

PROMOTION 2011

Mémoire présenté devant l'Institut des Actuaire Français

pour l'obtention du

Diplôme de l'Institut des Actuaire Français

**Par : Delphine DO HUU
Olivier LOPEZ**

Sur le sujet :

**ETUDE DE STABILITE ET ELABORATION DE PROCEDURES DE
DETECTION DE RUPTURE DANS LE CADRE DE LA
MODELISATION DE SINISTRES AUTOMOBILES**

Tuteur entreprise : Arnaud COHEN
Tuteur pédagogique : Christian ROBERT

CONFIDENTIEL

Table des matières

1	Présentation des données	13
1.1	L'assurance automobile	13
1.1.1	Description des garanties	13
1.1.2	Marché de l'automobile	14
1.1.3	Description des données	15
1.1.4	Présentation de la base sinistres	15
1.1.5	Présentation des variables	16
1.1.6	Autres données techniques disponibles	16
1.2	Analyse des garanties étudiées	17
1.2.1	Stabilité des cotisations	17
1.2.2	Responsabilité civile matériels	18
1.2.3	Responsabilité civile corporels	19
1.2.4	Vol	21
1.3	Conclusion	22
2	Détection de rupture : cadre théorique	23
2.1	Problématique de la détection de rupture	23
2.1.1	Formulation du problème	24
2.1.2	Cadre statique	24
2.1.3	Cadre dynamique	25
2.2	Résultats théoriques généraux	26
2.2.1	Résultats asymptotiques	26
2.2.2	Résultats à distance finie	27
2.2.3	Illustration graphique de phénomènes de rupture	29
3	Etude du nombre de sinistres	31
3.1	Objectifs et difficultés posées	31
3.1.1	Caractéristiques attendues de l'outil de détection	31
3.1.2	Stabilité du portefeuille	32
3.1.3	Présence de saisonnalité	33
3.2	Détection de ruptures multiples dans un processus de Poisson	33
3.2.1	Estimation d'une rupture dans un processus de Poisson	34
3.2.2	Sélection du nombre de ruptures	36
3.2.3	Implémentation	38
3.2.4	Discrétisation du problème	38
3.3	Application aux données	39

3.3.1	Décomposition de l'année	39
3.3.2	Détection de rupture	43
3.3.3	Etude par simulations	48
3.3.4	Apport d'information	52
3.4	Conclusion	52
4	Modélisation des coûts de sinistre	55
4.1	Maximum de vraisemblance gaussien	56
4.1.1	Expression de la vraisemblance	56
4.1.2	Rupture gaussienne et observations non gaussiennes	57
4.1.3	Impossibilité de la mise en oeuvre des tests gaussiens dans notre cadre	57
4.2	Méthode bootstrap basée sur la vraisemblance gaussienne	58
4.2.1	Bootstrap dans les résidus pour la détection de rupture	58
4.2.2	Détection dynamique	59
4.2.3	Application aux coûts moyens mensuels	60
4.2.3.1	Garantie Vol	61
4.2.3.2	Garantie Rccorp	65
4.2.3.3	Garantie Rcmat	67
4.3	Modèle de Pareto	69
4.3.1	Modélisation du coût de sinistre par une variable de Pareto	69
4.3.2	Méthode de détection de rupture pour le modèle de Pareto	70
4.3.3	Simulations	71
4.3.4	Application sur les données	73
4.3.4.1	Garantie Vol	73
4.3.4.2	Garantie Rccorp	74
4.3.4.3	Garantie Rcