the consequences of nonstationarity a bit further. First, suppose we know Ψ. Then, we can deal with possible nonstationarity by working in integrated-hazard time instead of contract time. Note that Ψ is increasing on the supports of T_1 and T_2. So, we can work with the transformed durations

$$T_1^* = \Psi(T_1) \qquad \text{and} \qquad T_2^* = \Psi(T_2)$$

instead of T_1 and T_2 without loss of information. This is convenient, as, for given λ, T_1^* and $T_2^* - T_1^*$ are again independent exponential random variables with parameters λ and $\beta(\lambda)\lambda$, respectively. We can directly apply the analysis for the exponential case above, provided that we know Ψ. Then, we can construct $T_{1,i}^* = \Psi(T_{1,i})$ and $T_{2,i}^* = \Psi(T_{2,i})$ and therefore

$$\hat{\pi}_n(\Psi) := \frac{1}{M_{2,n}} \sum_{i=1}^{n} I(T_{1,i}^* \geq T_{2,i}^* - T_{1,i}^*, N_i(T) = 2).$$

This is a generalization of $\hat{\pi}_n$ to arbitrary, but still known, nonstationarity.

In our application, we do not know Ψ and $\hat{\pi}_n(\Psi)$ is not feasible. However, recall from Subsection 3.4.1 that we can estimate Ψ consistently by $\hat{H}_{1,n}$ under the null that $\beta = 1$. This suggests substituting $\hat{H}_{1,n}$ for Ψ and using $\hat{\pi}_n(\hat{H}_{1,n})$ as our test statistic. Under the null $\beta = 1$, $\hat{\pi}_n(\hat{H}_{1,n})$ is asymptotically normal with expectation 1/2 and variance $1/(4np_2) + 1/(6np_1)$. The variance can be estimated consistently as $1/(4M_{2,n}) + 1/(6M_{1,n})$. Appendix A.3.3 provides details.

Substitution of $\hat{H}_{1,n}$ for Ψ comes at the price of lower power. This is due to the fact that $\hat{H}_{1,n}$ is only a consistent estimator of Ψ under the null and generally captures some of the occurrence-dependence effect if $\beta \neq 1$. We can again provide some insight by analyzing the local behavior of the test at $\beta = 1$. Let

$$\pi : h \in \mathcal{D} \mapsto \Pr(2h(T_1) \geq h(T_2) | N(T) = 2),$$

with $\mathcal{D}$ the set of all distribution functions on $[0, T]$ concentrated on $(0, T]$. Then, the population-equivalent of $\hat{\pi}_n(\hat{H}_{1,n})$ can be written as $\pi(H_1)$.

First, consider the population-equivalent $\pi(\Psi)$ of the infeasible statistic $\hat{\pi}_n(\Psi)$. As before, we can investigate the behavior of $\pi(\Psi)$, as a mapping $\beta \mapsto \pi(\Psi; \beta)$ for given Ψ, locally at $\beta = 1$. The directional derivative of $\pi(\Psi; \beta)$ at $\beta = 1$ in the direction $u > 0$ is given by

$$\frac{d}{dq}\left[\pi(\Psi;\, 1+qu)\right]_{q=0} = \frac{1}{12}\frac{-\mathscr{L}'''_u(1)}{\mathscr{L}''(1)} > 0. \tag{G2}$$

This reestablishes, now locally at $\beta = 1$, that $\pi_n(\Psi)$ can distinguish between $\beta < 1$ and $\beta > 1$ if we know Ψ.

For $\pi(H_1)$, as a mapping $\beta \mapsto \pi(H_1(\beta);\, \beta)$, we have instead that

$$\frac{d}{dq}\left[\pi(H_1(1+qu);\, 1+qu)\right]_{q=0} = \frac{1}{12}\left[\frac{-\mathscr{L}'''_u(1)}{\mathscr{L}''(1)} - \frac{\mathscr{L}''_u(1)}{-\mathscr{L}'(1)}\right]. \tag{G3}$$

The first term is again the effect in (G2) that works through the distribution of $(T_1,\, T_2)|(N(T) = 2)$. The second term is the counteracting effect on the time-transformation H_1. Note that the derivative in (G3) has the same sign as the derivative of $H_1^2 - H_2$ in Subsection 3.4.2. It follows that generally $\pi(H_1) < 1/2$ for homogeneous moral hazard alternatives close to the null. Furthermore, this result carries over to the moral-hazard alternative in which high-risk (high-λ) agents are more responsive to incentives (that is, have higher $|\beta(\lambda) - 1|$). Appendix A.3.4 provides details.

We find that $\hat{\pi}_n(\hat{H}_{1,n}) = 50.7$ percent in our data, confirming our earlier conclusion that we do not reject the null $\beta = 1$ against the alternative of moral hazard at any test size. The estimated asymptotic standard error is 3.1 percent, so that we do not reject the null against a two-sided alternative at reasonable sizes either.

3.5 Identification and Estimation

3.5.1 Identification. We now specialize the model in (M) by imposing homogeneity of the occurrence-dependence effects across contracts. Formally, we impose that β is a trivial function of λ and simply write

$$\theta(t|\lambda,\, N[0,\, t)) = \lambda\beta^{N(t-)}\psi(t), \tag{M$^+$}$$

with β now a positive scalar parameter. This additional structure will facilitate the identification and estimation of the model.

In this subsection we investigate identification. Appendix B.1 provides proofs and other details. Note that the model is fully characterized by the triple $(\beta, \Psi, \mathscr{L})$: each choice of the triple $(\beta, \Psi, \mathscr{L})$ maps into exactly one distribution of $N[0, T]$.[31] We say that $(\beta, \Psi, \mathscr{L})$ is "identified" if this mapping is one-to-one. Identification of certain features of $(\beta, \Psi, \mathscr{L})$ can be defined analogously. For example, the sign of $\beta - 1$ is identified if it is uniquely determined by the distribution of $N[0, T]$.

31. Recall that there is a one-to-one relation between G and its Laplace transform $\mathscr{L}$.

The following result implies that the null of no moral hazard is empirically distinguishable from the alternatives of occurrence dependence and moral hazard without further assumptions.

PROPOSITION 2. *The sign of* $\beta - 1$ *is identified.*

PROOF. See Appendix B.1. ■

In addition, we conjecture that the parameter β is point-identified without further assumptions. Define $q_k(t) := \Pr(N(t) = k)$ for $t \in [0, T]$. Suppose that $\beta < 1$, which we can tell from the data by Proposition 2 (the case $\beta > 1$ is similar). Key to the identification of β is the fact that it satisfies

$$q_1\{q_0^{-1}[(1 - \beta)q_1(t) + q_0(t)]\} = \beta q_1(t) + (1 - \beta^2)q_2(t) \qquad \text{(I)}$$

for all $t \in [0, T]$. Because the functions q_0, q_1, and q_2 are data, this provides a continuum of nonlinear restrictions on the scalar parameter β. Further analysis of this problem is beyond the scope of this paper.

Finally, we have a result on the identifiability of Ψ and $\mathcal{L}$ in the case that β is known. First note that we then also know

$$\tilde{q}(t) := (1 - \beta)q_1(t) + q_0(t) = \mathcal{L}(\beta\Psi(t)).$$

We have already seen in Subsection 3.4.1 that Ψ and $\mathcal{L}$ are identified if $\beta = 1$. In the case that $\beta \neq 1$, note that q_0 and $\tilde{q}$ jointly constitute the data of the restriction of a two-sample MPH model to $[0, T]$, with "treatment effect" β, "integrated baseline hazard" Ψ and Laplace transform $\mathcal{L}$ of the "mixing distribution." [32] Thus, we can identify $\mathcal{L}$ and Ψ along the lines of standard two-sample identification proofs for the MPH model (Elbers and Ridder 1982, and Kortram et al. 1995). These proofs rely on the additional assumption that $\mathbb{E}[\lambda] < \infty$. Thus, we have

PROPOSITION 3. *Suppose that* β *is known. Then,* $\mathcal{L}$ *and* Ψ *are identified in the class of models* $(\beta, \Psi, \mathcal{L})$ *such that* $-\mathcal{L}'(0+) = \mathbb{E}[\lambda] < \infty$.

PROOF. See Appendix B.1. ■

The assumption that $\mathbb{E}[\lambda] < \infty$ is not innocuous. Ridder (1990) provides extensive discussion in the context of single-spell MPH models.

3.5.2 Maximum-Likelihood Estimation. Finally, we have estimated parametric versions of the model ($M^{\dagger}$) by maximum likelihood. We have chosen a piece-

32. To be precise, suppose we observe two samples of durations. Then, q_0 and $\tilde{q}$ are the survival functions in the first and second sample in the case that the units in the first sample have hazards $\psi(t)\lambda$ conditional on λ, the units in the second sample hazards $\beta\psi(t)\lambda$ conditional on λ, and λ has the same distribution with Laplace transform $\mathcal{L}$ in both samples.

wise-constant specification of ψ. In particular, we partition the contract year [0, 1] in 12 months and set ψ to be constant within each month:

$$\psi(t) = \sum_{j=1}^{12} \psi_j I\left(\frac{j-1}{12} \leq t < \frac{j}{12}\right),$$

with $\psi_1, \ldots, \psi_{12} \geq 0$ parameters to be estimated, up to the normalization $\Psi(1) = (1/12) \sum_{j=1}^{12} \psi_j = 1$. For the distribution G of λ, we have experimented with various discrete distributions. We have found no evidence that G has more than 2 mass points. Therefore we present results for a model without unobserved heterogeneity, in which case we only estimate a constant, and results for a model with two points of support for λ. In the latter case, we have to estimate the support points λ^a, $\lambda^b \geq 0$, and one probability $\Pr(\lambda = \lambda^a) = 1 - \Pr(\lambda = \lambda^b)$. Appendix B.2 provides details on the construction of the likelihood.

Table 1 presents results for a specification in which the unobserved heterogeneity has 2 points of support and ψ consists of 12 monthly pieces. We find an estimate of β just below 1, with a large standard error. The point estimates are consistent with nondegenerate heterogeneity, but again the precision is low. If we estimate a model without unobserved heterogeneity, the estimate of β increases to 1.729 with a relatively small standard error of 0.091. This clearly illustrates the fact that unobserved heterogeneity causes spurious positive occurrence dependence.

In the bottom panel of Figure 1 we have plotted the estimated time effects (ψ) of Table 1. The estimates closely track the kernel estimates of the density of H_1 discussed earlier. Figure 2 plots the corresponding estimate of Ψ and its pointwise 95 percent confidence bounds. The uniform cumulative distribution function lies well within the latter. Indeed, we do not reject stationarity according to a Wald test. If we estimate a specification with heterogeneity that imposes stationarity, the estimate of β drops to 0.817 with an estimated standard error of 0.237. At conventional sizes, however, the estimate does not deviate significantly from 1.

All in all, the results are consistent with the nonparametric tests of the previous subsection. If we only control for heterogeneity and impose stationarity, we find a point estimate of β below 1. This mirrors the finding that generally $\hat{H}_{1,n}(t) > t/T$, which under the assumption of stationarity can be interpreted as evidence in favor of moral hazard (see Subsection 3.4.1). However, we do not find evidence of moral hazard once we control for both heterogeneity and nonstationarity. This confirms the test results in Subsections 3.4.2 and 3.4.3. It should be noted that the precision of, in particular, the maximum likelihood estimates is low if both flexible time effects and heterogeneity are included. We return to this in Section 4.

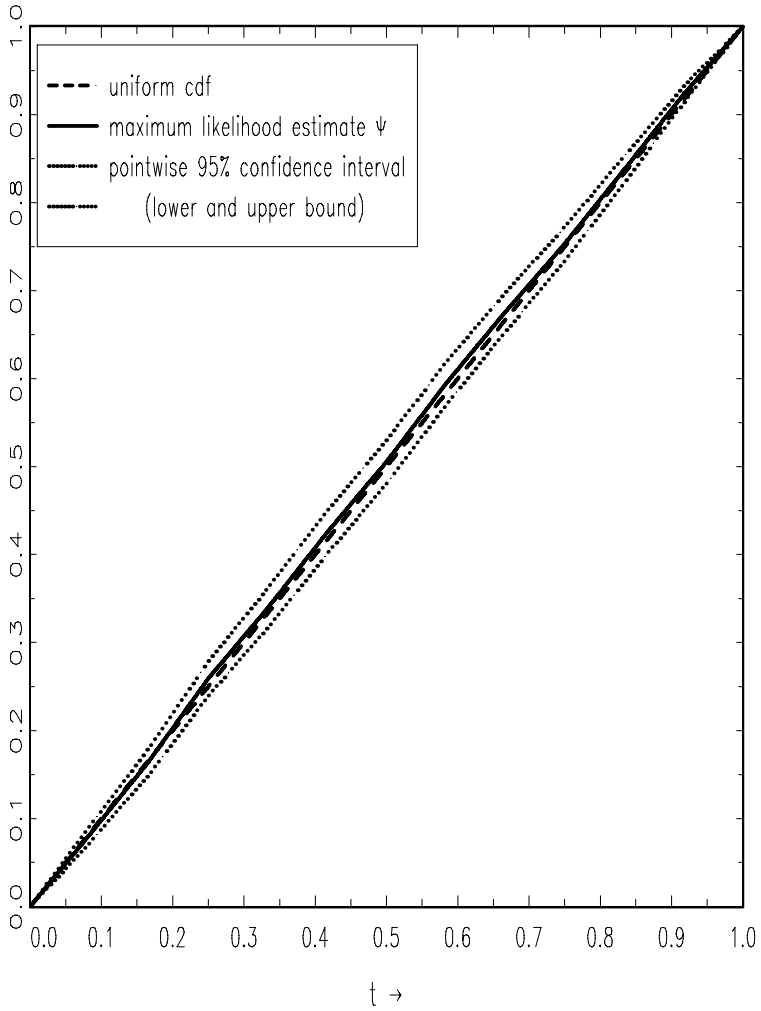


FIGURE 2. Maximum-Likelihood Estimate Ψ

Note: Based on sample and estimates from Table 1.

3.6 Sensitivity Analysis

3.6.1 The Fault-Status of Claims. So far, we have pooled claims at full fault and claims at partial fault, even though the financial consequences for the insuree differ quantitatively. To check whether this matters for our results, we have recomputed the tests and estimates of the previous subsections on data of claims at full fault only. There are 4,340 contracts with one claim at full fault, 230 contracts with two such claims, eleven contracts with three such claims and one contract with four such claims.

The test results are close to those based on all claims at fault. The stationarity tests are slightly less marginal in not rejecting stationarity, with all p-values now above 10 percent. All occurrence-dependence and moral-hazard tests accept the null that $\beta = 1$, with only slightly lower precisions. The three π-statistics are still just over 50 percent; $\widehat{\ln \beta_n}$ has switched sign to 0.026, but remains very close to 0. The estimation results confirm these results and are in line with the estimates on the pooled-claims data.

We have also rerun the tests and estimations on a data set that includes all claims, rather than just claims at (full or partial) fault. Even though the moral-hazard argument applies specifically to claims at fault, other theories may imply state dependence involving occurrence of not-at-fault claims as well. For example, agents may learn from accidents, even if they were not at fault, and this may lead to negative occurrence-dependence in itself (see Subsection 3.1). Obviously, given that we have not found any evidence of occurrence dependence so far, we do not expect to find any if we include all claims either. However, we should be careful as the precision of our tests and estimates will typically increase. After all, the number of claims increases considerably: there are 12,861 contracts with one, 1,996 contracts with two, 307 contracts with three, 43 contracts with four, seven contracts with five, one contract with six and one contract with seven claims.

We do not find much evidence of nonstationarity in the raw data. The p-values of the χ^2-tests are now 0.503 and 0.063 and the p-value of the KS-test is 0.077. The KS-test of occurrence dependence is highly insignificant. Surprisingly, $\widehat{\ln \beta_n} = 0.147$ with an estimated standard error under the null of 0.041. The more robust π-statistics are however consistent with the KS-test. Using contracts with exactly 2 claims only, we find $\hat{\pi}_n = 52.0$ percent (standard error 1.1 percent). If we use all contracts with at least 2 claims, we find 51.8 percent (1.0 percent). The corresponding p-values for a two-sided test are 0.081 and 0.080, respectively. This may seem slightly supportive of the result for $\widehat{\ln \beta_n}$, but recall that neither of these tests corrects for nonstationarity. The most general π-test does, and delivers $\hat{\pi}_n(\hat{H}_{1,n}) = 51.2$ percent (standard error 1.2 percent). The two-sided p-value is 0.327.

The parametric estimation results confirm this picture. The most appropriate specification seems to be one with two-point heterogeneity and 24 (half-monthly) time-intervals. Without regressors, the estimate of β is 1.117 with a standard error of 0.131. A Wald test with 23 degrees of freedom rejects stationarity at all reasonable sizes (p-value 0.005). As expected, overall the precision is much higher, even though we now have 24 instead of 12 time intervals.

In conclusion, we do not find evidence of occurrence dependence in data including all claims, even though the results are relatively precise. One of the reasons we have put forward for (negative) occurrence-dependence effects of

claims in general is learning. Thus, this suggests that there are no such learning effects. Note though that learning is particularly relevant for young and inexperienced drivers. If so, learning implies different occurrence-dependence parameters for young and old drivers. In the next subsection, we provide some results for samples stratified in this and other ways, both for data with at-fault claims only and for data with all claims.

3.6.2 Stratification with Respect to Some Regressors. First, consider again the data with claims at (full or partial) fault. We have stratified the data on respectively sex, age and experience (driver's license age) and have rerun the tests on each of the sub-samples.

We have 61,564 contracts with male insurees and 26,372 contracts with female insurees. The male test results closely resemble the results for males and females pooled, with the KS-test now accepting stationarity at all reasonable sizes. The results for females based on the χ^2-tests, $\widehat{\ln \beta_n}$ and the π-tests are also in line with the overall results. The precision is, understandably, low. Unlike the overall and male KS-statistics, the female KS-statistics are significant at low sizes. The KS-statistic for stationarity has a p-value of 0.004; the KS-statistic for occurrence dependence a p-value of 0.013. We find that $\hat{H}^2_{1,n} > \hat{H}_{2,n}$. Thus, for females we have an inconsistency between the χ^2-tests, $\widehat{\ln \beta_n}$ and the π-tests on the one hand and the KS-tests on the other hand. One explanation is that the low female sample size leaves room for outliers to affect the results.

Of all contracts for which the insuree's year of birth is observed, 26,372 are born in 1951 or later ("young") and 61,564 are born in 1950 or before ("old"). Recall that our data concerns contract years starting anytime during the year following October 1, 1987. We have deliberately constructed the young drivers to be truly young (and therefore a relatively small group), because we expect that any learning effects of accidents would quickly disappear with age. The test results are very similar between both age groups and are in line with the overall results. Again, the χ^2-tests are even less significant. One difference is that the KS-test on stationarity for young insurees is now highly significant, with a p-value of 0.004. This is somewhat comparable to the results for the similarly small sample of females.

These results suggest that learning effects, leading to relatively strong negative occurrence dependence for young drivers, are not important. It seems, though, that driving experience rather than age per se would interact with learning. Obviously, we do not observe actual driving experience, but we do know the years in which the insurees' driver's licenses were issued. We have divided the sample in 12,712 insurees with licenses issued in 1980 or later ("inexperienced") and 75,909 insurees with licenses issued in 1979 or before ("experienced"). We do not find evidence of nonstationarity, although the KS-statistic for (again, the small group of) inexperienced insurees has a p-value as low as 0.082. The KS-tests on occurrence dependence are highly insignificant for either experience level, but the other occurrence-dependence tests produce

some interesting results. For inexperienced drivers, we find that $\widehat{\ln \beta_n} = -0.296$ with a standard error of 0.222. The three π-statistics are in the range 40.3–41.2 percent, with standard errors 6.1–6.3 percent. For experienced drivers, on the other hand, we have that $\widehat{\ln \beta_n} = 0.041$ (0.127) and π-statistics in the range 53.0–53.7 percent (3.4–3.6 percent). The results for experienced drivers are consistent with the results for the pooled sample. Also, even if we use one-sided tests on moral hazard, we do not reject the null that $\beta = 1$ at a 5 percent size for inexperienced drivers. However, the differences between both experience levels are remarkable and suggest that, if anything, there is negative occurrence dependence for inexperienced drivers only. This points at learning rather than moral hazard effects of accidents.

Clearly, this conclusion cannot be drawn with any reasonable statistical significance because of the imprecision of our results. However, we have earlier argued that it may make sense to include claims that are not at fault in an analysis of learning. If this is correct, the resulting larger sample may be more informative on any differences in occurrence-dependence effects between experience levels.

The results for experienced drivers are consistent with the results for the pooled data. The occurrence-dependence tests that do not correct for nonstationarity are strongly in favor of $\beta > 1$. However, the KS-statistic on occurrence dependence is very insignificant and $\hat{\pi}_n(\hat{H}_{1,n}) = 51.8$ percent with a standard error of 1.3 percent and a two-sided p-value of 0.165. The results for inexperienced drivers are all insignificant, but indeed mostly pointing at $\beta < 1$. However, $\widehat{\ln \beta_n}$ is slightly positive and $\hat{\pi}_n(\hat{H}_{1,n}) = 49.6$ percent (standard error 2.8 percent). We conclude that the data including all claims are not supporting differences in occurrence-dependence effects between experience levels. For now, we can shelve the learning explanation of occurrence dependence.

4. Conclusions

In this paper, we show that the experience-rating structure commonly found in insurance contracts can be exploited in the empirical analysis of moral hazard. In particular, experience rating implies negative occurrence dependence of individual claim intensities under moral hazard. In other words, under moral hazard and experience rating individual claim intensities decrease with the number of past claims. In observed claim intensities, this negative occurrence dependence effect is confounded with a positive selection effect: an insuree with a large number of past claims is likely to be a bad driver and therefore to have a high future claim intensity. Thus, from an empirical perspective the distinction between moral hazard and (adverse) selection boils down to disentangling "true" state dependence and unobserved heterogeneity. This is a problem that has been studied at length in labor economics in the late 1970s and early 1980s. Following up on this literature, we

develop general tests of the null of no moral hazard. Our tests are nonparametric (except for a separability assumption), and generalize existing work by allowing for general nonstationarity of the claim intensity.

We have applied our tests to French car-insurance data and have found no evidence of moral hazard (or more general occurrence dependence). More precisely, we have not rejected the null of no moral hazard against the alternatives of moral hazard or general occurrence dependence at conventional levels. This result is confirmed by parametric estimates of a flexible model that allows for both occurrence dependence and selection on unobservables.

One remaining, practical concern is that observations of multiple claims by a single insuree are central to identifying moral hazard effects. Because claims are relatively rare and we focus on claims at fault within a single contract year, we observe multiple claims for relatively few of the many contracts in our data set. This translates into a fairly low precision of our empirical results. One solution would be to resort to low-dimensional parametric models, but this would artificially generate precision at the expense of robustness. We prefer to simply qualify our identification results by noting that even a large data set carries limited information on moral hazard effects. Note that this problem does not seem to be fundamental (i.e., it would be resolved if we would have a very large data set): a complementary analysis of a larger data set of all (at-fault and not-at-fault) claims yields results of satisfactory precision.

An obvious way to expand the data is to include claim histories beyond a single contract year. Our French data provide information for up to two years and, at least in principle, it should be possible to collect alternative data on many more years. Obviously, multiple claims will be more prevalent in longer claim histories. On the downside, however, using claim histories that extend beyond a single contract year introduces the problem of dynamic contract selection. This may imply nonignorable attrition.

This takes us to our final remark. The main contribution of this paper is to provide, in a dynamic context, tests of moral hazard that are valid in the presence of general unobserved heterogeneity. Our analysis is consistent with heterogeneity in accident rates that is symmetrically observed between the insurer and the agents. It also allows for adverse selection in the technical sense, which arises if agents are better informed about their risk than insurers. However, because we focus on claims and ignore other insurance events, we are not able to distinguish between both types of heterogeneity. In particular, we have not modelled changes in contracts (risk class, coverage, etc.). Such changes are observed across contract years, but also within contract years. Some contract changes may be forced by events that can safely be considered to be external to the claims process, but occasionally a case can be made for the endogeneity of contract changes to the agent's claim history. A common assumption in insurance theory is that both the agent and the insurer learn about the agent's ability, but that the learning process is asymmetric because the information available to the agent is much richer. If so, one would expect the agent's decisions about contract changes to be informative about her risk, even after controlling for the information available to the insurer (i.e., the agent's observable

characteristics and past history). Again, this (complex) problem is left for future research.

Appendices

A. Testing for Moral Hazard

This appendix provides results for Subsection 3.4. Note that the analysis in this subsection is based on the general econometric model (M).

A.1 Results for Subsection 3.4.1 (H_1)

A.1.1 Asymptotic Properties of $\hat{H}_{1,n}$. We use "$\Rightarrow$" to denote convergence in distribution (weak convergence) and "$\overset{\text{a.s.}}{\Rightarrow}$" to denote almost-sure convergence. Throughout, $\mathbb{G}_U$ is a uniform (on [0, 1]) Brownian bridge and, for any distribution H, $\mathbb{G}_H = \mathbb{G}_U \circ H$ a H-Brownian-bridge. The following properties of the estimator $\hat{H}_{1,n}$ are standard.

LEMMA 3.

$$\sup|\hat{H}_{1,n} - H_1| \xrightarrow{\text{a.s.}} 0 \text{ and } \sqrt{n}\,(\hat{H}_{1,n} - H_1) \Rightarrow \frac{1}{\sqrt{p_1}}\,\mathbb{G}_{H_1}$$

as $n \to \infty$.

PROOF. The result follows from the Glivenko-Cantelli and Donsker theorems (e.g., Van der Vaart 1998, Theorems 19.1 and 19.3), the law of large numbers and Slutsky's lemma. ■

Note that $H_1 = \Psi$ under the null that $\beta = 1$.

A.1.2 The Behavior of H_1 under the Alternative that $\beta \neq 1$. Under the null that $\beta = 1$, $H_1(t|\lambda) = \Psi(t)$ for $t \in [0, T]$, so that $H_1(\Psi^{-1}(z)|\lambda) = z$ for $z \in [0, 1]$. If $\beta(\lambda) \neq 1$, on the other hand,

$$\Pr(N(t) = 1, N(T) = 1|\lambda) = \int_0^t \lambda\psi(u)\mathrm{e}^{-\lambda[1-\beta(\lambda)]\Psi(u)-\lambda\beta(\lambda)}du$$

$$= \frac{\mathrm{e}^{-\lambda\beta(\lambda)}}{1-\beta(\lambda)}\left[1 - \mathrm{e}^{-\lambda[1-\beta(\lambda)]\Psi(t)}\right],$$

so that

$$H_1(t|\lambda) = \frac{\Pr(N(t) = 1, N(T) = 1|\lambda)}{\Pr(N(T) = 1|\lambda)} = \frac{1 - e^{-\lambda[1-\beta(\lambda)]\Psi(t)}}{1 - e^{-\lambda[1-\beta(\lambda)]}}.$$

Substituting $z = \Psi(t)$ gives

$$H_1(\Psi^{-1}(z)|\lambda) = \frac{1 - e^{-\lambda[1-\beta(\lambda)]z}}{1 - e^{-\lambda[1-\beta(\lambda)]}}.$$

$H_1(\Psi^{-1}(z)|\lambda)$ increases from 0 to 1 on [0, 1] and is strictly concave if $\beta(\lambda) < 1$ and strictly convex if $\beta(\lambda) > 1$. This implies that

$$H_1(\Psi^{-1}(z)|\lambda) = \begin{cases} > z & \text{if } \beta(\lambda) < 1 \text{ and} \\ < z & \text{if } \beta(\lambda) > 1 \end{cases}$$

for all $\lambda > 0$ and $z \in (0, 1)$. The inequalities in (H1†) in Subsection 3.4.1 follow.

A.1.3 Kernel Estimation of the Density of H_1. Here, we provide the details of the estimation procedure used to estimate the Lebesgue density h_1 corresponding to H_1. A standard kernel density estimator of h_1 is

$$\tilde{h}_1(t) := \frac{1}{b}\int k\left(\frac{t-x}{b}\right)d\hat{H}_{1,n}(x) = \frac{1}{bM_{1,n}}\sum_{i=1}^{n} I(N_i(T) = 1)k\left(\frac{t - T_{1,i}}{b}\right),$$

with $0 < b < ½$ the bandwidth and k the Epanechnikov kernel function

$$k(x) = \begin{cases} ¾\,(1 - x^2) & \text{if } |x| \le 1, \text{ and} \\ 0 & \text{if } |x| > 1. \end{cases}$$

Now, as H_1 has support $[0, T]$, $\tilde{h}_1(t)$ may have support on $[-b, T + b]$. The restriction of $\tilde{h}_1(t)$ to $[0, T]$ generally under-estimates h_1 on $[0, b]$ and $[T - b, T]$. The ad hoc solution we have used here is to "reflect" the mass of $\tilde{h}_1$ outside $[0, T]$ into $[0, T]$, and estimate h_1 by

$$\hat{h}_1(t) = \begin{cases} \tilde{h}_1(t) + \tilde{h}_1(-t) & \text{if } 0 < t < b, \\ \tilde{h}_1(t) & \text{if } b < t < T - b, \\ \tilde{h}_1(t) + \tilde{h}_1(2T - t) & \text{if } T - b < t < T, \\ 0 & \text{if } t \le 0 \text{ or } t \ge T. \end{cases}$$

A.1.4 Computing the Distributions of the Uniformity Tests under the Null. Time is grouped in 365 days in our sample. If we maintain that $\hat{H}_{1,n}$ is defined on the underlying continuous-time sample, this translates into observing $\hat{H}_{1,n}$ on a grid $\{T/365, 2T/365, \ldots, T\}$ only. Chi-square statistics can be straightforwardly computed using the natural grouping of the data in 365 days (or any coarser grouping). The computation of KS-statistics requires slightly more care. Grouped data only allow us to compute bounds on the continuous-time KS-statistic

$$\sup_{t\in[0,T]} |\hat{H}_{1,n}(t) - t/T|. \tag{1}$$

A sharp lower bound is given by the discrete-time KS-statistic

$$\max_{t\in\{T/365,2T/365,\ldots,T\}} |\hat{H}_{1,n}(t) - t/T|. \tag{2}$$

An upper bound is also easy to derive and the bounds can be expected to be narrow due to the small size of the intervals relative to the density of the claims (see the similar analysis in Appendix A.2.2). We could use standard distribution theory for the continuous-time statistic in (1) to derive corresponding bounds on the p-value for this statistic. In this case, however, it is easier to simply use the discrete-time statistic in (2) itself. Its distribution under the null is known and exact critical and p-values are easy to simulate by Monte Carlo methods.

The finite-sample p-values reported are conditional on the subsample size $M_{1,n}$, which is random even if n is not. It is easy to see that $\hat{H}_{1,n} \sim \hat{H}^*_{1,m}$ given $M_{1,n} = m \in \mathbb{N}$, with $\hat{H}^*_{1,m}$ the empirical distribution of a random sample of fixed size m from H_1.[33] Here and below, "$\sim$" denotes equality in distribution.

A.2 Results for Subsection 3.4.2 ($H_1^2 - H_2$)

A.2.1 The Behavior of $H_1^2 - H_2$ under Local Alternatives to $\beta = 1$. This appendix provides details on the directional derivative of $H_1(t;\beta)^2 - H_2(t;\beta)$ at $\beta = 1$ in the direction u. From Appendix A.1.2 we know that

$$H_1(t;\,1+qu) = \frac{\displaystyle\int \frac{e^{-\lambda[1+qu(\lambda)]}}{qu(\lambda)}[e^{\lambda qu(\lambda)\Psi(t)} - 1]dG(\lambda)}{\displaystyle\int \frac{e^{-\lambda[1+qu(\lambda)]}}{qu(\lambda)}[e^{\lambda qu(\lambda)} - 1]dG(\lambda)}.$$

33. For $l \in \mathbb{N}$ and $(t_1,\ldots,t_l) \in \mathbb{R}^l$, we can write

$$\Pr(\hat{H}_{1,n}(t_1) \le x_1,\ldots,\hat{H}_{1,n}(t_l) \le x_l | M_{1,n} = m)$$
$$= \mathbb{E}[\Pr(\hat{H}_{1,n}(t_1) \le x_1,\ldots,\hat{H}_{1,n}(t_l) \le x_l | N_1(T),\ldots,N_n(T), M_{1,n} = m) | M_{1,n} = m].$$

Note that this holds in particular for $l = 365$ and $t_j = Tj/365$. Because

$$\Pr(\hat{H}_{1,n}(t_1) \le x_1,\ldots,\hat{H}_{1,n}(t_l) \le x_l | N_1(T) = n_1,\ldots,N_n(T) = n_n)$$
$$= \Pr\left(\frac{\sum_{i=1}^m I(T_{1,i} \le t_1)}{m} \le x_1,\ldots,\frac{\sum_{i=1}^m I(T_{1,i} \le t_l)}{m} \le x_l \,\middle|\, N_1(T) = \cdots = N_m(T) = 1\right)$$

for each $(n_1,\ldots,n_n) \in \{0,1\}^n$ such that $\sum_{i=1}^n n_i = m$, it follows that

$$\Pr(\hat{H}_{1,n}(t_1) \le x_1,\ldots,\hat{H}_{1,n}(t_l) \le x_l | M_{1,n} = m)$$
$$= \Pr\left(\frac{\sum_{i=1}^m I(T_{1,i} \le t_1)}{m} \le x_1,\ldots,\frac{\sum_{i=1}^m I(T_{1,i} \le t_l)}{m} \le x_l \,\middle|\, N_1(T) = \cdots = N_m(T) = 1\right).$$

It is easy to derive that

$$\frac{d}{dq}\left[\int \frac{e^{-\lambda[1+qu(\lambda)]}}{qu(\lambda)}(e^{\lambda qu(\lambda)\Psi(t)}-1)dG(\lambda)\right]_{q=0} = -\frac{1}{2}\Psi(t)[2-\Psi(t)]\mathcal{L}_u''(1),$$

so that

$$\frac{d}{dq}[H_1(t;\,1+qu)^2]_{q=0} = -\Psi(t)^2[1-\Psi(t)]\frac{\mathcal{L}_u''(1)}{-\mathcal{L}'(1)}.$$

Next, for λ such that $\beta(\lambda) \neq 1$, we have that

$$\begin{aligned}
&\Pr(N(t)=2, N(T)=2|\lambda)\\
&\quad=\int_0^t\int_{t_1}^t \lambda^2\beta(\lambda)\psi(t_1)\psi(t_2)e^{-\lambda[1-\beta(\lambda)]\Psi(t_1)-\lambda\beta(\lambda)[1-\beta(\lambda)]\Psi(t_2)-\lambda\beta(\lambda)^2}dt_2dt_1\\
&\quad=\frac{e^{-\lambda\beta(\lambda)^2}(1-[1+\beta(\lambda)]e^{-\lambda\beta(\lambda)[1-\beta(\lambda)]\Psi(t)}+\beta(\lambda)e^{-\lambda[1-\beta(\lambda)^2]\Psi(t)})}{[1-\beta(\lambda)]^2[1+\beta(\lambda)]}.
\end{aligned}$$

Thus, for $q \neq 0$

$$\begin{aligned}
H_2(t;\,1+qu) = &\int \frac{e^{-\lambda[1+qu(\lambda)]^2}}{q^2u(\lambda)^2[2+qu(\lambda)]}(1-[2+qu(\lambda)]e^{\lambda qu(\lambda)[1+qu(\lambda)]\Psi(t)}\\
&+[1+qu(\lambda)]e^{\lambda qu(\lambda)[2+qu(\lambda)]\Psi(t)})dG(\lambda)\\
&\Big/\int \frac{e^{-\lambda[1+qu(\lambda)]^2}}{q^2u(\lambda)^2[2+qu(\lambda)]}(1-[2+qu(\lambda)]e^{\lambda qu(\lambda)[1+qu(\lambda)]}\\
&+[1+qu(\lambda)]e^{\lambda qu(\lambda)[2+qu(\lambda)]})dG(\lambda).
\end{aligned}$$

Now, using that

$$\begin{aligned}
&\frac{d}{dq}\left[\int \frac{e^{-\lambda[1+qu(\lambda)]^2}}{q^2u(\lambda)^2[2+qu(\lambda)]}(1-[2+qu(\lambda)]e^{\lambda qu(\lambda)[1+qu(\lambda)]\Psi(t)}\right.\\
&\qquad\left.+[1+qu(\lambda)]e^{\lambda qu(\lambda)[2+qu(\lambda)]\Psi(t)})dG(\lambda)\right]_{q=0}\\
&\qquad=\frac{1}{2}\Psi(t)^2\{[2-\Psi(t)]\mathcal{L}_u'''(1)+\mathcal{L}_u''(1)\}
\end{aligned}$$

we find that

$$\frac{d}{dq}\left[H_2(t;\,1+qu)\right]_{q=0} = -\Psi(t)^2[1-\Psi(t)]\frac{-\mathcal{L}_u'''(1)}{\mathcal{L}''(1)}.$$

Conclude that

$$\frac{d}{dq}\left[H_1(t;\,1+qu)^2 - H_2(t;\,1+qu)\right]_{q=0}$$
$$= \Psi(t)^2(1-\Psi(t))\left[\frac{-\mathcal{L}_u'''(1)}{\mathcal{L}''(1)} - \frac{\mathcal{L}_u''(1)}{-\mathcal{L}'(1)}\right]. \quad \text{(G1)}$$

The following two results provide sufficient conditions for (G1) to be positive on $(0, T)$. First, if u is constant (trivial) then $\mathcal{L}_u^{(k)}(1) = u\mathcal{L}^{(k)}(1)$ (here, the superscript (k) denotes the k-th derivative). Then, Lemma 4 implies that (G1) is positive if u is positive and G has at least two positive points of support.

LEMMA 4. *If $\mathcal{L}$ is the Laplace transform of a distribution G with nonnegative support such that $G(0) < 1$, then*

$$-\frac{d}{ds}\ln\left[\frac{\mathcal{L}''(s)}{-\mathcal{L}'(s)}\right] = \frac{-\mathcal{L}'''(s)}{\mathcal{L}''(s)} - \frac{\mathcal{L}''(s)}{-\mathcal{L}'(s)} \geq 0, \quad (3)$$

with equality holding if and only if G is either degenerate or has two points of support of which one is 0.

PROOF. For given $s \in (0, \infty)$, $x \in [0, \infty) \mapsto \mathcal{L}(s + x)/\mathcal{L}(s)$ is the Laplace transform of a distribution $\tilde{G}$ that has the same support as G. The k-th moment of $\tilde{G}$ exists and is given by $\tilde{\mu}_k := (-1)^k\mathcal{L}^{(k)}(s)/\mathcal{L}(s)$. Thus, (3) has the sign of $\tilde{\mu}_3\tilde{\mu}_1 - (\tilde{\mu}_2)^2$ and the claimed result follows from standard results for the Stieltjes moment problem (e.g., Shohat and Tamarkin 1943, Theorem 1.3). ■

Second, for positive and increasing u we can apply

LEMMA 5. *If u is positive and increasing, then*

$$\frac{-\mathcal{L}_u'''(1)}{\mathcal{L}''(1)} - \frac{\mathcal{L}_u''(1)}{-\mathcal{L}'(1)} > 0.$$

PROOF. Define

$$\bar{u} := \frac{\mathcal{L}_u''(1)}{\mathcal{L}''(1)} \text{ and } \tilde{u}(\lambda) := u(\lambda) - \bar{u}.$$

Then $\mathscr{L}''_{\tilde{u}}(1) = 0$, so that

$$\mathscr{L}'''_u(1)\mathscr{L}'(1) - \mathscr{L}''_u(1)\mathscr{L}''(1) = \frac{\mathscr{L}''_u(1)}{\mathscr{L}''(1)}\left[\mathscr{L}'''(1)\mathscr{L}'(1) - \mathscr{L}''(1)\mathscr{L}''(1)\right] + \mathscr{L}'''_{\tilde{u}}(1)\mathscr{L}'(1). \quad (4)$$

We have to show that (4) is positive. The first term on the right-hand-side of (4) is nonnegative by Lemma 4. Next, note that $[\lambda - u^{-1}(\bar{u})]\tilde{u}(\lambda) > 0$ for all $\lambda \neq u^{-1}(\bar{u})$. This implies that

$$-\mathscr{L}'''_{\tilde{u}}(1) = \int \lambda^3 \tilde{u}(\lambda)\mathrm{e}^{-\lambda} dG(\lambda) = \int \lambda^2[\lambda - u^{-1}(\bar{u})]\tilde{u}(\lambda)\mathrm{e}^{-\lambda} dG(\lambda) > 0.$$

Thus, the second term on the right-hand side of (4) is positive. ■

A.2.2 Computing the Distribution of K_n under the Null. The finite-sample distribution of K_n conditional on the relevant subsample sizes $(M_{1,n}, M_{2,n})$ follows from a simple quantile transformation.

PROPOSITION 4. *Under the null $\beta = 1$ and conditional on $(M_{1,n}, M_{2,n}) = (m_1, m_2)$,*

$$K_n \sim \sup_{u\in[0,1]} |\hat{U}_{1,m_1}(u)^2 - \hat{U}_{2,m_2}(u^2)|.$$

Here, $\hat{U}_{1,m_1}$ and $\hat{U}_{2,m_2}$ are independent uniform empirical distribution functions with m_1 and m_2 points of support, respectively, for given $m_1, m_2 \in \mathbb{N}$.

PROOF. Along the lines of Footnote 33 it is easy to show that $(\hat{H}_{1,n}, \hat{H}_{2,n}) \sim (\hat{H}^*_{1,m_1}, \hat{H}^*_{2,m_2})$ conditional on $(M_{1,n}, M_{2,n}) = (m_1, m_2)$, with $\hat{H}^*_{1,m_1}$ and $\hat{H}^*_{2,m_2}$ independent empirical distributions of random samples of sizes m_1 and m_2 from H_1 and H_2, respectively. Therefore, conditional on $(M_{1,n}, M_{2,n}) = (m_1, m_2)$

$$\begin{aligned} K_n \sim \sup_t |\hat{H}^*_{1,m_1}(t)^2 - \hat{H}^*_{2,m_2}(t)| &\sim \sup_t |(\hat{U}_{1,m_1} \circ H_1(t))^2 - \hat{U}_{2,m_2} \circ H_2(t)| \\ &= \sup_t |(\hat{U}_{1,m_1} \circ H_1(t))^2 - \hat{U}_{2,m_2} \circ H_1(t)^2| \\ &= \sup_{u\in[0,1]} |\hat{U}_{1,m_1}(u)^2 - \hat{U}_{2,m_2}(u^2)|. \quad \blacksquare \end{aligned}$$

Next, recall from Appendix A.1.4 that durations are rounded to integer days in our sample. Formally, we can only compute KS-statistics for the discretized distributions on $\{T/365, 2T/365, \ldots, T\}$. Again, the resulting statistic is not distribution-free. In the present case, in which the null hypothesis does not specify the distribution H_1 (or H_2), this complicates the computation of exact p-values. However, as argued before, the effect of the discretization is likely to

be small. To check this, we have computed (sharp) bounds on the (continuous-time) KS-statistic and its p-value imposed by observations of $\hat{H}_1$ and $\hat{H}_2$ on $\{T/365, 2T/365, \ldots, T\}$. The discrete (and reported) KS-statistic provides a sharp lower bound. In the main text, we report a statistic of 0.067 with a p-value of 0.423. Note that this p-value provides an upper bound on the p-value under continuous observation. An upper bound on the KS-statistic is easily found by including distances of $\hat{H}_1(t)^2$ and $\hat{H}_2(t')$ for t and t' one day apart in the comparison. The upper bound on the statistic is 0.070. The corresponding p-value, 0.358, provides a lower bound on the p-value under continuous observation. As expected, the bounds are narrow and justify the conclusion that the results are not affected by the discretization.

A.3 Results for Subsection 3.4.3 ($\widehat{\ln \beta_n}$, $\pi(\Psi)$, and $\pi(H_1)$)

A.3.1 Properties of $\widehat{\ln \beta_n}$ and $\hat{\pi}_n$. Let $\widehat{\ln \beta_n}(\Psi)$ be defined as $\widehat{\ln \beta_n}$ for the transformed durations $T^*_{1,i}$ and $T^*_{2,i} - T^*_{1,i}$. Note that $(T^*_{1,i}, T^*_{2,i} - T^*_{1,i})$ is uniformly distributed on $\{(t_1, t_2) : 0 \leq t_1 < 1, t_1 < t_2 \leq 1\}$ conditional on $N_i(T) = 2$ under the null. Using this, the asymptotic distribution of $\widehat{\ln \beta_n}(\Psi)$ under the null (and therefore $\widehat{\ln \beta_n}$ under the assumption of stationarity) is easy to derive. Let $\mathcal{N}(\mu, \sigma^2)$ denote a normal random variable with mean μ and variance σ^2. Then, we have

PROPOSITION 5. *Under the null $\beta = 1$,*

$$\sqrt{n}\,\widehat{\ln \beta_n}(\Psi) \Rightarrow \mathcal{N}\left(0, \frac{\pi^2}{3p_2}\right)$$

as $n \to \infty$.

PROOF. Let $\beta = 1$. Then $\mathbb{E}[(\ln(T^*_{1,i}) - \ln(T^*_{2,i} - T^*_{1,i}))I(N_i(T) = 2)] = 0$ and

$$\mathbb{E}[(\ln(T^*_{1,i}) - \ln(T^*_{2,i} - T^*_{1,i}))^2 I(N_i(T) = 2)]$$
$$= 2p_2 \int_0^1 \int_{t_1}^1 (\ln(t_1) - \ln(t_2 - t_1))^2 dt_2 dt_1 = p_2 \frac{\pi^2}{3}.$$

Also, $n^{-1}M_{2,n} \xrightarrow{\text{a.s.}} p_2$ as $n \to \infty$ by the law of large numbers. The result follows by the central limit theorem and Slutsky's lemma. ■

The distributional properties of $\hat{\pi}_n(\Psi)$ under the null (and therefore of $\hat{\pi}_n$ under the assumption of stationarity) are standard.

PROPOSITION 6. *Under the null* $\beta = 1$,

$$\Pr\left[\hat{\pi}_n(\Psi) = \frac{i}{m}\middle| M_{2,n} = m\right] = \binom{m}{i}2^{-m}$$

for $i = 0, \ldots, m$, *and*

$$\sqrt{n}\left(\hat{\pi}_n(\Psi) - \frac{1}{2}\right) \Rightarrow \mathcal{N}\left(0, \frac{1}{4p_2}\right)$$

as $n \to \infty$.

PROOF. This follows from the well-known application of the central limit theorem to the binomial distribution, $n^{-1}M_{2,n} \xrightarrow{\text{a.s.}} p_2$ as $n \to \infty$ and Slutsky's lemma. ■

Finally, we can sign $\pi(\Psi) - \frac{1}{2}$ under the alternative using that

$$\Pr(T_1^* \geq T_2^* - T_1^* | \lambda, N(T) = 2) = \left(\frac{\beta(\lambda)}{2\beta(\lambda)+1}\right)\frac{[2\beta(\lambda)+1]e^{-\lambda} - 2[\beta(\lambda)+1]e^{-(1/2)[\beta(\lambda)+1]\lambda} + e^{-\beta(\lambda)^2\lambda}}{\beta(\lambda)e^{-\lambda} - [\beta(\lambda)+1]e^{-\beta(\lambda)\lambda} + e^{-\beta(\lambda)^2\lambda}}$$

for $\beta(\lambda) \neq 1$.

A.3.2 Adapting $\hat{\pi}_n$ *to Selection on* $N_i(T) \geq 2$. Consider the alternative case in which we select on $N_i(T) \geq 2$. It is easy to check that the distribution of $(T_1^*, T_2^* - T_1^*)$ conditional on $N(T) \geq 2$, unlike that conditional on $N(T) = 2$, depends on the distribution G of λ under the null. However,

$$\Pr(T_1^* \geq T_2^* - T_1^* | \lambda, N_i(T) \geq 2) = \begin{cases} \left(\dfrac{\beta(\lambda)}{\beta(\lambda)+1}\right)\dfrac{[\beta(\lambda)+1]e^{-\lambda} - 2e^{-(1/2)[\beta(\lambda)+1]\lambda} + 1 - \beta(\lambda)}{\beta(\lambda)e^{-\lambda} - e^{-\beta(\lambda)\lambda} + 1 - \beta(\lambda)} & \text{if } \beta \neq 1, \text{ and} \\ \dfrac{1}{2} & \text{if } \beta = 1. \end{cases}$$

Under the null $\beta = 1$, we again have that $\Pr(T_1^* \geq T_2^* - T_1^* | \lambda, N(T) \geq 2) = \frac{1}{2}$ is known and independent of λ. The distributional properties of the equivalent of $\hat{\pi}_n(\Psi)$ that conditions on $N_i(T) \geq 2$ are analogous to Proposition 6, with $\sum_{k=2}^{\infty} M_{k,n}$ replacing $M_{2,n}$ and $\sum_{k=2}^{\infty} p_k$ replacing p_2. Finally,

$$\Pr(T_1^* \geq T_2^* - T_1^* | \lambda, N(T) \geq 2) = \begin{cases} < \frac{1}{2} & \text{if } \beta(\lambda) < 1, \text{ and} \\ > \frac{1}{2} & \text{if } \beta(\lambda) > 1 \end{cases}$$

for all $\lambda > 0$.

One attractive difference with selection on $N(T) = 2$ is that the baseline case without selection arises if we take the limit $T \to \infty$:

$$\lim_{T\to\infty} \Pr(T_1^* \geq T_2^* - T_1^* | \lambda, N(T) \geq 2) = \frac{\beta(\lambda)}{\beta(\lambda) + 1}.$$

A.3.3 Asymptotic Properties of $\hat{\pi}_n(\hat{H}_{1,n})$ under the Null $\beta = 1$. The main result is

PROPOSITION 7. *Under the null $\beta = 1$,*

$$\sqrt{n}\left(\hat{\pi}_n(\hat{H}_{1,n}) - \frac{1}{2}\right) \Rightarrow \mathcal{N}\left(0, \frac{1}{4p_2} + \frac{1}{6p_1}\right)$$

as $n \to \infty$.

PROOF. Let

$$\pi : h \in \mathcal{D} \mapsto \Pr(2h(T_{1,i}) \geq h(T_{2,i}) | N_i(T) = 2), \tag{5}$$

with $\mathcal{D}$ the set of all distribution functions on $[0, T]$ concentrated on $(0, T]$. Note that π does not depend on i and that $\pi(\Psi) = \frac{1}{2}$ if $\beta = 1$ (as is maintained throughout). It is convenient to write

$$\sqrt{n}(\hat{\pi}_n(\hat{H}_{1,n}) - \tfrac{1}{2}) = \sqrt{n}(\hat{\pi}_n(\hat{H}_{1,n}) - \pi(\hat{H}_{1,n})) + \sqrt{n}(\pi(\hat{H}_{1,n}) - \tfrac{1}{2})$$

and analyze the two terms in the right-hand side. Lemmas 6 and 7 below imply that

$$\sqrt{n}(\hat{\pi}_n(\hat{H}_{1,n}) - \pi(\hat{H}_{1,n})) \Rightarrow \mathcal{N}_1\left(0, \frac{1}{4p_2}\right)$$

and

$$\sqrt{n}\left(\pi(\hat{H}_{1,n}) - \frac{1}{2}\right) \Rightarrow \mathcal{N}_2\left(0, \frac{1}{6p_1}\right)$$

as $n \to \infty$, with $\mathcal{N}_1(0, 1/(4p_2))$ and $\mathcal{N}_2(0, 1/(6p_1))$ independent normal random variables. The claimed result follows. ■

It remains to state and prove Lemmas 6 and 7.

LEMMA 6. *Under the null $\beta = 1$,*

$$\sqrt{n}(\hat{\pi}_n(\hat{H}_{1,n}) - \pi(\hat{H}_{1,n})) \Rightarrow \mathcal{N}\left(0, \frac{1}{4p_2}\right)$$

conditional on $\{\hat{H}_{1,i}\}$ almost surely as $n \to \infty$.

PROOF. Denote $Y_{n,i} := I(2\hat{H}_{1,n}(T_{1,i}) \geq \hat{H}_{1,n}(T_{2,i}), N_i(T) = 2)$. Note that

$$\mathbb{E}[Y_{n,i}|\{N_j(T),\hat{H}_{1,j}\}] = I(N_i(T)=2)\pi(\hat{H}_{1,n}) \text{ and}$$
$$\operatorname{var}(Y_{n,i}|\{N_j(T),\hat{H}_{1,j}\}) = I(N_i(T)=2)\pi(\hat{H}_{1,n})(1-\pi(\hat{H}_{1,n}))$$

almost surely, that π is continuous at Ψ, that $\sup|\hat{H}_{1,n} - \Psi| \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$ (Lemma 3) and that $n^{-1}M_{2,n} \xrightarrow{\text{a.s.}} p_2$ as $n \to \infty$. It follows that for all $\varepsilon > 0$

$$0 \le \sum_{i=1}^{n} \mathbb{E}\left[\left(\sqrt{n}\,\frac{Y_{n,i}}{M_{2,n}}\right)^2 I\left(\sqrt{n}\,\frac{Y_{n,i}}{M_{2,n}} > \varepsilon\right)\middle|\{N_j(T),\hat{H}_{1,j}\}\right]$$
$$\le \left(\frac{n}{M_{2,n}}\right)^2 n^{-1}\sum_{i=1}^{n}\Pr\left(Y_{n,i} > \sqrt{n}\,\varepsilon\,\frac{M_{2,n}}{n}\middle|\{N_j(T),\hat{H}_{1,j}\}\right) \to 0$$

and

$$\sum_{i=1}^{n}\operatorname{var}\left(\sqrt{n}\,\frac{Y_{n,i}}{M_{2,n}}\middle|\{N_j(T),\hat{H}_{1,j}\}\right) = \frac{n}{M_{2,n}}\pi(\hat{H}_{1,n})(1-\pi(\hat{H}_{1,n})) \to \frac{\pi(\Psi)(1-\pi(\Psi))}{p_2}$$
$$= \frac{1}{4p_2}$$

almost surely as $n \to \infty$. As a consequence, by the Lindeberg-Feller central limit theorem (Van der Vaart 1998, Proposition 2.27)

$$\sqrt{n}\sum_{i=1}^{n}\left(\frac{Y_{n,i}}{M_{2,n}} - \pi(\hat{H}_{1,n})\right) \Rightarrow \mathcal{N}\left(0, \frac{1}{4p_2}\right)$$

conditional on $\{N_j(T), \hat{H}_{1,j}\}$ almost surely as $n \to \infty$. Note that the limit does not depend on $\{N_j(T), \hat{H}_{1,j}\}$ and that

$$\hat{\pi}_n(\hat{H}_{1,n}) = \sum_{i=1}^{n}\frac{Y_{n,i}}{M_{2,n}}.$$

The claimed result follows. ■

LEMMA 7. *Under the null* $\beta = 1$,

$$\sqrt{n}\left(\pi(\hat{H}_{1,n}) - \frac{1}{2}\right) \Rightarrow \mathcal{N}\left(0, \frac{1}{6p_1}\right)$$

as $n \to \infty$.

PROOF. We need some notation and conventions. Let $\mathcal{B}(Z)$ be the set of bounded $\mathbb{R}$-valued functions on $Z \subset \mathbb{R}$. We abbreviate $\mathcal{B}[0, T] := \mathcal{B}([0, T])$, etc. We equip $\mathcal{B}(Z)$ with the uniform norm and the product space $\mathcal{B}(Z) \times \mathcal{B}(Z)$ with the

norm $(g_1, g_2) \in (\mathcal{B}(Z) \times \mathcal{B}(Z)) \mapsto \max\{\sup|g_1|, \sup|g_2|\}$. All subsets of $\mathcal{B}(Z)$ and $\mathcal{B}(Z) \times \mathcal{B}(Z)$ inherit the corresponding metrics. Multiplication of elements of $\mathcal{B}(Z)$ is defined as point-wise multiplication. $\mathcal{C}(Z)$ is the set of uniformly continuous functions in $\mathcal{B}(Z)$. Throughout, we take $\beta = 1$.

Below, we show that the Hadamard derivative of π at Ψ tangentially to $\mathcal{C}[0, T]$ is

$$\pi'_\Psi : u \in \mathcal{C}[0, T] \mapsto 2 \int_0^{1/2} [2u(\Psi^{-1}(z)) - u(\Psi^{-1}(2z))]dz.$$

With the facts that $\sqrt{n}(\hat{H}_{1,n} - \Psi) \Rightarrow (1/\sqrt{p_1})\mathbb{G}_\Psi$ (Lemma 3) and that $\mathbb{G}_\Psi$ assumes values in $\mathcal{C}[0, T]$, the functional Delta method (Van der Vaart 1998, Theorem 20.8) implies that

$$\begin{aligned}\sqrt{n}(\pi(\hat{H}_{1,m}) - \pi(\Psi)) \Rightarrow \frac{1}{\sqrt{p_1}}\pi'_\Psi(\mathbb{G}_\Psi) &= \frac{2}{\sqrt{p_1}}\int_0^{1/2} [2\mathbb{G}_\Psi(\Psi^{-1}(z)) \\ &\quad - \mathbb{G}_\Psi(\Psi^{-1}(2z))]dz \\ &= \frac{2}{\sqrt{p_1}}\int_0^{1/2} [2\mathbb{G}_U(z) - \mathbb{G}_U(2z)]dz \\ &= \mathcal{N}\left(0, \frac{1}{6p_1}\right)\end{aligned}$$

as $n \to \infty$.

It remains to show that π is Hadamard differentiable at Ψ tangentially to $\mathcal{C}[0, T]$ with the derivative π'_Ψ given above. Note that the Lebesgue density of $(T_1, T_2)|(N(T) = 2)$ at (t_1, t_2) equals $2\psi(t_1)\psi(t_2)$ if $0 \le t_1 < t_2 \le T$ and 0 otherwise. Using this, we can rewrite (5) as

$$\pi(h) = 2\int_0^T \left[\int_{h^{-1}(h(t)/2)}^t \psi(\tau)d\tau\right]\psi(t)dt = 1 - 2\int \left(\Psi \circ h^{-1} \circ \left(\frac{h}{2}\right)\right)d\Psi.$$

where $h^{-1}(p) = \inf\{t : h(t) \ge p\}$ is the generalized inverse of h. The functional $\pi : \mathcal{D} \to \mathbb{R}$ can be decomposed as

$$\begin{aligned}h \in \mathcal{D} &\overset{\rho}{\mapsto} \left(\frac{h}{2}, h^{-1}\right) \in \mathcal{D}_1 \times \mathcal{I} \\ &\overset{\sigma}{\mapsto} h^{-1} \circ \left(\frac{h}{2}\right) \in \mathcal{D}_T \overset{\tau}{\to} \Psi \circ \left(h^{-1} \circ \left(\frac{h}{2}\right)\right) \in \mathcal{B}[0, T] \\ &\overset{\upsilon}{\mapsto} 1 - 2\int \left(\Psi \circ h^{-1} \circ \left(\frac{h}{2}\right)\right)d\Psi \in \mathbb{R}.\end{aligned}$$

Recall that $\mathcal{D} \subset \mathcal{B}[0, T]$ is the set of distribution functions on $[0, T]$ concentrated on $(0, T]$. Also, $\mathcal{D}_1 \subset \mathcal{B}[0, T]$ and $\mathcal{D}_T \subset \mathcal{B}[0, T]$ are the sets of, respectively, $[0, 1)$-valued and $[0, T]$-valued functions on $[0, T]$ and $\mathcal{I} \subset \mathcal{B}[0, 1)$ is the set of $[0, T]$-valued functions on $[0, 1)$.

We first show that each of the maps ρ, σ, τ and υ is (tangentially) Hadamard differentiable and then derive π'_Ψ by the chain rule.

(i). *Hadamard derivative of* $\rho : \mathcal{D} \to \mathcal{D}_1 \times \mathcal{I}$ *at* Ψ *tangentially to* $\mathcal{C}[0, T]$
By Van der Vaart (1998), Lemma 21.4, the inverse map $h \in \mathcal{D} \subset \mathcal{B}[0, T] \mapsto h^{-1} \in \mathcal{I}$ is Hadamard differentiable at Ψ tangentially to $\mathcal{C}[0, T]$ with derivative $u \in \mathcal{C}[0, T] \mapsto -(u/\psi) \circ \Psi^{-1}$. Thus, the Hadamard derivative of ρ at Ψ tangentially to $\mathcal{C}[0, T]$ is

$$\rho'_\Psi : u \in \mathcal{C}[0, T] \mapsto \left(\frac{u}{2}, -\left(\frac{u}{\psi}\right) \circ \Psi^{-1}\right) \in \mathcal{B}[0, T] \times \mathcal{C}[0, 1).$$

(ii). *Hadamard derivative of* $\sigma : \mathcal{D}_1 \times \mathcal{I} \to \mathcal{D}_T$ *at* $\rho(\Psi) = (\Psi/2, \Psi^{-1})$ *tangentially to* $\mathcal{B}[0, T] \times \mathcal{C}[0, 1)$
By the assumption that Ψ is continuously differentiable with positive derivative ϕ on $[0, T]$, Ψ^{-1} is uniformly differentiable on $[0, 1)$ with uniformly bounded derivative $(1/\psi) \circ \Psi^{-1}$. Thus, by Van der Vaart and Wellner (1996), Lemma 3.9.27, the Hadamard derivative of σ at $(\Psi/2, \Psi^{-1})$ tangentially to $\mathcal{B}[0, T] \times \mathcal{C}[0, 1)$ is[34]

34. It may be helpful here to construct the proof of this lemma for our special case. Let $(u_q, v_q) \to (u, v) \in \mathcal{B}[0, T] \times \mathcal{C}[0, 1)$ uniformly as $q \downarrow 0$ and $\rho(\Psi) + q(u_q, v_q) \in \mathcal{D}_1 \times \mathcal{I}$ for all q. We have that

$$0 \leq \left| \frac{(\Psi^{-1} + qv_q) \circ \left(\frac{\Psi}{2} + qu_q\right) - \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)}{q} - v \circ \left(\frac{\Psi}{2}\right) - \frac{u}{\psi \circ \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)} \right|$$

$$\leq \left| \frac{\Psi^{-1} \circ \left(\frac{\Psi}{2} + qu_q\right) - \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)}{q} - \frac{u_q}{\psi \circ \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)} \right|$$

$$+ \left| \frac{u_q - u}{\psi \circ \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)} \right| + \left| v \circ \left(\frac{\Psi}{2} + qu_q\right) - v \circ \left(\frac{\Psi}{2}\right) \right| + \left| (v_q - v) \circ \left(\frac{\Psi}{2} + qu_q\right) \right|.$$

The first term in the right-hand side converges (uniformly) to 0 because of the uniform differentiability of Ψ^{-1}. The second term converges to 0 because $\sup|u_q - u| \to 0$ and the denominator is bounded away from 0. The third term converges to 0 because v is uniformly continuous. The fourth term converges to 0 because $\sup|v_q - v| \to 0$.

$$\sigma'_{\rho(\Psi)} : (u, v) \in \mathcal{B}[0, T] \times \mathcal{C}[0, 1) \mapsto v \circ \left(\frac{\Psi}{2}\right) + \frac{u}{\psi \circ \Psi^{-1} \circ \left(\frac{\Psi}{2}\right)} \in \mathcal{B}[0, T].$$

(iii). *Hadamard derivative of* $\tau : \mathcal{D}_T \to \mathcal{B}[0, T]$ *at* $\sigma(\rho(\Psi)) = \Psi^{-1} \circ (\Psi/2)$
Let $u_q \to u \in \mathcal{B}[0, T]$ uniformly as $q \downarrow 0$ and $\sigma(\rho(\Psi)) + qu_q \in \mathcal{D}_T$ for all q. Abbreviate $s := \sigma(\rho(\Psi))$ and consider

$$\begin{aligned} 0 &\leq \left| \frac{\tau(s + qu_q) - \tau(s)}{q} - (\psi \circ s)u \right| \\ &= \left| \frac{\Psi \circ (s + qu_q) - \Psi \circ s}{q} - (\psi \circ s)u_q + (\psi \circ s)(u_q - u) \right| \\ &\leq \left| \frac{\Psi \circ (s + qu_q) - \Psi \circ s}{q} - (\psi \circ s)u_q \right| + (\psi \circ s)|u_q - u|. \end{aligned}$$

The first term in the last line converges uniformly to 0 as $q \downarrow 0$ by the uniform differentiability of Ψ. The second term converges uniformly to 0 because ψ is uniformly bounded. Thus, the right-hand side of the first line converges uniformly to 0 as $q \downarrow 0$ and the Hadamard derivative of $\tau : \mathcal{D}_T \to \mathcal{B}[0, T]$ at $\sigma(\rho(\Psi)) = \Psi^{-1} \circ (\Psi/2)$ is

$$\tau'_{\sigma(\rho(\Psi))} : u \in \mathcal{B}[0, T] \mapsto \left(\psi \circ \Psi^{-1} \circ \left(\frac{\Psi}{2}\right) \right) u \in \mathcal{B}[0, T].$$

(iv). *Hadamard derivative of* $v : \mathcal{B}[0, T] \to \mathbb{R}$ *at* $\tau(\sigma(\rho(\Psi))) = \Psi/2$
Let $u_q \to u \in \mathcal{B}[0, T]$ uniformly as $q \downarrow 0$ and $\tau(\sigma(\rho(\Psi))) + qu_q \in \mathcal{B}[0, T]$ for all q. Then,

$$\frac{v(\Psi/2 + qu_q) - v(\Psi/2)}{q} = -2 \int u_q d\Psi \to -2 \int u d\Psi$$

as $q \downarrow 0$ because the sequence u_q is uniformly bounded. So, the Hadamard derivative of $v : \mathcal{B}[0, T] \to \mathbb{R}$ at $\tau(\sigma(\rho(\Psi))) = \Psi/2$ is

$$v'_{\tau(\sigma(\rho(\Psi)))} : u \in \mathcal{B}[0, T] \mapsto -2 \int u d\Psi.$$

By the chain rule (Van der Vaart 1998, Proposition 20.9), the Hadamard derivative of $\pi = v \circ \tau \circ \sigma \circ \rho$ at Ψ tangentially to $\mathcal{C}[0, T]$ is $v'_{\tau(\sigma(\rho(\Psi)))} \circ \tau'_{\sigma(\rho(\Psi))}$

$\circ\, \sigma'_{\rho(\Psi)} \circ \rho'_{\Psi}$ (note that the tangent spaces are properly lined up). Substitution of the derivative maps derived in (i)–(iv) gives the desired result. ■

A.3.4 The Behavior of $\pi(H_1)$ under Local Alternatives to $\beta = 1$. This appendix provides details on the directional derivatives of $\pi(\Psi;\, \beta)$ and $\pi(H_1(\beta);\, \beta)$ at $\beta = 1$ in the direction u.

First consider $\pi(\Psi;\, \beta)$. We have earlier seen that the Lebesgue density of $(T_1,\, T_2)|(N(T) = 2)$ at $(t_1,\, t_2)$ equals $2\psi(t_1)\psi(t_2)$ if $0 \leq t_1 < t_2 \leq \mathrm{T}$ and 0 otherwise if $\beta = 1$. For $\beta(\lambda) \neq 1$, the density of $(T_1,\, T_2)|(\lambda,\, N(T) = 2)$ at $(t_1,\, t_2)$ equals

$$\frac{\lambda^2\beta(\lambda)\psi(t_1)\psi(t_2)\mathrm{e}^{-\lambda[1-\beta(\lambda)]\Psi(t_1)-\lambda\beta(\lambda)[1-\beta(\lambda)]\Psi(t_2)-\lambda\beta(\lambda)^2}}{\int_0^T\int_0^{\tau_2}\lambda^2\beta(\lambda)\psi(\tau_1)\psi(\tau_2)\mathrm{e}^{-\lambda[1-\beta(\lambda)]\Psi(\tau_1)-\lambda\beta(\lambda)[1-\beta(\lambda)]\Psi(\tau_2)-\lambda\beta(\lambda)^2}d\tau_1 d\tau_2}$$

for $0 \leq t_1 < t_2 \leq T$ and 0 elsewhere. Thus, for $\beta = 1 + qu$ and $q \neq 0$ we have that

$$\pi(\Psi;\, \beta) = \frac{\int\int_0^1\int_{\tau_2/2}^{\tau_2}\lambda^2[1 + qu(\lambda)]\mathrm{e}^{\lambda qu(\lambda)\tau_1+\lambda[1+qu(\lambda)]qu(\lambda)\tau_2-\lambda[1+qu(\lambda)]^2}d\tau_1 d\tau_2 dG(\lambda)}{\int\int_0^1\int_0^{\tau_2}\lambda^2[1 + qu(\lambda)]\mathrm{e}^{\lambda qu(\lambda)\tau_1+\lambda[1+qu(\lambda)]qu(\lambda)\tau_2-\lambda[1+qu(\lambda)]^2}d\tau_1 d\tau_2 dG(\lambda)}$$

Using that

$$\frac{d}{dq}\left[\int\int_0^1\int_{\tau_2/2}^{\tau_2}\lambda^2[1 + qu(\lambda)]\mathrm{e}^{\lambda qu(\lambda)\tau_1+\lambda[1+qu(\lambda)]qu(\lambda)\tau_2-\lambda[1+qu(\lambda)]^2}d\tau_1 d\tau_2 dG(\lambda)\right]_{q=0} = \frac{5}{24}\mathcal{L}'''_u(1) + \frac{1}{4}\mathcal{L}''_u(1)$$

and that

$$\frac{d}{dq}\left[\int\int_0^1\int_0^{\tau_2}\lambda^2[1 + qu(\lambda)]\mathrm{e}^{\lambda qu(\lambda)\tau_1+\lambda[1+qu(\lambda)]qu(\lambda)\tau_2-\lambda[1+qu(\lambda)]^2}d\tau_1 d\tau_2 dG(\lambda)\right]_{q=0} = \frac{1}{2}\mathcal{L}'''_u(1) + \frac{1}{2}\mathcal{L}''_u(1)$$

we find that

$$\frac{d}{dq}\left[\pi(\Psi;\,1+qu)\right]_{q=0}=\frac{1}{12}\,\frac{-\mathscr{L}_u'''(1)}{\mathscr{L}''(1)}. \tag{G2}$$

Next, consider $\pi(H_1(\beta);\,\beta)$. The analysis in A.2.1 implies that

$$\frac{d}{dq}\left[H_1(1+qu)\right]_{q=0}=-\frac{1}{2}\,\frac{\mathscr{L}_u''(1)}{-\mathscr{L}'(1)}\,\Psi(1-\Psi)$$

so that

$$\frac{d}{dq}\left[\pi(H_1(1+qu);\,1)\right]_{q=0}=\pi'_\Psi\left(-\frac{1}{2}\,\frac{\mathscr{L}_u''(1)}{-\mathscr{L}'(1)}\,\Psi(1-\Psi)\right)=-\frac{1}{12}\,\frac{\mathscr{L}_u''(1)}{-\mathscr{L}'(1)}.$$

Thus, it follows that

$$\begin{aligned}\frac{d}{dq}\left[\pi(H_1(1+qu);\,1+qu)\right]_{q=0}&=\frac{d}{dq}\left[\pi(\Psi;\,1+qu)\right]_{q=0}\\&\quad+\frac{d}{dq}\left[\pi(H_1(1+qu);\,1)\right]_{q=0}=\frac{1}{12}\left[\frac{-\mathscr{L}_u'''(1)}{\mathscr{L}''(1)}-\frac{\mathscr{L}_u''(1)}{-\mathscr{L}'(1)}\right].\end{aligned} \tag{G3}$$

B. Identification and Estimation

This appendix provides results for Subsection 3.5. Note that the identification and estimation analyses in this subsection assume that β is homogeneous, as in (M^+).

B.1 Results for Subsection 3.5.1 (Identification)

Proof of Proposition 2. Denote the subdensity of $(T_1,\ldots,T_k)$ on $N(T)=k$ by f_k (i.e., $f_1(t)=d\Pr(T_1\le t,N(T)=1)/dt$, etc.). We first show that the cases $\beta=1$ and $\beta\neq 1$ can be distinguished from data on f_1 and f_2. With the normalization $\Psi(T)=1$, we have that

$$\frac{f_2(t_1,\,t_2)}{f_1(t_1)\,f_1(t_2)}=\beta\,\frac{\mathscr{L}''((1-\beta)(\Psi(t_1)+\beta\Psi(t_2))+\beta^2)}{\mathscr{L}'((1-\beta)\Psi(t_1)+\beta)\mathscr{L}'((1-\beta)\Psi(t_2)+\beta)}.$$

Straightforward calculations show that

$$\left[\frac{\partial\,\ln\left(\dfrac{f_2(t_1,\,t_2)}{f_1(t_1)\,f_1(t_2)}\right)}{\partial t_1}-\frac{\partial\,\ln\left(\dfrac{f_2(t_1,\,t_2)}{f_1(t_1)\,f_1(t_2)}\right)}{\partial t_2}\right]_{t_1=t_2=t}$$

$$=(1-\beta)^2\psi(t)\,\frac{\mathscr{L}'''((1-\beta^2)\Psi(t)+\beta^2)}{\mathscr{L}''((1-\beta^2)\Psi(t)+\beta^2)},$$

which equals 0 (for all t) if and only if $\beta=1$. This establishes whether $\beta=1$.

In the case that $\beta \neq 1$, it remains to distinguish between $\beta < 1$ and $\beta > 1$. To this end, note that the claim intensity at t conditional on $(T_1 = t_1, T_2 \geq t)$, $t > t_1$, is given by

$$\theta(t|T_1 = t_1, T_2 \geq t) = \beta\psi(t)\,\frac{\mathscr{L}''((1-\beta)\Psi(t_1) + \beta\Psi(t))}{-\mathscr{L}'((1-\beta)\Psi(t_1) + \beta\Psi(t))}.$$

Lemma 4 in Appendix A.2 implies that the function $\mathscr{L}''/(-\mathscr{L}')$ is trivial if λ either is degenerate or has two points of support of which one is 0. Otherwise, $\mathscr{L}''/(-\mathscr{L}')$ is (strictly) decreasing. Clearly, if $\theta(t|T_1 = t_1, T_2 \geq t)$ is decreasing in t_1, we know that $\beta < 1$. Similarly, if $\theta(t|T_1 = t_1, T_2 \geq t)$ is increasing in t_1 then we can conclude that $\beta > 1$. If $\theta(t|T_1 = t_1, T_2 \geq t)$ is constant in t_1, we know (because $\beta \neq 1$) that λ either is degenerate or has two points of support of which one is 0. In that case,

$$\mathscr{L}(s) = \Pr(\lambda = 0) + (1 - \Pr(\lambda = 0))\exp(-\mathbb{E}[\lambda|\lambda > 0]s)$$

and $\theta(t|T_1 = t_1, T_2 \geq t) = \mathbb{E}[\lambda|\lambda > 0]\beta\psi(t)$. With the normalization $\Psi(T) = 1$, this identifies $\mathbb{E}[\lambda|\lambda > 0]\beta$ and Ψ. Then,

$$\Pr(T_1 > t) = \mathscr{L}(\Psi(t)) = \Pr(\lambda = 0) + (1 - \Pr(\lambda = 0))\exp(-\mathbb{E}[\lambda|\lambda > 0]\Psi(t))$$

identifies $\Pr(\lambda = 0)$ and $\mathbb{E}[\lambda|\lambda > 0]$ and therewith β. ■

Note that the proof only uses data on first and second claim times (that is, the sub-distributions of T_1 on $N(T) = 1$ and T_2 on $N(T) = 2$ and the claim intensities at 0 and 1 claim). A central role is played by the way $\theta(t|T_1 = t_1, T_2 \geq t)$ depends on t_1. Conditional on λ however, the claim intensity only depends on the occurrence, and not the timing, of past claims. Thus, the dependence of $\theta(t|T_1 = t_1, T_2 \geq t)$ on t_1 works by way of the heterogeneity. This explains that a special role is played by the cases that λ is degenerate and that λ has two points of support of which one is 0. In either case, λ is degenerate at the non-zero point of support $\mathbb{E}[\lambda|\lambda > 0]$ conditional on past occurrence of a claim and there is no heterogeneity conditional on $(T_1 = t_1, T_2 \geq t)$. We have seen these cases appearing in the analyses of the behavior of the two general tests in Subsections 3.4.2 and 3.4.3 under homogeneous-β local alternatives. Note that these alternatives fit the special model ($M^{\dagger}$) that we are studying here.

Next, consider the key identifying equation (I). Recall that $q_0(t) = \mathscr{L}(\Psi(t))$ and note that

$$q_1(t) = \frac{\mathscr{L}(\beta\Psi(t)) - \mathscr{L}(\Psi(t))}{1-\beta} \qquad \text{and}$$

$$q_2(t) = \frac{\beta\mathscr{L}(\Psi(t)) - (1+\beta)\mathscr{L}(\beta\Psi(t)) + \mathscr{L}(\beta^2\Psi(t))}{(1-\beta^2)(1-\beta)}.$$

if $\beta \neq 1$. Also, define $\tilde{q}(t) := (1-\beta)q_1(t) + q_0(t) = \mathscr{L}(\beta\Psi(t))$. If $\beta \neq 1$ then

$$q_1\{q_0^{-1}[\tilde{q}(t)]\} = \frac{\mathcal{L}(\beta^2\Psi(t)) - \mathcal{L}(\beta\Psi(t))}{1-\beta},$$

which indeed equals $\beta q_1(t) + (1-\beta^2)q_2(t)$.

Proof of Proposition 3. The case $\beta = 1$ has been covered in Subsection 3.4.1. So, consider the case that $\beta \neq 1$. If we know β we can compute the function $\tilde{q}$ defined above. The remainder of the proof closely follows Kortram et al. (1995), in particular a version thereof by Abbring (2002) that can directly be applied to our finite-support setup here. Suppose that $\beta > 1$ (the case $\beta < 1$ is similar). We have that, for any $t \in [0, T]$, $y := \tilde{q}(t) = \mathcal{L}(\beta\Psi(t)) \Leftrightarrow \Psi(t) = \beta^{-1}\mathcal{L}^{-1}(y)$. Also, $t = \tilde{q}^{-1}(y)$, so that

$$q_0(\tilde{q}^{-1}(y)) = q_0(t) = \mathcal{L}(\Psi(t)) = \mathcal{L}(\beta^{-1}\mathcal{L}^{-1}(y)).$$

By induction, it follows that, for $y \in [\tilde{q}(T), 1]$,

$$(q_0 \circ \tilde{q}^{-1})^{(n)}(y) = \mathcal{L}(\beta^{-n}\mathcal{L}^{-1}(y)), \tag{6}$$

where $q_0 \circ \tilde{q}^{-1}$ is the composition of q_0 with $\tilde{q}^{-1}$, and superscript (n) denotes the n-fold composition. Now, l'Hôspital's rule gives

$$\mathbb{E}[\lambda]\mathcal{L}^{-1}(y) = \lim_{n\to\infty} \frac{1 - \mathcal{L}(\beta^{-n}\mathcal{L}^{-1}(y))}{\beta^{-n}} = \lim_{n\to\infty} \frac{(1 - (q_0 \circ \tilde{q}^{-1})^{(n)}(y))}{\beta^{-n}} \tag{7}$$

on $[\tilde{q}(T), 1]$. Evaluating (7) at $q_0(T) \in [\tilde{q}(T), 1]$ gives $\mathbb{E}[\lambda]$, because $\mathcal{L}^{-1}(q_0(T)) = \Psi(T) = 1$. So, $\mathcal{L}^{-1}$ is uniquely determined on $[\tilde{q}(T), 1]$. We then have that $\Psi(t) = \mathcal{L}^{-1}(q_0(t)) = \beta^{-1}\mathcal{L}^{-1}(\tilde{q}(t))$ on $[0, T]$. From $\mathcal{L}^{-1}$ on $[\tilde{q}(T), 1]$ we can identify $\mathcal{L}$ on $[0, \mathcal{L}^{-1}(\tilde{q}(T)] = [0, \beta\Psi(T)]$. Finally, $\mathcal{L}$ can be analytically extended to $[0, \infty)$. ■

B.2 Results for Subsection 3.5.2 (Estimation)

In this appendix we construct the likelihood on which the estimates in Subsection 3.5.2 are based. Choose some marginal density $g(\cdot\,; \xi)$ of λ_i and some contract-time function $\psi(\cdot\,; \alpha)$. As before, we define $\Psi(t; \alpha) := \int_0^t \psi(u; \alpha)du$ and make sure that the normalization $\Psi(T; \alpha) = 1$ is satisfied. Both ξ and α are finite-dimensional parameter vectors.

Consider the likelihood contribution L_i of contract i. For contract i, we observe a claim history $N_i[0, T]$ (we ignore the discretization of time in days here). First consider the likelihood contribution for the case that we observe λ_i. Now, recall that $T_{k,i}$ is the time of the k-th claim and let $T_{0,i} := 0$. For contract i, we observe the number of claims $N_i(T)$ in the contract year, and, if $N_i(T) \geq 1$, the times $T_{1,i}, \ldots, T_{N_i(T),i}$ of these claims. Straightforward calculations show that this "full information" likelihood contribution of contract i is

$$L_i(\beta, \alpha, \xi; N_i[0, T], \lambda_i) = \left[\prod_{k=1}^{N_i(T)} \lambda_i \beta^{k-1} \psi(T_{k,i}; \alpha) e^{-\lambda_i \beta^{k-1}[\Psi(T_{k,i};\alpha) - \Psi(T_{k-1,i};\alpha)]} \right] \times e^{-\lambda_i \beta^{N_i(T)}[1 - \Psi(T_{N_i(T),i};\alpha)]} g(\lambda_i; \xi)$$

if $N_i(T) \geq 1$, and

$$L_i(\beta, \alpha, \xi; N_i[0, T], \lambda_i) = e^{-\lambda_i} g(\lambda_i; \xi)$$

if $N_i(T) = 0$. Thus, the marginal likelihood contribution for the case we do not observe λ_i is

$$L_i(\beta, \alpha, \xi; N_i[0, T]) = \left[\prod_{k=1}^{N_i(T)} \beta^{k-1} \psi(T_{k,i}; \alpha) \right] (-1)^{N_i(T)} \mathcal{L}^{(N_i(T))}[\beta^{N_i(T)}[1 - \Psi(T_{N_i(T),i}; \alpha)] + \sum_{k=1}^{N_i(T)} \beta^{k-1}[\Psi(T_{k,i}; \alpha) - \Psi(T_{k-1,i}; \alpha)]; \xi]$$

if $N_i(T) \geq 1$, and

$$L_i(\beta, \alpha, \xi; N_i[0, T]) = \mathcal{L}(1; \xi)$$

if $N_i(T) = 0$. Here, $\mathcal{L}^{(k)}(\cdot\,; \xi)$ is the k-th derivative of the Laplace transform of $g(\cdot\,; \xi)$.

References

Abbring, J. H. (2002). "Stayers Versus Defecting Movers: A Note on the Identification of Defective Duration Models." *Economics Letters*, 74, pp. 327–331.

Araujo, A. and H. Moreira (2002). "Adverse Selection Without the Single Crossing Condition," mimeo. Rio de Janeiro, Brazil: Getulio Vargas Foundation.

Bates, G. E. and J. Neyman (1952). "Contributions to the Theory of Accident Proneness II: True or False Contagion." *University of California Publications in Statistics*, 1, pp. 255–275.

Cardon, J. and I. Hendel (1998). "Asymmetric Information in Health Insurance: Evidence from the National Health Expenditure Survey," mimeo. Princeton, New Jersey: Princeton University.

Chamberlain, G. (1985). "Heterogeneity, Omitted Variable Bias, and Duration Dependence." In *Longitudinal Analysis of Labor Market Data*, edited by J. J. Heckman and B. Singer. Cambridge, England: Cambridge University Press.

Chassagnon, A. and P. A. Chiappori (2000). "Insurance under Moral Hazard and Adverse Selection: The Case of Pure Competition," mimeo. Paris, DELTA.

Chiappori, P. A. (2000). "Econometric Models of Insurance under Asymmetric Information." In *Handbook of Insurance*, edited by G. Dionne. Boston: Kluwer Academic Publishers.

Chiappori, P. A., F. Durand, and P. Y. Geoffard (1998). "Moral Hazard and the Demand for Physician Services: First Lessons from a French Natural Experiment." *European Economic Review*, 42, pp. 499–511.

Chiappori, P. A., P. Y. Geoffard, and E. Kyriadizou (1998). "Cost of Time, Moral Hazard, and the Demand for Physician Services," mimeo. Chicago, Illinois: University of Chicago.
Chiappori, P. A., B. Jullien, B. Salanié, and F. Salanié (2001). "Asymmetric Information in Insurance: A Robust Characterization," mimeo. Chicago, Illinois: University of Chicago.
Chiappori, P. A., I. Macho, P. Rey, and B. Salanié (1994). "Repeated Moral Hazard: Memory, Commitment, and the Access to Credit Markets." *European Economic Review*, 38, pp. 1527–1553.
Chiappori, P. A. and B. Salanié (1997). "Empirical Contract Theory: The Case of Insurance Data." *European Economic Review*, 41, pp. 943–950.
Chiappori, P. A. and B. Salanié (2000). "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy*, 108, pp. 56–78.
Chiappori, P. A., B. Salanié, and J. Valentin (1999). "Early Starters vs Late Beginners." *Journal of Political Economy*, 107, pp. 731–760.
Dionne, G. and N. Doherty (1994). "Adverse Selection, Commitment and Renegotiation: Extension to and Evidence from Insurance Markets." *Journal of Political Economy*, 102, pp. 210–235.
Dionne, G., C. Gouriéroux, and C. Vanasse (1999). "Evidence of Adverse Selection in Automobile Insurance Markets." In *Automobile Insurance: Road Safety, Insurance Fraud and Regulation*, edited by G. Dionne and C. Laberge-Nadeau. Boston: Kluwer Academic Publishers.
Dionne, G., C. Gouriéroux, and C. Vanasse (2001). "Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment." *Journal of Political Economy*, 109, pp. 444–453.
Dionne, G., M. Maurice, J. Pinquet, and C. Vanasse (2001). "The Role of Memory in Long-Term Contracting: Empirical Evidence in Automobile Insurance," mimeo. Montreal: Hautes Etudes Commerciales (HEC).
Dionne, G. and C. Vanasse (1992). "Automobile Insurance Ratemaking in the Presence of Asymmetrical Information." *Journal of Applied Econometrics*, 7, pp. 149–165.
Elbers, C. and G. Ridder (1982). "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model." *Review of Economic Studies*, 64, pp. 403–409.
Feller, W. (1943). "On a General Class of 'Contagious' Distributions." *Annals of Mathematical Statistics*, 14, pp. 389–400.
Finkelstein, A. and J. Poterba (2002). "Adverse Selection in Insurance Markets: Policy-holder Evidence from the U.K. Annuity Market," mimeo. Cambridge, Massachusetts: Harvard University.
de Garidel, T. (1997). "Pareto Improving Asymmetric Information in a Dynamic Insurance Market." Discussion Paper 266, London School of Economics, London.
Heckman, J. J. (1981). "Heterogeneity and State Dependence." In *Studies in Labor Markets*, edited by S. Rosen. Chicago, Illinois: University of Chicago Press.
Heckman, J. J. and G. J. Borjas (1980). "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence." *Economica*, 47, pp. 247–283.
Heckman, J. J., R. Robb, and J. R. Walker (1990). "Testing the Mixture of Exponentials Hypothesis and Estimating the Mixing Distribution by the Method of Moments." *Journal of the American Statistical Association*, 85, pp. 582–589.
Hendel, I. and A. Lizzeri (1999). "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance." Working Paper, Princeton University, Princeton, New Jersey.
Holly, A., L. Gardiol, G. Domenighetti, and B. Bisig (1998). "An Econometric Model of Health Care Utilization and Health Insurance in Switzerland." *European Economic Review*, 42, pp. 513–522.
Holt, J. D. and R. L. Prentice (1974). "Survival Analysis in Twin Studies and Matched Pair Experiments." *Biometrika*, 61, pp. 17–30.

Jullien, B., B. Salanié, and F. Salanié (1999). "Should More Risk-Averse Agents Exert More Effort?" *Geneva Papers on Risk and Insurance Theory*, 24, pp. 19–28.

Kortram, R. A., A. J. Lenstra, G. Ridder, and A. C. M. van Rooij (1995). "Constructive Identification of the Mixed Proportional Hazards Model." *Statistica Neerlandica*, 49, pp. 269–281.

Mossin, J. (1968). "Aspects of Rational Insurance Purchasing." *Journal of Political Economy*, 76, pp. 553–568.

Puelz, R. and A. Snow (1994). "Evidence on Adverse Selection: Equilibrium Signalling and Cross-subsidization in the Insurance Market." *Journal of Political Economy*, 102, pp. 236–257.

Richaudeau, D. (1999). "Automobile Insurance Contracts and Risk of Accident: An Empirical Test Using French Individual Data." *The Geneva Papers on Risk and Insurance Theory*, 24, pp. 97–114.

Ridder, G. (1990). "The Non-Parametric Identification of Generalized Accelerated Failure-Time Models." *Review of Economic Studies*, 57, pp. 167–182.

Ridder, G. and I. Tunalı (1999). "Stratified Partial Likelihood Estimation." *Journal of Econometrics*, 92, pp. 193–232.

Rochet, Jean-Charles and L. Stole (2000). "The Economics of Multidimensional Screening." In *Advances in Economic Theory, 7th World Congress of the Econometric Society*, Seattle, Washington.

Shohat, J. A. and J. D. Tamarkin (1943). *The Problem of Moments.* New York: American Mathematical Society.

Van den Berg, G. J. (2001). "Duration Models: Specification, Identification, and Multiple Durations." In *Handbook of Econometrics, Volume V*, edited by J. J. Heckman and E. Learner. Amsterdam: North Holland.

Van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge, England: Cambridge University Press.

Van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes.* New York: Springer.

Widder, D. V. (1946). *The Laplace Transform.* Princeton, New Jersey: Princeton University Press.

INCENTIVE MECHANISMS FOR SAFE DRIVING: A COMPARATIVE ANALYSIS WITH DYNAMIC DATA

Georges Dionne, Jean Pinquet, Mathieu Maurice, and Charles Vanasse*

Abstract—Road safety policies often use incentive mechanisms based on traffic violations to promote safe driving—for example, fines, experience rating, and point-record driver's licenses. We analyze the effectiveness of these mechanisms in promoting safe driving. We derive their theoretical properties with respect to contract time and accumulated demerit points. These properties are tested empirically with data from the Quebec public insurance plan. We find evidence of moral hazard, which means that drivers who accumulate demerit points become more careful because they are at risk of losing their license. An insurance rating scheme introduced in 1992 reduced the frequency of traffic violations by 15%. We use this result to derive monetary equivalents for traffic violations and license suspensions.

I. Introduction

SINCE the 1970s fatality rates due to road traffic accidents have decreased steadily in developed countries, although risk exposure increased concomitantly (OECD, 2005). For example, over the past ten years, the road fatality rate decreased by 40% in France. However, the implied social cost of road accidents remains high (Doyle, 2005). A major reason for the improvement of the situation in the OECD has been the development of incentives for safe driving. Experience rating schemes used by the insurance industry have incentive properties (see Boyer & Dionne, 1989; Abbring, Chiappori, & Pinquet, 2003). They are supplemented by point-record driver's licenses based on traffic violations. In many countries, each convicted traffic offense is filed with a specific number of demerit points. When the accumulated number of points exceeds a given threshold, the driver's license is suspended. Point removal clauses are added so that this penalty can be avoided in the long run.[1] A point-record driver's license implemented in Quebec in 1978, together with a no-fault insurance plan for bodily injuries, replaced a tort system.[2] The road fatality rate decreased by 50% during the fifteen years that followed.[3]

In Quebec, the Société de l'Assurance Automobile du Quebec (SAAQ) is a public monopoly that provides coverage for bodily injury. The SAAQ is also in charge of accident prevention and control, including the management of driver's licenses. Before 1992, the rating structure for bodily injury insurance was completely flat. Since December 1, 1992, the public authorities in Quebec have implemented an experience rating system based on accumulated demerit points. This mechanism was added to other incentives, such as fines, the point-record driver's license in force since 1978, and the private sector insurance rating for coverages other than bodily injury.

This paper analyzes the incentive properties of fines, point-record driver's licenses, and experience rating based on traffic violations. Studies on incentive mechanisms for road safety have appeared in the economic literature for many years (Peltzman, 1975; Landes, 1982; Boyer & Dionne, 1987). In the presence of asymmetric information, insurers use partial insurance or experience rating to improve resource allocation. Both systems have proved to be efficient for handling moral hazard and adverse selection. Empirical tests have measured the effectiveness of such mechanisms for road safety (Sloan, Reilly, & Schenzler, 1995; Boyer & Dionne, 1989) and the presence of residual asymmetric information problems in insurers' portfolios (Chiappori & Salanié, 2000; Dionne, Gouriéroux, & Vanasse, 2001). More recently, Abbring et al. (2003) designed a new test based on the dynamics of insurance contracts to detect the presence of moral hazard. Their model makes it possible to separate the moral hazard effect on accidents from unobserved heterogeneity. They found no evidence of moral hazard in the French car insurance market. The convex structure of the French *bonus-malus* system is used to show that the optimal effort level exerted by a rational policyholder increases after a claim at fault. In our study, insurance pricing is not the major incentive scheme but rather a measure used to complement fines and the point-record driver's license. Moreover, the pricing scheme of the Quebec public automobile insurance is an increasing step function of past demerit points.

Insurance pricing may not suffice as a tool for designing an optimal road safety policy because it may not create the appropriate incentives for reckless drivers (Sloan et al., 1995). Bourgeon and Picard (2007) show how point-record driver's licenses provide incentives for road safety among

Received for publication July 21, 2005. Revision accepted for publication October 8, 2009.

* Dionne: HEC Montréal; Pinquet: Université Paris Ouest-Nanterre and Ecole Polytechnique; Maurice: HEC Montréal; Vanasse: TD Asset Management.

This research was financed by the FCAR, the SAAQ, the MTQ (Québec), the CREST, the PREDIT, the FFSA, and the AXA "Large Risks in Insurance" Chair (France). A previous version of the paper was entitled "Point-Record Incentives, Asymmetric Information and Dynamic Data." Different versions have been presented at many conferences. We thank J. M. Bourgeon, P. A. Chiappori, J. C. Cloutier, B. Fortin, C. Gourieroux, A. Monfort, C. Montmarquette, P. Ouellette, F. Pichette, P. Picard, J. Robert, B. Salanié, P. Viala, and two referees for their comments. We also thank R. Marshall and A. M. Lemire for their very efficient collaboration in the preparation of this unique data set. Our conclusions do not involve these persons and organizations. Claire Boisvert improved the presentation of the manuscript significantly.

[1] These clauses and their incentive properties are detailed in section III.

[2] North America preceded Europe in the design of such systems. Point-record driver's licenses were introduced in 1947 in the United States. Germany, France, and Spain implemented these mechanisms in 1974, 1992, and 2005, respectively.

[3] We do not have information on factors different from the incentive schemes analyzed in this paper that might have explained this decrease. This study is performed in a no-fault environment. For a recent comparison of strict liability and a negligence rule for risk incentives trade-off, see Fagart and Fluet (2009).

The Review of Economics and Statistics, February 2011, 93(1): 218–227

normal drivers (those who respond to the usual incentive schemes) when the judicial system or the insurance market fails to provide optimal incentives. Point-record driver's licenses also allow the government to incapacitate reckless drivers because fines for traffic violations are bounded above in many jurisdictions. Bounded fines exist for different reasons: many offenders are judgment proof and are unable to pay optimal fines; many drivers are expected to escape from paying tickets issued by the authorities when fines are very high; and society thinks it is unfair that rich and reckless drivers will pay high fines and continue to drive dangerously (Shavell, 1987a, 1987b). However, fines do reinforce the effectiveness of the point-record mechanism by providing normal drivers with more incentives. In Bourgeon and Picard's model, which uses only two levels of prevention, the optimal fine must be set at the maximal level and must be neither progressive nor regressive. These authors also discuss the optimality of point-removal mechanisms as a screening device. Public intervention can be justified when there is a significant difference between the private and the social cost of human lives (Viscusi, 1993). Finally, drivers may be unaware of their own accident or infraction probabilities or may misunderstand some features of the incentive environment.

We present the database in section II, as well as our first empirical results related to the introduction of the new pricing policy implemented in 1992. The point-record mechanisms (driver's license suspensions and insurance pricing) are described in section III, and their incentive properties are investigated in an optimal behavior model in which time and effort are continuous. If incentives are caused by fines and by the point-record driver's license, we show that the optimal effort level increases globally with the number of demerit points accumulated and decreases with the seniority of nonredeemed traffic violations, if any. Traffic violation risk varies conversely, as it decreases with the effort level.

These results are compared with the data in section IV. The observed dynamics on the drivers result from the incentive effects and the revelation with time of hidden features in risk distributions (that is, an unobserved heterogeneity effect). Let us compare these two effects at the time and event (accident or traffic violation) level of the data dynamics. Drivers with more traffic violations committed during a given period are riskier with respect to hidden features in risk distributions. Hence, unobserved heterogeneity entails a risk reassessment after each event, and the event effect of risk revelation counters the corresponding incentive effect of the point-record driver's license. The revelation effect on the risk level of a period without traffic violations is negative.[4] The time effect of unobserved heterogeneity is also converse to that of the incentive effect, which raises an identification issue in the interpretation of the observed dynamics on the drivers. Abbring et al.'s (2003) test for moral hazard can be used if there are no time effects in the incentives. Because the time effects are important in Quebec's point-record system, we use another approach. We test for an increasing link between traffic violation risk and the number of demerit points. Rejecting this assumption amounts to finding evidence of moral hazard (ab absurdo, because of the increasing event effect created by unobserved heterogeneity). In section IVB, the incentives created by the threat of driver's license suspension are found to increase with accumulated demerit points. These findings confirm the theoretical analysis.

The insurance rating scheme introduced in 1992 reduced the frequency of traffic violations by 15%. In section IVB, we link the incentive levels of the three point-record mechanisms as derived from the theoretical analysis of section III and the observed efficiency of these mechanisms. Monetary equivalents for traffic violations and license suspensions are derived from this analysis. Conclusions are drawn in section V and technicalities are relegated to an appendix available online at http://hal.archives-ouvertes.fr/hal-00414479/fr/.

II. Presentation of the Database and Preliminary Empirical Results

Our database represents roughly 1% of the SAAQ portfolio. The panel covers the period from January 1, 1983, to December 31, 1996. A first sample of 40,000 license holders was selected at random at the beginning of 1983, and about 300 young drivers were added each following year.[5] The attrition rate per year is close to 1.5%, which is very low compared with the private sector. Due to the monopolistic status of the SAAQ, leaving the company means that drivers lose their license, which explains the low attrition rate. The endogenous attrition is not very high. It was estimated from a bivariate probit model on traffic offenses and departures from the sample. A score test for the nullity of the correlation coefficient between the two equations was performed with the regression components set used in section IV.[6] The null hypothesis was not rejected at a 5% significance level. Hence, the attrition risk adds no significant information to the assessment of traffic violation risk.

The personal characteristics of each driver are available on the driver's license for the current period. These characteristics are used as regression components in the empirical study. Several types of events are recorded in the database:

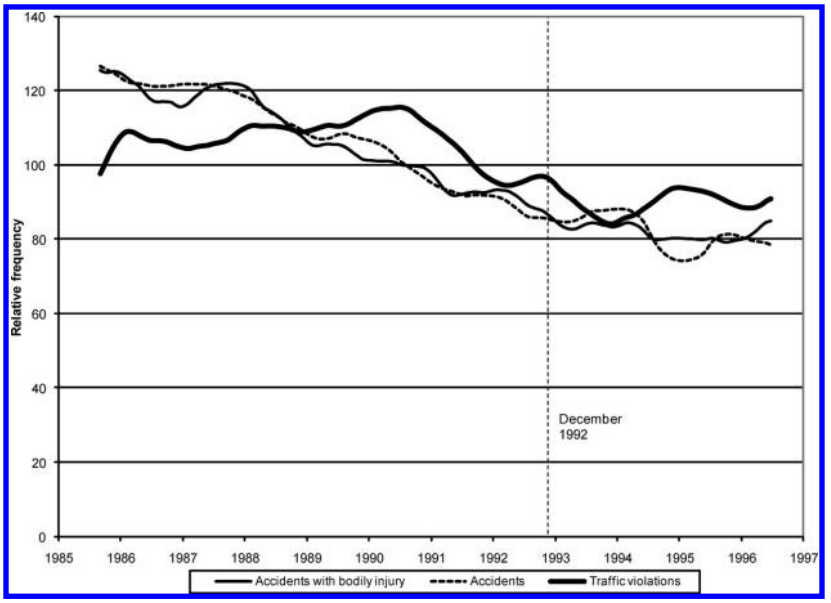
- Accidents that have led to a police report. Only those with bodily injury are compensated by the SAAQ.

[4] Increases in premium after an event and "no claims discount" clauses observed in non–life insurance can be explained by the revelation effect of unobserved heterogeneity.

[5] Selecting at random 1% of the new license holders every year would evidently have been a preferable sampling procedure. One thousand new license holders would then have been selected every year, as the entry rate in the SAAQ portfolio is close to 2.5%.

[6] Binary variables related to traffic offenses and attrition were created on a monthly basis and explained with the covariates used in section IVA. The score test statistic is equal to 0.34. Hence, we do not reject the nullity of the correlation coefficient at the usual significance levels.

FIGURE 1.—RELATIVE FREQUENCIES (IN PERCENTAGE) FOR TRAFFIC VIOLATIONS AND ACCIDENTS



Frequencies derived form a one-year centered moving average. The reform introducing the insurance experience rating structure based on demerit points took effect on December 1, 1992.

- Convicted violations of the Highway Safety Code, together with the number of demerit points used in the point-record mechanisms. The number of demerit points is based on the severity of the traffic violation. Their distribution is given in section IIID.
- Driver's license suspensions, which are spells rather than events.
- Premium payments, which since the 1992 reform are related to accumulated demerit points. These payments are made every two years on the policyholder's birthday.

Between 1985 and 1996, the average yearly frequencies of accidents with bodily injuries, accidents of all types (not including joint reports filed with private insurers), and traffic violations are 1.4%, 6.7%, and 16.9%, respectively. Figure 1 represents the relative frequencies derived from a one-year-centered moving average.[7]

There is an overall decline in the frequency of accidents, whereas the frequency of traffic violations remains more stationary. This may seem surprising, but it is explained by the evolution of the traffic control environment. For instance, the number of traffic control devices such as radars increased during the 1980s and 1990s. An increase in the rate of traffic offenses recorded by devices or police officers among those committed explains this relationship. Figure 1 shows evidence of several periods where the frequency of traffic violations increased along with opposite variations in the frequencies of accidents. A step-up in traffic control during these periods may well explain such observations.

A traffic violation committed by a driver must be selected twice in order to be filed with demerit points. It must first be recorded by a control device or a police officer. We already mentioned that the related selection rate had previously increased. Second, the offender must be convicted of the recorded traffic violation. The filing of a traffic violation is somewhat discretionary. Since the 1992 reform, for instance, drivers are being forced to pay more in premiums given demerit points, and we might expect police to be more hesitant to hand them out and to give warnings instead.[8] The conviction rate is less likely to vary with time than the recording rate.

In figure 1, a downturn is also observed in the frequency of traffic violations just before the date of the reform that introduced the experience rating structure based on demerit points (December 1, 1992). The reform was announced on August 1, 1992, which may explain why the downturn starts slightly before December 1, 1992. The experience-rated premium was first applied at the contract anniversary following December 1, 1992, and used the complete two-year history of demerit points. Hence the new incentives for safe driving took effect for most of the drivers on August 1, 1992 (drivers with their next birthday after December 1, 1992).[9] On average, the annual frequency of traffic violations was equal to 17.6%

[7] We begin in 1985 in order to match the regressions that follow because a two-year history is needed to obtain the accumulated demerit points. Data are first averaged over one year, to account for strong seasonal effects. A centered moving average derived on five fortnights is then performed twice in order to reduce the volatility of the series.

[8] We thank a referee for suggesting this interpretation.

[9] The SAAQ had the information on the drivers' traffic violation history since 1978. Drivers with a contract anniversary falling between the announcement of the reform and its enforcement are not incited by the experience-rated premium before this anniversary. Incentives exist otherwise (for these drivers after their birthday and for all the other drivers since the announcement of the reform). A referee suggested using this natural experiment in order to disentangle the incentive effects of the reform from calendar effects. We did not obtain significant results. Four months is a short period, and on average only one driver out of twelve was not incited by the rating scheme during the period.

before the reform and 15.4% afterward, which corresponds to a 12.5% decrease. The 1992 reform can be interpreted as a laboratory experiment to test whether experience rating based on demerit points reduces traffic violations. Nonetheless, the lower rate of traffic violations following the 1992 Quebec reform may be due to the changes in other factors that influence driver behavior. Identifying the influence of these factors necessitates a control group that is not affected by the policy change. Unfortunately, we do not have access to such a control group because the insurer is a monopoly and bodily injury insurance is compulsory.[10] In section IVB, we link the average decrease in the frequency of traffic violations before and after the 1992 reform to the overall effectiveness of incentive schemes.

Monetary and nonmonetary incentives for safe driving are based on traffic violations as well as the optimal behavior models presented in Section III. However, the actual social cost of road traffic is caused by accidents. To reconcile these two approaches, two results are worth mentioning. First, demerit points are good predictors of accidents. This is well documented in the literature and is confirmed in our data in section IVA. Second, the global stationarity of convicted traffic violation frequency observed in figure 1 concurs with a probable decrease in the frequency of committed traffic violations (see the already noted developments on selection rates). Lowering traffic violation risk through point-record mechanisms should also lower accident risk and the related social cost.

Finally, figure 1 shows that accidents with bodily injuries evolve in much the same way as those recorded in the SAAQ file. We include accidents of all types in the empirical analysis in order to obtain more stable results.[11]

III. Incentive Effects of Point-Record Mechanisms

A. *Point-Record Mechanisms in Quebec*

In this section, we describe Quebec's point-record mechanisms, which are based on traffic violations, both monetary (insurance premiums) and nonmonetary (point-record driver's license). Comparisons are made with respect to the mechanisms used in other countries.[12] We investigate the incentive properties of point-record mechanisms in sections IIIB to IIID.

In many countries, driver's license suspensions are based on demerit points. In Quebec, demerit points are assigned to convictions for traffic offenses, and their number depends on the traffic violation severity. When the accumulated number of demerit points reaches or exceeds a given threshold, the driver's license is suspended. Before January 1990 this threshold was set at twelve in Quebec and has since been increased to fifteen.

In order to mitigate the social cost of license suspensions, point-removal systems exist for most point-record driver's licenses. In Quebec, the demerit points related to a given driving offense are removed after two years. Hence, driver's license suspensions depend on the demerit points recorded during the previous two years. Most of point-removal systems used in American states follow the same approach. The average number of demerit points per convicted offense is 2.4 in Quebec. It takes about six traffic violations within two years to trigger a license suspension, an unlikely outcome when the annual traffic violation frequency is 16.9%. However, the heterogeneity of risks is high, and a point-record driver's license is also an incapacitating device for risky and reckless drivers through license suspensions. Another point-removal system consists of cancelling all the demerit points after a given period of violation-free driving. This mechanism was recently implemented in Spain, with a two-year seniority. The French driver's license uses the two removal systems.

The experience rating structure introduced by the SAAQ on December 1, 1992, links each premium paid every two years to the demerit points accumulated over the previous two years. The rating structure is given in section IIID. Once the premium is paid, the driver is reinstated with an unblemished record. Thus the length of the record relevant to the derivation of optimal behavior never exceeds two years.

B. *Basic Model for a Point-Record Driver's License without Point Removal*

Bourgeon and Picard (2007) analyze the incentive effects of point-record driver's licenses using a binary effort variable. We extend their approach with a continuous effort level. Hence, the effectiveness of effort may also be a continuous function of contract time, a desirable property for empirical validation. We show that under fairly general conditions, a rational policyholder's safe driving effort will increase with the number of demerit points accumulated.

We assume that the driver's license is revoked when the driver reaches a total of N demerit points. For simplicity, each convicted traffic violation is linked to one supplementary demerit point in this section. A driver with a suspended driver's license is reinstated after a period D with an unblemished record like that of a beginner.[13] The duration D may be fixed or random in the model. In Quebec, a license suspension is of random length because drivers must pass a new exam after a given period before recovering their license.[14] A rational driver maximizes his or her expected lifetime utility expressed in dollars and derived from:

[10] See Manning et al. (1987) for the use of a control group in the assessment of a cost-sharing modification in the health insurance market.

[11] Important variables in the regressions such as the number of accumulated demerit points have low frequencies for the highest values. It is hard to make an accurate estimate if the frequency of events is low, as is the case for accidents with bodily injury.

[12] We focus on license suspensions in the comparisons. However, financial penalties are also associated with demerit points in some U.S. states.

[13] This reinstatement can be seen as a removal of demerit points. In this paper, we consider a point-removal mechanism to be a cancellation of demerit points applied before the suspension of the driver's license.

[14] The failure rate at the first exam after the license suspension is 25%.

- An instantaneous driving utility, d_u.
- A time-dependent disutility of effort, denoted as $e(t)$.[15] This effort level is linked to an instantaneous traffic violation frequency risk, denoted as $\lambda(e(t))$. The hazard function $\lambda(e)$ corresponds to a probability $p(e)$ in static or discrete time incentive models and is assumed to be a positive, decreasing, and strictly convex function of the effort level.

In this section, we suppose that there is no point-removal mechanism. In that case, the lifetime expected utility (we assume an infinite horizon) depends on only the number n of accumulated demerit points and is denoted as u_n. The Bellman equation on the expected utility leads to

$$u_n = \frac{d_u}{r} - \frac{c_\lambda(u_n - u_{n+1})}{r},\ (0 \leq n < N), \tag{1}$$

where r is a subjective discount rate and c_λ is defined as

$$c_\lambda(\Delta u) \underset{def}{=} \min_{e \geq 0} e + [\lambda(e) \times \Delta u]. \tag{2}$$

Technical details can be found in appendix A1. From equation (2), incentives are effective if we have $\Delta u > -1/\lambda'(0)$. In this equation, Δu is the lifetime utility loss between the current state and the one reached after an additional traffic offense. Once quantified, Δu is the monetary equivalent of this traffic violation. Values of this type are derived in section IVB. The function c_λ minimizes the sum $e + [\lambda(e) \times \Delta u]$, which is the disutility flow of both effort (short-term component) and the expected lifetime utility loss (long-term component). All u_n are lower than $u_{\max} = d_u/r$, the private lifetime driving utility without the point-record driver's license. Equation (1) means that $c_\lambda(u_n - u_{n+1})/r$ is the minimal private utility cost of the point-record mechanism for a driver with n demerit points. The cycle of lifetime utilities is closed with a link between u_0 and u_N, the lifetime expected utility just after the suspension of the driver's license. For instance, if the private disutility of driver's license suspension is only the loss of driving utility during a period D, we have that

$$u_N = \beta u_0,\ \beta = E[\exp(-rD)]. \tag{3}$$

The utilities are then derived from the recurrence equations (1) and (3). Optimal effort depends on the variation of lifetime utility Δu because it minimizes the function defined in equation (2). The variable Δu (the argument of c_λ in equation (2)), which determines optimal effort, will be referred to later as the incentive level. In this setting, optimal effort depends on the number n of accumulated demerit points but not on time, and we denote it as e_n. We show in appendix A1 that e_n increases with n for any given value of N. The related frequency of violations $\lambda_n = \lambda(e_n)$ thus decreases with n. The intuition behind this result is the following. A given reduction in traffic violation risk is more efficient as the threat of the license suspension gets closer. Hence, the efficiency of effort increases with the number of demerit points accumulated, and we obtain an increasing link between this number and the optimal effort level. A parallel can be drawn with the "three strikes and you're out" rule enforced in California to deter crime. The deterrence effect increases from one to two strikes (Helland & Tabarrok, 2007).[16]

Fines represent another monetary incentive scheme applied in Quebec throughout the period investigated in this study. Let us denote $\overline{fa}$ as the average fine for a traffic violation conviction. Given that fines and premiums are low in comparison to average wealth, risk aversion is not significant. If fines are combined with the preceding point-record driver's license, the incentive level is equal to $\overline{fa} + u_n - u_{n+1}$. This means that the average fine is added to the utility loss in the variable, which determines optimal effort. Besides, the optimal effort still increases with n for a given value of the average fine.

C. *Point-Record Driver's Licenses with Point Removal*

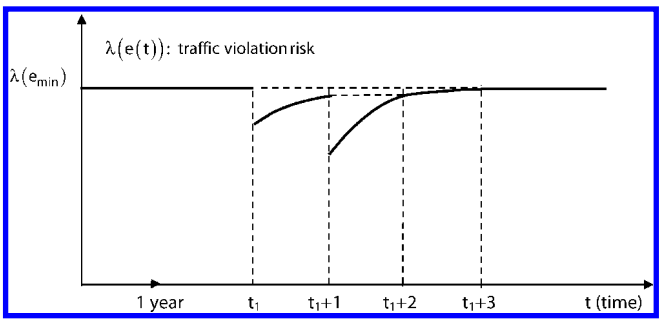
In Quebec, each traffic violation is redeemed at the end of a two-year period. Integrating this feature in the optimal behavior model is difficult because all the seniorities of nonredeemed driving offenses must be included as state variables in the dynamic programming equations. Lifetime utility is expected to increase with time for a given number of demerit points accumulated. Optimal effort depends on the difference between the existing utility and a substitute utility (that reached after an additional traffic violation). With the point-removal system in force in Quebec, the substitute utility increases with time, as does the existing utility. Time should have more value for worse situations, which implies that the substitute utility should increase faster than the existing utility. Optimal effort should thus decrease with time. Appendix A2 provides proof that optimal effort is continuous at the time of a point removal. Optimal effort is then expected to increase with each traffic violation in order to compensate for the decreasing link between time and effort. To summarize, we expect the effort level to increase globally with the number of demerit points accumulated and to decrease with the seniority of nonredeemed traffic violations if any.

Figure 2 illustrates the changes in the optimal traffic violation risk during a period that includes two violations and their point removal. The continuous time effect of incentives

[15] Safe driving effort can also reduce the expected disutility of accidents. If $e \rightarrow \delta(e)$ is the implied decrease in the disutility flow, replacing e by $e - \delta(e)$ in the model includes the influence of safe driving effort on accident disutility.

[16] We thank a referee for suggesting this connection. A "strike" is a conviction for a serious felony, the equivalent of a demerit point. Three strikes entail a prison sentence of 25 years to life, which in our context would be the license suspension. Comparing the time to rearrest for criminals with one or two strikes in California, Helland and Tabarrok (2007) conclude that those with two strikes are less risky than those convicted twice, with one strike only. They thus observe an increase in the deterrence effect at the second strike.

FIGURE 2.—TIME EVOLUTION OF TRAFFIC VIOLATION RISK RELATED TO OPTIMAL EFFORT: EXAMPLE WITH TWO TRAFFIC OFFENSES



The function $t \to \lambda(e(t))$ is represented by the thick line. In this example, the first traffic violation occurs at t_1, the second at $t_1 + 1$. The two demerit points are removed at $t_1 + 2$ and $t_1 + 3$. The optimal effort level is denoted as $e(t)$. It is determined by all the seniorities of nonredeemed traffic violations, if any. The minimum effort level is denoted as $e_{\min}$ and may be greater than 0, depending on the individual characteristics of the driver. After each traffic violation, there is an upward jump in the optimal effort level and an implied drop in traffic violation risk if incentives are effective. The continuous time effect of incentives counters the event-driven effect ($e(t)$ decreases with t and $\lambda(e(t))$ increases). Before t_1, the effort level is minimal. It increases after the first traffic violation. The drop in traffic violation risk is the opposite of the unobserved heterogeneity effect. The effort is maximum after the second offense, and then we have $e(t_1 + 2) = e(t_1 + 1)$ at the time of the first point removal (one nonredeemed traffic violation with a one-year seniority in both cases). We also have $e(t_1 + 3) = e_{\min}$ at the time of the second removal.

balances the downward jump in risk after each violation. The incentive effects at the time and event level are the opposite of those created by the revelation of unobserved heterogeneity. The incentive properties of the point-removal system in force in Quebec are similar to those of a mechanism without point removal concerning the number of demerit points accumulated. In section IV, we use the model of section III B as a proxy for the incentive environment in Quebec.

D. *Incentive Effects of Premiums Indexed on Demerit Points in Quebec*

Table 1 presents the rating structure enforced for each driver's license on the first contract anniversary following December 1, 1992. The premium paid every two years after this date depends on the number of demerit points accumulated in the previous two years. It does not represent the total premium for bodily injury insurance but rather the additional premium related to demerit points. This average premium, equal to $54.60, complements a yearly driver's license fee for insurance coverage equal to $107.

In this section, the incentive properties of this rating structure are analyzed separately from the point-record driver's license. An important input is the distribution of demerit points for a given driving offense, which we left out in section IIIB. Denoting f_j as the proportion of traffic violations with j demerit points, we have the following values:

$$f_1 = 4.71\%;\ f_2 = 52.32\%;\ f_3 = 38.34\%;\ f_4 = 2.83\%;\ f_5 = 1.80\%. \quad (4)$$

Note that f_5 refers to offenses with five or more points. Table 1 shows that the premium is a step function of the accumulated demerit points. Because of the local nonconvexity of the premium, the incentives may not always increase with the number of demerit points accumulated. Consider, for

TABLE 1.—SAAQ INSURANCE PREMIUMS FOR BODILY INJURY AS A FUNCTION OF ACCUMULATED DEMERIT POINTS SINCE THE LAST CONTRACT ANNIVERSARY

Accumulated Demerit Points (Last Two Years)	Premium for the Next Two Years (Canadian $)	Frequency (%)
0–3	50	93.7
4–7	100	4.9
8–11	174	1.1
12–14	286	0.2
15 or more	398	0.1

The premium is a step function of accumulated demerit points. The incentives for safe driving depend on the accumulated demerit points. For example, we document in the text that the incentive level is stronger with two accumulated demerit points than with four because the probability of climbing a step in the rating structure after a traffic violation is higher.

instance, a policyholder just before her contract anniversary. The incentive level will be stronger with two accumulated demerit points than with four. With four points, it is indeed less than likely that the next traffic offense will trigger an increase in premium. The corresponding probability is $2.83 + 1.80 = 4.63\%$, if we assume that the distribution of the f_j is independent of the accumulated demerit points. The incentives for safe driving are stronger at a two-point level because the probability of climbing a step in the rating structure after a traffic offense is close to 1. The result is in contrast to that obtained by Abbring et al. (2003) for the French *bonus-malus* scheme and its exponential structure. Hence, step pricing schemes may have poor incentive properties. They are employed because they are simple.

Let us design an optimal behavior model based on this rating structure. Once the premium is paid, the driver is reinstated with an unblemished record. Hence, the optimal control model can be designed with the next contract anniversary as the horizon. Let π_n be the premium paid for n demerit points accumulated during a period of $T = 2$ years between two contract anniversaries. As we disregard the point-record driver's license and its possible deprivation, the driving utility is no longer a parameter. However, we retain fines in the incentives for safe driving. We denote $v_n(t)$ as the expected disutility of premiums and fines paid until the next contract anniversary, where t is the seniority of the last anniversary and n is the number of demerit points accumulated since that date. We have the terminal conditions

$$v_n(T) = \pi_n,\ \forall n = 0, \ldots, N. \quad (5)$$

If incentives are related to fines and insurance premiums, the incentive level is the sum of the average fine $\overline{fa}$ and the expected variation of $v_n(t)$ after a traffic offense (see appendix A3). Hence, optimal effort is determined by $\overline{fa} + \Delta v_n(t)$, with $\Delta v_n(t) = \left(\sum_{j/f_j>0} f_j\, v_{\min(n+j,N)}(t)\right) - v_n(t)$. Derivations show that the average of $\Delta v_n(t)$ with respect to n does not vary much with time. The terminal values of Δv_n are derived from equations (4) and (5) and table 1. For $n = 0, 2, 4$, we obtain

$$\Delta v_0(T) = \$2.32;\ \Delta v_2(T) = \$47.65;\ \Delta v_4(T) = \$3.43. \quad (6)$$

The incentives with two points accumulated are much stronger than with four points, which confirms the analysis

following equation (4). The overall average of $\Delta v_n(t)$ with respect to t and n is close to \$12, a 9% increase for a \$130 average fine. In section IVB, this increase will be compared with the variation in traffic violation frequency before and after the reform.

IV. Empirical Results on the Incentive Effects of Point-Record Mechanisms

A. *Point-Record Driver's License*

In this section, we analyze the data before the 1992 reform that introduced the experience rating scheme based on demerit points. Thus, the point-record driver's license interacts only with fines. Regressions are performed from January 1985 (we need a two-year history to derive the accumulated demerit points) to December 1992, the date of the reform enforcement. We try to confirm the theoretical findings of sections IIIB and IIIC (that the effort level increases globally with the number of demerit points accumulated and decreases with the seniority of nonredeemed traffic violations, if any), and to find evidence of moral hazard in the data.

We estimate the hazard functions of convicted traffic offenses and accidents with a proportional hazards model (Cox, 1972). We retained the following specification:

$$\lambda_i^j(t) = \exp((x_i(t)\beta_j) + g_j(adp_i(t))) \times h_j(c_i(t)). \qquad (7)$$

In equation (7), $\lambda_i^j(t)$ is the hazard function of type j ($j = 1$: traffic violation or $j = 2$: accident) for driver i at calendar time t. Regression components are denoted by the line vector $x_i(t)$. We retained the gender, driver's license class, place of residence, age of the driver, calendar effects related to years and months, and the number of past license suspensions.[17] The number of demerit points accumulated in the previous two years is denoted as $adp_i(t)$. Let us comment on the expected shape of g_1, which links traffic violation risk to accumulated demerit points. The revelation effect of unobserved heterogeneity is increasing in *adp* and is converse to the incentive effect. In the absence of moral hazard, g_1 should then be increasing. We will test the null hypothesis that g_1 is weakly increasing against the alternative that it is not (g_1 strictly decreases for at least some values), rejecting the null amounts to finding evidence of moral hazard.[18]

Contract time $c_i(t)$ is integrated into the baseline hazard function h_j. The function c_i is set at 0 at the beginning of the period. Then it is reset to 0 at each event that triggers a variation of the accumulated demerit points (traffic violation or point removal). This event-driven operation should eliminate interactions between calendar and contract-time effects.

Table 2 shows that $\widehat{g_1}$ is increasing until *adp* reaches the seven-point threshold, at which point the increasing property is no longer fulfilled, and $\widehat{g_1}$ reaches a maximum at seven points. It is worth mentioning that the SAAQ warns the policyholders when their accumulated demerit points increase beyond a seven-point threshold.[19] When the accumulated demerit points are less than or equal to seven, the unobserved heterogeneity effects outweigh the incentive effects because $\widehat{g_1}$ is increasing. Controlling for other covariates, a driver with seven accumulated demerit points is $\exp(0.974) = 2.65$ times riskier than a driver with a violation-free record. Hence, unobserved heterogeneity effects strongly outweigh incentives to drive carefully between zero and seven points. This result is reversed beyond seven points. For instance, a driver with seven accumulated demerit points who gets a supplementary two-point offense has a traffic violation risk that is 19% lower. As the unobserved heterogeneity effect is increasing with *adp*, the estimated risk reduction due to the incentive effect is greater than 19%. A 20% value for the risk reduction after a traffic offense will be retained in the next section in order to derive monetary equivalents for traffic violations and license suspensions. Table 2 presents tests where the null assumption is $g_1(7) \leq g_1(adp)$, for $adp > 7$. The null is rejected for these values of *adp* at significance levels greater than 1.3%, except for $adp = 11$. We also performed a global test on the weakly increasing shape of g_1. The test statistic is a squared distance between the unconstrained estimation of g_1 given in table 2 and the set of weakly increasing functions of *adp* (see appendix A4 for more details). A graphical illustration is given in figure 3. We reject the increasing shape assumption for every usual significance level. As an absence of incentive effects entails an increasing shape for g_1 (unobserved heterogeneity effect), rejecting the null shows evidence of moral hazard.

TABLE 2.—ESTIMATION OF THE HAZARD FUNCTION FOR TRAFFIC VIOLATION AND ACCIDENT FREQUENCY RISKS

Variable	Level	Frequency (%)	Traffic Violation Risk (g_1)	P-value (H0 : $g_1(7) \leq g_1(adp)$, $adp > 7$)	Accident risk (g_2)
adp:	0 point[a]	76.60	0		0
Number of	1 point	0.39	0.311		0.243
accumulated	2 points	9.36	0.500		0.328
demerit points	3 points	6.23	0.522		0.449
(last two years)	4 points	1.92	0.794		0.496
	5 points	2.09	0.854		0.595
	6 points	1.25	0.873		0.647
	7 points	0.72	0.974		0.601
	8 points	0.55	0.876	0.0130	0.736
	9 points	0.43	0.764	< 0.0001	0.797
	10 points	0.32	0.786	0.0002	0.842
	11 points	0.06	0.906	0.2328	0.522
	12 points	0.04	0.378	< 0.0001	0.810
	13–14 points	0.04	0.571	0.0004	0.222

Controlling for other variables, a driver with seven accumulated demerit points who gets a supplementary two-point offense has a traffic violation risk reduced by $1 - \exp(0.764 - 0.974) = 19\%$. As the unobserved heterogeneity effect is increasing with *adp*, the estimated risk reduction due to the incentive effect is greater than 19%. The p-values refer to tests for the null assumption: $g_1(7) \leq g_1(adp)$ (with $adp > 7$) versus the alternative, with one-sided rejection regions. For instance, the null assumption related to $adp = 8$ is rejected at the level α if α is greater than the p-value 1.3%, and accepted otherwise. The results suggest that g_1 is not increasing beyond $adp = 7$, which shows evidence of moral hazard. A global test for the increasing shape of g_1 is presented in figure 3. Additional regression variables are gender, driver's license class (nine levels), place of residence (sixteen levels), age of the driver (five slopes), number of past suspension spells (three levels), as well as calendar effects related to years (eight levels), and months (twelve levels). Number of observations: 3,492,868 duration-event indicator pairs, derived from 41,290 driver's licenses. The durations are bounded above by one month.

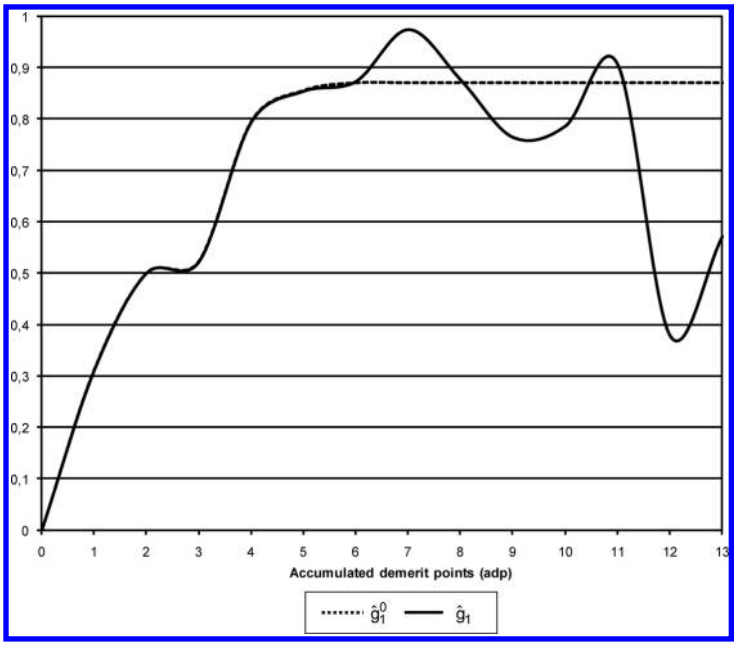
[a] The reference level.

[17] Comprehensive regressions based on two-year periods can be found in Dionne et al. (2001).

[18] We thank a referee for suggesting this formulation of the test.

[19] Drivers are not informed when offenses are redeemed.

FIGURE 3.—GLOBAL TEST FOR THE INCREASING SHAPE OF THE TRAFFIC VIOLATION RISK AS A FUNCTION OF ACCUMULATED DEMERIT POINTS



This figure presents a graphical test where the null is the weakly increasing shape of g_1 as a function of *adp*, the number of accumulated demerit points. The test statistic is a squared distance between the unconstrained estimation of g_1 (thick line) given in table 2, and the set of weakly increasing functions of *adp*. The metric is that of the Wald asymptotic test, and the weakly increasing function closest to g_1 is represented by the dashed line. The statistic is equal to 43.18, and its asymptotic distribution is less than that of a $\chi^2(13)$ in the sense of first-order stochastic dominance. This leads us to reject the increasing shape assumption for every significance level greater than 42×10^{-6}. As an absence of incentive effects entails an increasing shape for g_1 (unobserved heterogeneity effect), rejecting the null shows evidence of moral hazard.

Table 2 shows less evidence of moral hazard on accident risk than on traffic violation risk, as the estimation of g_2 increases until $adp = 10$. A possible interpretation is that we cannot separate at-fault from no-fault accidents. In the literature, the incentive effect is usually higher with at-fault accidents. Besides, drivers nearing the license suspension threshold might also apply opportunistic strategies regarding traffic violations (for example, paying more attention to radar) without otherwise modifying their attitude toward road traffic risk.

The estimated hazard function from the traffic violation equation globally decreases with time.[20] This means that the continuous time effect created by the revelation of unobserved heterogeneity outweighs the incentive effects (see figure 2 and the discussion that follows). This is not surprising, given that the average increase of g_1 in table 2 (a vast majority of drivers have fewer than seven demerit points) must be balanced by a negative time effect in order to obtain stationary risk levels on average. As effort is expected to decrease with time only if the number of demerit points accumulated is greater than zero, the statistical analysis can be refined with a stratified proportional hazards model.[21] Two strata can be created, depending on whether the variable $adp_i(t)$ is equal to 0 or not. On both strata, we observe a decreasing shape for the estimated hazard function.

[20] Results are available from the authors on request.

[21] Stratification in a proportional hazards model means that Cox likelihoods (of a multinomial logit type) are derived for each stratum and then multiplied together. In other words, an individual with an observed event is assumed to have competed only with other individuals in the same stratum and at risk at the same date. However, the same coefficients for the covariates are used across all strata.

B. *Incentive Effects of the 1992 Reform and Monetary Equivalents for Traffic Violations and License Suspensions*

In this section, we assess the efficiency of the 1992 reform. We compare its incentive level to those of fines and of the point-record driver's license. We relate the efficiency of the reform to its relative weight in the three incentive mechanisms, and we infer results on the shape of the link between the cost and efficiency of effort. Finally, we derive monetary equivalents for traffic violations and license suspensions from the estimated efficiency of incentives created by a supplementary traffic violation.

In section II, we mentioned a 12.5% decrease in the average frequency of traffic violations before and after the reform, which introduced the experience rating structure based on demerit points. This result changes slightly if we control with the regression components used in table 2. A regression estimated from 1985 to 1996 with the covariates of section IVA and a dummy related to the period following December 1, 1992, associates the reform with a 15% decrease.[22] The results of section IIID (for instance, equation (6)) suggest that the number of demerit points accumulated since the previous anniversary should influence the effectiveness of the 1992 reform. However, we did not obtain significant results in this direction. The drivers' limited knowledge of the environment could explain this poor result, a point developed later.

[22] We retained the covariates used in table 2 except for dummies related calendar effects and to the number of past license suspension spells. The estimated parameter for the reform dummy is equal to −0.163, and the related standard deviation is equal to 0.008. Hence, the reform effect is conclusive at the usual tests significance levels.

We perform an overall comparison of the three incentive schemes. We use the model without point removal of section IIIB to analyze the incentives for drivers in Quebec. Before the 1992 reform, fines were supplemented by a point-record driver's license. Optimal effort after n nonredeemed traffic violations depends on the incentive level—that is, the argument of the function c_λ defined in equation (2). This level is equal to $\overline{fa} + u_n - u_{n+1}$ (see section IIIB). The average fine $\overline{fa}$ is equal to \$130, and the 1992 reform entails an average increase in the incentive level equal to \$12 from section IIID. At this point, it seems interesting to relate optimal frequency risk and the incentive level. This relation can be assessed from the elasticity between the former and the latter variable. When the incentives are effective, it can be shown that this elasticity is less than -1 if and only if $\log(\lambda)$ is a concave function of effort (elasticity and concavity are considered locally; see appendix A5 for a proof). A global elasticity equal to -1 is linked to an exponential decay of λ. With $\lambda(e) = \lambda(0) \times \exp(-\alpha e)$, the optimal risk level as a function of $\overline{fa} + \Delta u$ (the incentive level) is equal to $1/(\alpha \times (\overline{fa} + \Delta u))$ if the incentives are effective.

We now apply this result to the 1992 reform. As the reform entailed a significant reduction in traffic violation risk regardless of the number of demerit points accumulated, we can assume that incentives are effective for a representative driver.[23] This leads us to analyze the elasticity between traffic violation risk and the incentive level. Suppose that we leave out the modifications of incentive levels due to the aggregation of incentive mechanisms. Then we can relate:

- On the one hand, a 15% reduction in the frequency of traffic violations since the 1992 reform.
- On the other hand, a relative increase in the incentive level ranging between 9% and 10%. Indeed, the 1992 reform entails a \$12 average increase in the incentive level. This increase supplements the other components of the incentives—the \$130 average fine and the utility variation for the point-record driver's license. In table 2, the point-record driver's license offers significant incentives beyond a seven-point threshold, a result corresponding to only a minority of drivers (1.4%). The contribution of the point-record driver's license to the incentives is low compared with that of fines.

This suggests that the elasticity between the optimal frequency risk and the incentive level is less than -1 in this case. This result is linked to a locally concave shape of $\log(\lambda)$ for the representative driver. However, external effects could also explain the reduction in the frequency of traffic violations. We cannot eliminate these effects because there is no control group that is not affected by the reform. Besides, the elasticity would be modified if the distribution of demerit points for a given driving offense (see equation (4)) was wrongly perceived by the drivers. The \$12 contribution of the reform to the incentive level is low because of the high frequency of drivers without demerit points since the previous birthday (87%) and the low incentive level of the reform for these drivers (see equation (6)). This level depends largely on the probability of moving up a step in the premium schedule after an additional traffic violation, which must be associated with four demerit points or more. If the perceived frequency of corresponding traffic violations was greater than the actual one ($f_4 + f_5 = 2.83 + 1.80 = 4.63\%$), the variation of the incentive level induced by the reform would increase. In that case, the elasticity would be closer to 0.

Finally, let us assess monetary equivalents for a traffic violation and a license suspension. The monetary equivalent of a traffic violation is the component of the incentive level related to the point-record driver's license. It is equal to the loss of lifetime utility after a traffic violation, which depends on the number of traffic violations accumulated. A value can be derived from the effectiveness of effort estimated in table 2 and from the link between efficiency of effort and the incentive level. An additional traffic violation beyond seven accumulated demerit points entails a reduction of traffic violation frequency of close to 20%. Although these drivers cannot be seen as representative, we will apply the elasticity derived from the preceding developments. If the 9% increase in the incentive level induced by the 1992 reform entails a 15% reduction in the frequency of traffic violations, a 20% decrease in traffic violation frequency is associated with a 12% increase in the incentive level. The implied loss of lifetime utility depends on the traffic violation frequency risk λ but mostly on the discount rate r (see appendix A6). With $\lambda = 0.15$, the monetary equivalent of an additional traffic violation for these drivers belongs to the interval [\$120, \$195] if $r = 3\%$, and to [\$41.1, \$55.7] if $r = 6\%$. Besides, the growth rate of this monetary equivalent with respect to the number n of nonredeemed traffic violations falls between r/λ_n and r/λ_{n+1}, where λ_n is the optimal traffic violation frequency related to n. Monetary costs for license suspensions are then obtained by adding the costs of traffic violations until the crossing of the demerit point threshold. Starting from a zero-point record and assuming that six traffic violations are needed to trigger a license suspension, the monetary cost of a license suspension is bounded by \$700 and \$1,178 if $r = 3\%$ and if $\lambda_0 = 0.17$; $\lambda_1 = 0.17$; $\lambda_2 = 0.16$; $\lambda_3 = 0.15$; $\lambda_4 = 0.12$; $\lambda_5 = 0.09$.[24] A misperception of the environment could modify the monetary equivalents of traffic violations and license suspensions. In the discussion on the elasticity between the optimal frequency risk and the incentive level, we argued that an overestimation of the frequency of severe traffic violations would increase the perceived incentive effect of the 1992 reform. The \$12 average effect of the reform on the incentive level should then be upgraded, and the monetary equivalents of traffic violations and license suspensions should also be upgraded.

[23] From section IIIB, a sufficient condition to have this result is that the average fine is higher than the threshold $-1/\lambda'(0)$ beyond which the incentives are effective.

[24] These values correspond to an effort level that increases with the number of demerit points accumulated.

V. Conclusion

In this paper, we analyze the properties of policies designed to promote safe driving. Three important incentive mechanisms for road safety are used in Quebec. When the Quebec government introduced a no-fault insurance regime in 1978, it implemented a point-record driving license to maintain incentives for road safety. Because prevention activities are not observable by the public insurer, the new mechanism uses drivers' accumulated demerit points to approximate past safety activities and increase incentives to reduce infractions. Under moral hazard, drivers who accumulate demerit points become more careful because they are at threat of losing their license. This result was confirmed from the data. Otherwise, we would have accepted an increasing link assumption between accumulated demerit points and traffic violation risk.

Fines are on average the most efficient device, but they are associated with a uniform incentive level. We designed our incentive models with a representative driver, but there is, of course, heterogeneity in the individual parameters, such as the threshold beyond which the incentives are effective. We did not have wealth variables at hand, and an interesting empirical issue would have been to cross such variables with a reform dummy in risk assessment.

The experience-rated premium based on accumulated demerit points is a monetary point-record mechanism. The empirical results exhibit a rather uniform effectiveness after its enforcement in 1992—a 15% decrease in the frequency of traffic violations. Its incentive effects do not strictly increase with the accumulated demerit points, however, because of the steps in the rating structure. The actual incentive effect of the reform looks more like that induced by an increase in the average fine. The SAAQ modified its rating policy in 2008, with a premium increase from the first demerit point.

Regulations are not self-enforcing. To successfully achieve a reform, the regulator must implement enforcement activities and bear the costs thereof. In our application to road safety, the police is the most important enforcer of the different incentive schemes. Police officers have a limited capacity to modify the penalties for a given infraction but may affect the detection probabilities based on the aggregate convictions cycles, the driver's accumulated demerit points, or other tasks. When analyzing our results, we must consider the effect of these potential behaviors. For example, is it possible that when accumulated demerit points are greater in a given file, police officers have an incentive to issue them more or less often because of risk of the driver's losing his license?

We did not have precise information to control for these factors. It is well documented that the police collaborate very well with the Quebec public insurer and government. There is, consequently, a low probability that aggregate safety scores would have affected police behavior without the agreement of the government, which is assumed to be fully committed to its road safety policies. Regarding the effect of accumulated demerit points, we do not believe there is a general policy in the police force on the behavior to adopt based on a driver's past experience, although some individuals may modify their behavior in one direction or the other. This effect is implicitly considered a purely random effect in our analysis and should not have affected our results on the point-record incentives.

In this study, we have not examined the long-term evolution of accidents in detail, because we did not have access to the control variables of interest. In recent years, many road safety initiatives have had an impact on accidents but did not necessarily have any effect on violations. These initiatives include measures such as occasional campaigns to prevent fatal accidents, increased police patrols to reduce speeding, and designated-driver campaigns to prevent drinkers from getting behind the wheel. The decline in deaths and serious injuries can also be explained by vehicular improvements and the wearing of seat belts. All such measures are complementary to those studied in this article.

REFERENCES

Abbring, Jaap, Pierre-André Chiappori, and Jean Pinquet, "Moral Hazard and Dynamic Insurance Data," *Journal of the European Economic Association* 1 (2003), 767–820.

Boyer, Marcel, and Georges Dionne,"The Economics of Road Safety," *Transportation Research* 21B:5 (1987), 413–431.

——— "An Empirical Analysis of Moral Hazard and Experience Rating," this REVIEW 71 (1989), 128–134.

Bourgeon, Jean-Marc, and Pierre Picard, "Point-Record Driving License and Road Safety: An Economic Approach," *Journal of Public Economics* 91 (2007), 235–258.

Chiappori, Pierre-André, and Bernard Salanié, "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy* 108 (2000), 56–78.

Cox, David R., "Regression Models and Life Tables," *Journal of the Royal Statistical Society*, Series B, 34 (1972), 187–220.

Dionne, Georges, Christian Gouriéroux, and Charles Vanasse, "Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment," *Journal of Political Economy* 109 (2001), 444–453.

Dionne, Georges, Mathieu Maurice, Jean Pinquet, and Charles Vanasse, "The Role of Memory in Long-Term Contracting with Moral Hazard: Empirical Evidence in Automobile Insurance," working paper no. 01-05, HEC Montréal (2001), http://neumann.hec.ca/gestiondesrisques/01-05.pdf.

Doyle, Joseph J., "Health Insurance, Treatment and Outcomes: Using Auto Accidents as Health Shocks," this REVIEW 87 (2005), 256–270.

Fagart, Marie-Cécile, and Claude Fluet, "Liability Insurance under the Negligence Rule," *RAND Journal of Economics* 40 (2009), 486–509.

Helland, Eric, and Alexander Tabarrok, "Does Three Strikes Deter? A Nonparametric Estimation," *Journal of Human Resources* 52 (2007), 309–330.

Landes, Elisabeth M. "Insurance, Liability, and Accidents: A Theoretical and Empirical Investigation of the Effect of No-Fault Accidents," *Journal of Law and Economics* 25 (1982), 49–65.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, Arleen Leibowitz, and Susan Marquis, "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment," *American Economic Review* 77 (1987), 251–277.

OECD, International Road Traffic and Accident Data Base (2005), http://cemt.org/IRTAD/.

Peltzman, Sam, "The Effects of Automobile Regulation," *Journal of Political Economy* 83 (1975), 677–725.

Shavell, Steven, "The Optimal Use of Nonmonetary Sanctions as a Deterrent," *American Economic Review* 77 (1987a), 584–592.

——— *Economic Analysis of Accident Law* (Cambridge, MA: Harvard University Press, 1987b).

Sloan, Frank A., Bridget A. Reilly, and Christoph Schenzler, "Effects of Tort Liability and Insurance on Heavy Drinking and Driving," *Journal of Law and Economics* 38 (1995), 49–77.

Viscusi, Kip W., "The Value of Risks to Life and Health," *Journal of Economic Literature* 31 (1993), 1912–1946.

A Appendix

A.1 Incentive effects of point-record driver's licenses: Model without point removal

The Bellman equation on the expected utility is

$$u_n = \max_{e\geq 0} (d_u - e)dt + (\exp(-rdt) \times [((1-\lambda(e)dt) \times u_n) + (\lambda(e)dt \times u_{n+1}) + o(dt)]) .$$

We then obtain

$$0 = \max_{e\geq 0} (d_u - e) - (r + \lambda(e))u_n + \lambda(e)\, u_{n+1},$$

and equation (1).

We give the main properties of the function

$$c_\lambda(\Delta u) \underset{def}{=} \min_{e\geq 0} e + [\lambda(e) \times \Delta u] = \min_{e\geq 0} h(\Delta u, e) ,$$

with λ a positive, decreasing and strictly convex hazard function. The related optimal effort level is equal to

$$e_{opt}(\Delta u) = \arg\min_{e\geq 0} h(\Delta u, e) \Rightarrow$$

$$e_{opt}(\Delta u) = 0 \text{ if } \Delta u \leq \frac{-1}{\lambda^{'}(0)};\ e_{opt}(\Delta u) = \left(\lambda^{'}\right)^{-1}\left(\frac{-1}{\Delta u}\right) \text{ if } \Delta u \geq \frac{-1}{\lambda^{'}(0)}. \quad (8)$$

Hence the function c_λ is defined on the real line as the optimal effort. From the last equation, we obtain

$$\Delta u \leq \frac{-1}{\lambda^{'}(0)} \Rightarrow c_\lambda(\Delta u) = \lambda(0) \times \Delta u, \quad (9)$$

and c_λ is linear in the neighborhood of 0, which corresponds to no effort. The function c_λ is strictly increasing because λ is strictly positive. If $\Delta u \geq 0$, we have that:

$$c_\lambda(\Delta u) = h(\Delta u, e_{opt}(\Delta u)) \geq e_{opt}(\Delta u) \Rightarrow \lim_{\Delta u\to+\infty} c_\lambda(\Delta u) \geq \lim_{\Delta u\to+\infty} e_{opt}(\Delta u) = +\infty.$$

Hence c_λ is an increasing homeomorphism on the real line.

The function c_λ is concave. From the envelope theorem, we have

$$h'_{\Delta u}(\Delta u, e) = \lambda(e) \Rightarrow c'_\lambda(\Delta u) = h'_{\Delta u}(\Delta u, e_{opt}(\Delta u)) = \lambda(e_{opt}(\Delta u)). \tag{10}$$

Hence c_λ is concave from the assumptions on λ and from the properties of e_{opt}.

We give a proof of the increasing property of the optimal effort level as a function of accumulated demerit points. From equation (1), we obtain

$$u_n - u_{n+1} = c_\lambda^{-1}\left(r(u_{\max} - u_n)\right),\ u_{\max} = \frac{d_u}{r}\ (0 \leq n < N). \tag{11}$$

The sequence $(u_n)_{0\leq n\leq N}$ is decreasing because we have $u_{\max} \geq u_n$. Plugging this result into equation (11) implies that the sequence $(u_n - u_{n+1})_{0\leq n<N}$ is increasing. The optimal effort level is denoted as e_n, and expressed as

$$e_n = \arg\ \min_{e\geq 0}\ e + [\lambda(e) \times (u_n - u_{n+1})] = e_{opt}(u_n - u_{n+1}),$$

for $0 \leq n < N$, where e_{opt} is defined from (8). As e_{opt} is an increasing function, the optimal effort is an increasing function of the number of demerit points for any given value of the license suspension threshold.

Let us specify the condition under which incentives are effective. From (11) and (9), we obtain

$$e_n > 0 \Leftrightarrow u_n - u_{n+1} = \frac{d_u - ru_n}{\lambda(0)} > \frac{-1}{\lambda'(0)} = \underline{\Delta u}. \tag{12}$$

If fines are included in the incentives, u_{n+1} is replaced by $u_{n+1} - \overline{fa}$ in equation (1), which leads to the recurrence equation (see Figure 4)

$$d_u - ru_n = c_\lambda(u_n - u_{n+1} + \overline{fa})$$

$$\Leftrightarrow u_{n+1} = u_n + \overline{fa} - c_\lambda^{-1}(d_u - ru_n) = g(u_n). \tag{13}$$

The fixed point of g is the lifetime driving utility if fines were the only incentive scheme, i.e.

$$\widetilde{u}_{\max} = \frac{d_u - c_\lambda(\overline{fa})}{r}.$$

We evidently assume that $d_u > c_\lambda(\overline{fa})$, i.e. $\widetilde{u}_{\max} > 0$. If the two incentives are mixed, we have $u_n \leq \widetilde{u}_{\max}$ and we deduce from (13) the properties of utilities and of optimal effort levels as functions of n that we obtained in the first place. In addition, we have

$$e_n > 0\ , \forall n,\ \Leftrightarrow \overline{fa} + u_n - u_{n+1} > \underline{\Delta u}\ , \forall n.$$

This condition is fulfilled if

$$\overline{fa} > \underline{\Delta u} = -1/\lambda^{'}(0),$$

in which case the incentives are effective at every level.

A.2 Incentive effects of point-record driver's licenses: Model with point removal

The Bellman equation on a holistic incentive model can be written as follows

$$d_u - ru(S) + \left(\frac{d}{dt}\left[u(S_t)\right]\right)_{t=0^+} = c_\lambda(\overline{fa} + u(S) - E\left[u(TR(S))\right]). \qquad (14)$$

The state variables S are the seniorities of each non-redeemed traffic offense (if any), the related demerit points and the seniority of the last contract anniversary if the premium is included in the incentives. The related lifetime utility is $u(S)$. The state S_t is reached from S with an eventless history (no traffic offense, point removal or contract anniversary) of duration t. The parameters d_u and $\overline{fa}$ are the driving utility flow and the average fine, and $E\left[u(TR(S))\right]$ is the lifetime utility averaged with transition probabilities on the state(s) reached from S after a traffic offense. Continuity equations on utility at the time of a point removal or of a contract anniversary (in the latter case, the increase in lifetime utility is equal to the disutility of the premium) and the equation linking the utility of a beginner and the utility just after a license suspension define the solution together with equation (14).

Let us prove the continuity of optimal effort after a point removal in a system where each traffic violation is redeemed beyond a given seniority threshold, equal to T. We suppose that each traffic violation is associated with one demerit point, and that incentives are related to fines and to the point-record driver's license. The state variables are then the seniorities of each non-redeemed traffic offense, if any. Let us denote these variables as

$$S = (t_1, \ldots, t_n),\ 0 \leq t_1 < \ldots < t_n < T.$$

The corresponding optimal effort is denoted as $e(S)$. The states reached without traffic offense before the next point removal are then

$$S_t = (t_1 + t, \ldots, t_n + t),\ 0 \leq t < T - t_n.$$

We denote the state reached from S after an additional traffic offense (if $n < N$) as $(0, t_1, \ldots, t_n) = TR(S)$. As the lifetime utility is continuous after a point removal, we have the following result:

$$n \geq 1 : \lim_{t \to (T-t_n)^-} u(S_t) = u(S^R),\ S^R = (t_1 + T - t_n, \ldots, t_{n-1} + T - t_n).$$

The state S^R is reached from S if there is no traffic offense before the first point removal. Then it is easily seen that

$$\lim_{t \to (T-t_n)^-} u\left[TR(S_t)\right] = u\left[TR(S^R)\right] = u(0, t_1 + T - t_n, \ldots, t_{n-1} + T - t_n).$$

This means that the left continuity at $T - t_n$ of the map $t \to u(S_t)$ also holds for the map $t \to u\left[TR(S_t)\right]$, which is associated with the states reached after an additional traffic offense. The reason is that removal of past offenses occurs regardless of the future individual history.

From the three last equations, we obtain

$$\lim_{t \to (T-t_n)^-} e(S_t) = e(S^R)$$

and the continuity property of the optimal effort level. Since we expect a global increasing link between optimal effort and the accumulated demerit points, the time-effect should be decreasing globally in order to fulfill this continuity property.

A.3 Incentive effects of the experience rating system

Let us derive the Bellman equation on the expected disutility function given in (15), including an average fine of fa_j for a j demerit point traffic violation. The optimal disutility function is obtained from the program

$$v_n(t) = \min_{e\geq 0} edt + (\exp(-rdt) \times (1-\lambda(e)dt) \times v_n(t+dt))$$

$$+ \left(\exp(-rdt) \times \left[\sum_{j \ / \ f_j>0} f_j \, \lambda(e)dt \times \left[v_{\min(n+j,N)}(t+dt) + fa_j \right] \right] \right) + o(dt),$$

which leads to

$$0 = v_n'(t) + c_\lambda \left(\overline{fa} + \left(\sum_{j \ / \ f_j>0} f_j \, v_{\min(n+j,N)}(t) \right) - v_n(t) \right) - r v_n(t),$$

with $\overline{fa} = \sum_{j \ / \ f_j>0} \ f_j \times fa_j$ the average fine. Then we obtain the Bellman equation

$$v_n'(t) = r v_n(t) - c_\lambda \left(\overline{fa} + \Delta v_n(t) \right), \ (0 \leq n \leq N). \tag{15}$$

The argument of c_λ, $\overline{fa} + \Delta v_n(t)$, is the incentive level.

A.4 Test for the increasing shape of a vector of parameters, and for evidence of moral hazard

Let θ be the vector of values reached by the function g_1, where g_1 is defined in equation (7). We have $\theta_{adp} = g_1(adp)$, $1 \leq adp \leq 13$, with adp the number of demerit points accumulated in the last two years. In what follows, θ can be

replaced by g_1, but we use here the usual notations for asymptotic tests. The level $adp = 13$ also includes the 14 point total. As $adp = 0$ is the default level in the regression, the increasing shape of g_1 is related to the null assumption

$$g_1(0) = 0 \leq \theta_1 \leq \ldots \leq \theta_{13}. \tag{16}$$

Rejecting this assumption amounts to finding evidence of moral hazard. The null assumption is obviously expressed as a set of inequalities of the type $f_i(\theta) \geq 0$ $(i = 1, \ldots, 13)$, with

$$f(\theta) = \underset{i}{vec}\,(f_i(\theta)) = J \times \theta,\ \underset{13,13}{J} = \begin{pmatrix} 1 & 0 & \ldots & \ldots & 0 \\ -1 & 1 & \ddots & 0 & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & 0 & \ddots & 1 & 0 \\ 0 & \ldots & 0 & -1 & 1 \end{pmatrix}.$$

A usual asymptotic test for the nullity of $f(\theta)$ derived from $\widehat{\theta}$, the maximum likelihood estimation (m.l.e.) of θ – estimated here together with nuisance parameters – is the Wald test. The related statistic writes as follows

$$W = ||f(\widehat{\theta})||^2_M = d^2_M(f(\widehat{\theta}), 0),\ M = \left(\widehat{V}\left[f(\widehat{\theta})\right]\right)^{-1} = \left[J_f(\widehat{\theta}) \times \widehat{V}(\widehat{\theta}) \times {}^t J_f(\widehat{\theta})\right]^{-1},$$

where $J_f(\widehat{\theta})$ is the Jacobian of f ($J_f(\widehat{\theta}) = J$ as f is linear), and where $\widehat{V}(\widehat{\theta})$ is derived from the asymptotic efficiency of the m.l.e. Under the null, the Wald statistic follows a $\chi^2(13)$ distribution asymptotically. The null (i.e. $f(\theta) = 0 \Leftrightarrow \theta = 0$ as f is one-to-one; as J is invertible, we have $W = ||\widehat{\theta}||^2_{[\widehat{V}(\widehat{\theta})]^{-1}}$) is rejected from our data at every usual level, as $W = 1933.12$.

The increasing shape of g_1 is the null to be tested for with inequalities, and is expressed as follows

$$H_0 : f_i(\theta) \geq 0\ \forall i = 1, \ldots, 13 \Leftrightarrow f(\theta) \in O^+,$$

where O^+ is the positive orthant of $\mathbb{R}^{13}$. In order to test for H_0 against the alternative, a natural statistic derived from the Wald test uses the distance

between the unconstrained estimation of positivity constraints and the positive orthant, i.e.

$$W^+ = d^2_M(f(\widehat{\theta}), O^+) = \min_{z \in O^+} ||f(\widehat{\theta}) - z||^2_M = ||f(\widehat{\theta}) - z_0||^2_M,\ z_0 = P_{O^+}(f(\widehat{\theta})).$$

The metric M was already used for the Wald statistic. As 0 belongs to the positive orthant, we have $W^+ \leq W$. The derivation of the projection $P_{O^+}(f(\widehat{\theta}))$ of $f(\widehat{\theta})$ on the positive orthant O^+ can be seen numerically as a quadratic programming problem. As this projection belongs to the frontier of the positive orthant, the issue is to find the set of coordinates that are strictly positive.

The asymptotic properties of the statistic W^+ under the null are described by Wolak (1991). The statistic is not asymptotically distribution free under the null, as is the case in the equality setting. The limit distribution is a mixture of $\chi^2(i)$ distributions ($0 \leq i \leq 13$), where $\chi^2(0)$ is the unit mass (Dirac) at zero. If $f(\theta)$ belongs to the interior of O^+, the limit distribution of W^+ is obviously equal to $\chi^2(0)$, due to the pointwise convergence of $f(\widehat{\theta})$ towards $f(\theta)$. Otherwise the weights of the mixture depend on the position of $f(\theta)$ at the frontier of the positive orthant. As a $\chi^2(i+1)$ variable is obtained as the sum of a $\chi^2(i)$ variable and of a nonnegative variable, the $\chi^2(i)$ distributions increase with i in the sense of first order stochastic dominance. Then all the limit distributions of W^+ will be first order-dominated by a $\chi^2(13)$ distribution, owing to their definition as mixtures. This result is also simply obtained from the inequality $W^+ \leq W$.

On our data, the projection $P_{O^+}(f(\widehat{\theta}))$ is found equal to the projection of $f(\widehat{\theta}) = J \times \widehat{\theta}$ on the subspace of $\mathbb{R}^{13}$ defined by $\{z \in \mathbb{R}^{13} \ / \ z_7 = \ldots = z_{13} = 0\}$. We have: $P_{O^+}(f(\widehat{\theta})) = f(\widehat{\theta}^0) \Rightarrow \widehat{\theta}^0 = J^{-1} \times \left[P_{O^+}(J \times \widehat{\theta})\right]$, with

$$^t\widehat{\theta} = \begin{pmatrix} 0.311 & 0.500 & 0.522 & 0.794 & 0.854 & 0.873 & 0.974 & ..... & 0.571 \end{pmatrix};$$
$$^t\widehat{\theta}^0 = \begin{pmatrix} 0.312 & 0.500 & 0.523 & 0.795 & 0.855 & 0.870 & 0.870 & ..... & 0.870 \end{pmatrix};$$

The missing components of $\widehat{\theta}$ can be found in Table 2, and we have $\widehat{\theta}^0_6 = \ldots = \widehat{\theta}^0_{13}$. Hence $\widehat{\theta}^0$ is the vector of parameters with a weakly increasing shape that is the

closest to our estimation with respect to the metric ${}^tJMJ = [\widehat{V}(\widehat{\theta})]^{-1}$. Both vectors are represented in Figure 3 (with θ replaced by g_1). From our data, the statistic W^+ is equal to

$$W^+ = ||f(\widehat{\theta}) - f(\widehat{\theta}^0)||^2_M = ||\widehat{\theta} - \widehat{\theta}^0||^2_{[\widehat{V}(\widehat{\theta})]^{-1}} = 43.18,$$

The statistic W^+ is a squared distance between the unconstrained estimation of g_1 and the set of weakly increasing functions of adp. If F is the distribution function of a $\chi^2(13)$ variable, we have that $F(43.18) = 0.999958$. If the null assumption: $f(\theta) \in O^+$ is tested for with rejection regions of the type $[W^+ > a]$, it will be rejected at any significance level greater than $1 - 0.999958 = 42 \times 10^{-6}$. To see this, it is enough to apply the first order stochastic dominance property given before. Then the increasing shape of g_1 is rejected at every usual significance level, which shows evidence of moral hazard.

A.5 Elasticity between the optimal frequency risk and the incentive level

We relate the shape of λ and the elasticity between optimal frequency risk and the incentive level. Let us perform a local expansion around a value Δu^0 of the incentive level, in a situation where the incentives are effective (i.e. $\Delta u^0 > \underline{\Delta u} = -1/\lambda'(0)$). If we write

$$e^0 = e_{opt}(\Delta u^0),\ e^0 + de = e_{opt}(\Delta u^0 + d\Delta u),$$

the equations

$$1 + \lambda'(e^0)\Delta u^0 = 0;\ 1 + \left[\lambda'(e^0 + de)\left(\Delta u^0 + d\Delta u\right)\right] = 0$$

lead to

$$de = \frac{-\lambda'(e^0)}{\lambda''(e^0)\Delta u^0} d\Delta u + o\left(d\Delta u\right),$$

and to

$$\frac{d\lambda}{\lambda(e^0)} = \frac{\lambda^{'}(e^0)}{\lambda(e^0)}de = \frac{\left[-\lambda^{'}(e^0)\right]^2}{\lambda(e^0) \times \lambda^{''}(e^0)} \times \frac{d\Delta u}{\Delta u^0}.$$

Hence the aforementioned elasticity is equal to $\left(\lambda^{'}\right)^2/\lambda\lambda^{''}$. Now we have that

$$(\log \lambda)^{''} = \frac{\lambda^{''}}{\lambda} - \left(\frac{\lambda^{'}}{\lambda}\right)^2 = \frac{\lambda^{''}}{\lambda}\left(1 + \frac{\left(\lambda^{'}\right)^2}{\lambda\lambda^{''}}\right).$$

Then the conclusions given in Section 4.2 are easily obtained.

A.6 Monetary equivalents of traffic violations

Let us suppose that the increase in the argument of c_λ is close to 12% after a traffic violation. This is the value retained in Section 4.2 for a driver with seven demerit points accumulated, which corresponds to $n = 3$ traffic violations on average. As the argument of c_λ in the model without point removal and with fines is equal to $\overline{fa} + u_n - u_{n+1}$ (see equation 13) we have that

$$\overline{fa} + u_{n+1} - u_{n+2} = 1.12 \times \left(\overline{fa} + u_n - u_{n+1}\right). \tag{17}$$

We shall compare the utility losses $u_{n+1} - u_{n+2}$ and $u_n - u_{n+1}$ from the recurrence equation on lifetime utility, and obtain a monetary equivalent of an additional traffic violation from a derivation of the utility loss $u_n - u_{n+1}$. We have

$$u_{n+1} = g(u_n),\ g(u) = \overline{fa} + u - c_\lambda^{-1}(d_u - ru)$$

(see Figure 4). From the equality $c_\lambda^{'}(\Delta u) = \lambda(e_{opt}(\Delta u))$ (see equation (10)), we obtain

$$g^{'}(u_n) = 1 + \frac{r}{\lambda_n},\ \lambda_n = \lambda(e_{opt}(\overline{fa} + u_n - u_{n+1})).$$

The parameter λ_n is the frequency risk corresponding to the optimal effort exerted with n traffic violations accumulated. As λ_n decrease with n, we have that

$$1 + \frac{r}{\lambda_n} \leq \frac{u_{n+1} - u_{n+2}}{u_n - u_{n+1}} = \frac{g(u_n) - g(u_{n+1})}{u_n - u_{n+1}} \leq 1 + \frac{r}{\lambda_{n+1}}. \tag{18}$$

From equations (17) and (18), we obtain

$$\left(\frac{r}{\lambda_n^{'}} - 0.12\right) \times (u_n - u_{n+1}) = 0.12 \times \overline{fa} = 15.6\,\$,\ \lambda_{n+1} \leq \lambda_n^{'} \leq \lambda_n.$$

The monetary equivalent of an additional traffic violation is then bounded as follows:

$$\Leftrightarrow \frac{15.6\,\$}{\frac{r}{\lambda_{n+1}} - 0.12} \leq u_n - u_{n+1} \leq \frac{15.6\,\$}{\frac{r}{\lambda_n} - 0.12}.$$

Section 4.2 provides numerical examples with $\lambda_n = 0.15,\ \lambda_{n+1} = 0.12$. The monetary cost of a license suspension follows from a sum of the items related to traffic violations and from the inequalities given in (18).

References

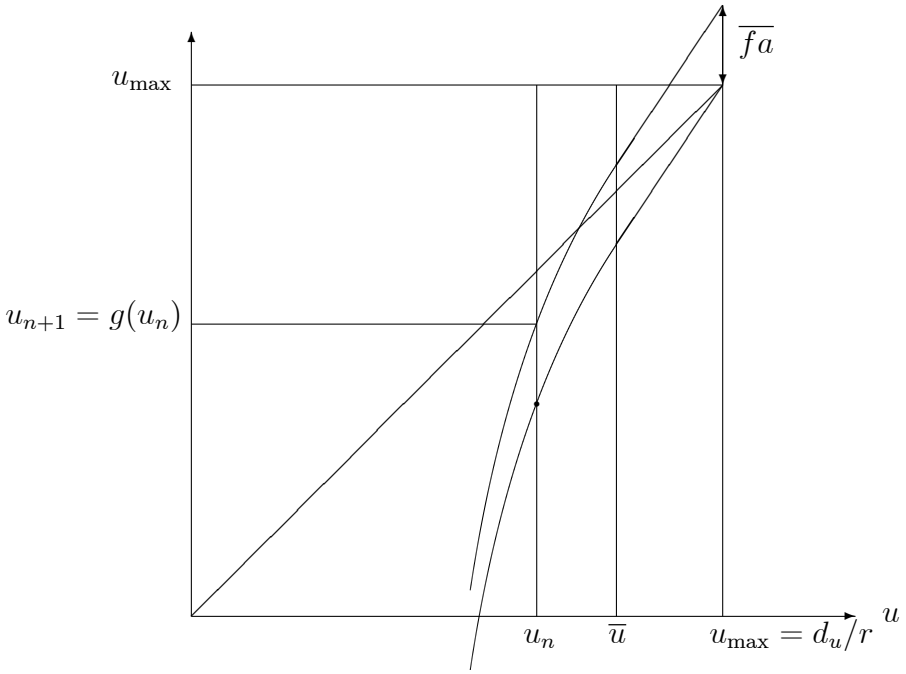
[1] Wolak, Frank A., "The Local Nature of Hypothesis Tests Involving Inequality Constraints in Nonlinear Models," *Econometrica* 59 (1991), 981-995.

Figure 4

Recurrence equation on the lifetime utility function

Point-record driver's license without fines: $u_{n+1}^0 = f(u_n^0)$

Point-record driver's license with fines: $u_{n+1} = g(u_n)$



$$f(u) = u - c_\lambda^{-1}(r(u_{\max} - u)),\ g(u) = f(u) + \overline{fa}.$$

Effective incentives condition with and without fines

$$e_n > 0 \Leftrightarrow u_n < \overline{u} = u_{\max}\left(1 + \frac{\lambda(0)}{\lambda'(0) \times d_u}\right).$$

1 Incentive effects of point-record mechanisms

1.1 Models for point-record driving licenses with redemption

In Quebec, each traffic violation is redeemed at the end of a two year period. Integrating this feature to the optimal behavior model is difficult, as all the seniorities of non-redeemed driving offenses must be included as state variables in the dynamic programming equations. Another redeeming system consists in cancelling all the demerit points after violation-free driving of a given duration, say T. Such mechanisms are also enforced in the real world (see Section **??**). An optimal behavior model is easier to design in this framework since only the seniority of the last convicted driving offense must be added to the number of demerit points accumulated as a state variable. Later on, we will denote the latter redeeming mechanisms as of Type I, and the Quebec-like systems as of Type II.
The model of the preceding section can be extended to a Type I redeeming system (see Appendix 1.2). The conclusions are the following.

- For every given number of demerit points, optimal effort increases continuously with time. If optimal effort is greater than zero, it is strictly increasing.
- When all the demerit points are redeemed (i.e. after a violation-free driving record of duration T), the effort collapses to the minimum level.
- Effort variation after a convicted driving offense may be positive or negative. If the last traffic violation immediately follows another one, the variation is positive, but decreases with the duration between the two last offenses. A drop in the effort level is expected if the duration is large enough.

The time-effect of point-record driving licenses on optimal driving behavior is very different with a Type II redeeming system like the one in Quebec. We do not have rigorous proofs to provide in that intricate framework, but we use the

preceding results to guess the main qualitative properties of Type II redeeming mechanisms.

For both types, lifetime utility is expected to increase with time for a given number of demerit points accumulated. However the link between utility and effort is different according to the type. Optimal effort depends on the difference between the present utility and a substitute utility (i.e. that reached after an additional traffic violation). With a Type I mechanism, the substitute utility only depends on the number of demerit points accumulated as the time variable is reset to zero after a traffic violation. Hence we obtain an increasing link between time and effort. With a Type II redeeming system, all the seniorities of past traffic violations are kept as state variables after an additional traffic violation, and the substitute utility increases with time as the present utility. Time should have more value for worse situations, hence the substitute utility should increase faster than the present utility. Thus optimal effort should decrease with time. Besides, we prove in Appendix 1.2 that optimal effort is continuous before and after a redemption, a property which does not hold for a type I mechanism. This property will be tested empirically in Section **??**. Optimal effort is then expected to increase at each traffic violation in order to compensate the decreasing link between time and effort. On the whole, the incentive properties of a Type II system are closer to those of a mechanism without redemption than to those of a Type I system.

1.2 Incentive effects of point-record driving licenses: Models with redemption

We derive below the incentive properties of the Type I redeeming systems presented in Section 1.1.

The expected utility is denoted as $u_n(t)$, where t $(0 \leq t \leq T)$ is the seniority of the last convicted traffic violation. The expected utility for $n = 0$ does not depend on time, and is denoted as u_0. All the demerit points are redeemed if $t = T$, and we have $u_n(T) = u_0$ for $n = 1, \ldots, N-1$. With the assumptions and notations of Section **??**, the Bellman equation is

$$u_n^{'}(t) = (ru_n(t) - d_u) + \lambda_*(u_n(t) - u_{n+1}(0)) \ (0 \leq n < N)$$

$$\Leftrightarrow u_n^{'} = f_{u_{n+1}(0)}(u_n),\ f_v(u) = (ru - d_u) + \lambda_*(u - v). \tag{1}$$

Hence $u_n(t)$ is defined implicitly by the equation

$$\int_{u_n(t)}^{u_n(T)=u_0} \frac{du}{f_{u_{n+1}(0)}(u)} = T - t,\ \forall t \in [0, T]. \tag{2}$$

The functions f_v are strictly increasing and the integral of $1/f_v$ diverges in the neighborhood of the value which nullifies f_v because we have

$$r < f_v^{'}(u) \leq r + \lambda(0)\ \forall u, v.$$

Hence $u_{n+1}(0)$ is defined from $u_n(0)$ from equation (2) with $t = 0$, which implies that

$$f_{u_{n+1}(0)}(u_n(0)) > 0.$$

This condition holds if $f_v(u_n(0)) > 0$ for positive values of v, which amounts to

$$u_n(0) > \underline{u},\ \text{with}\ f_0(\underline{u}) = (r\underline{u} - d_u) + \lambda_*(\underline{u}) = 0.$$

Then

$$f_{u_{n+1}(0)}(u) > 0\ \forall u \in [u_n(0),\ u_n(T) = u_0]$$

and u_n is strictly increasing on $[0, T]$ from (1).
A result with an obvious economic interpretation is that the sequence $(u_n(0))_{n=0,\ldots,N}$ is decreasing. Indeed, we have

$$f_{u_{n+1}(0)}(u_n(0)) > 0 \Leftrightarrow u_n(0) - u_{n+1}(0) > \lambda_*^{-1}(d_u - ru_n(0)) > 0. \tag{3}$$

Let us prove now the concavity of the sequence $(u_n(0))_{n=0,\ldots,N}$, which means that the sequence $(u_n(0) - u_{n+1}(0))_{n=0,\ldots,N}$ is increasing. The recurrence equation

$$u_{n+1}(0) = w(u_n(0)),\ \text{with}\ \int_v^{u_0} \frac{du}{(ru - d_u) + \lambda_*(u - w(v))} = T \tag{4}$$

follows from (2) with $t = 0$. Consider v_0, v_1 with $\underline{u} < v_0 < v_1 < d_u/r$.
We have

$$\int_{v_1}^{u_0} \frac{du}{(ru - d_u) + \lambda_*(u - w(v_1))} = T;\ \int_{v_0}^{u_0} \frac{du}{(ru - d_u) + \lambda_*(u - w(v_1) + v_1 - v_0)} > T.$$

The first equality results from (4). The second integrand is greater than the first one if the comparison starts from the lower bound of the integral, and the wider range of the second integration reinforces the inequality. As these integrals increase with w, we have that

$$v_0 < v_1 \Rightarrow w(v_1) + v_0 - v_1 > w(v_0).$$

Hence the function $v \rightarrow v - w(v)$ is decreasing. We obtain the desired result with $v_0 = u_{n+1}(0)$ and $v_1 = u_n(0)$.
We can then prove the results given in Section 1.1 on the optimal effort level, which we denote as $e_n(t)$. We have

$$e_n(t) = e_{opt}(u_n(t) - u_{n+1}(0)), \tag{5}$$

where e_{opt} is the increasing function defined in equation (**??**). Since the functions u_n are strictly increasing on $[0, T]$, equation (5) implies the optimal effort level increases with time for a given number of demerit points. From the definition of e_{opt}, the optimal effort is strictly increasing if it is greater than zero.
If the duration between demerit points n and $n + 1$ is equal to t, the transition between the optimal effort levels is

$$e_n(t) = e_{opt}(u_n(t) - u_{n+1}(0)) \rightarrow e_{n+1}(0) = e_{opt}(u_{n+1}(0) - u_{n+2}(0)).$$

If the last traffic violation immediately follows another one, the variation is positive because of the concavity of the sequence $(u_n(0))_{n=0,\ldots,N}$. Since u_n is increasing, the variation of optimal effort after a convicted traffic violation decreases with the duration between the two last offenses.
The following picture gives an example where

$$u_n(t) - u_{n+1}(0) > u_{n+1}(0) - u_{n+2}(0),$$

which implies a drop of the effort level and an increase in risk if the driver is convicted with a supplementary demerit point. This result is in contrast with what is expected with a Type II redeeming mechanism.

Insert Figure 5 about here

Figure 5: Utility functions and variation of the optimal effort level

Optimal effort decreases after a traffic violation in the state (n, t) if and only if

$$u_n(t) - u_{n+1}(0) > u_{n+1}(0) - u_{n+2}(0) \Longleftrightarrow t > t_0.$$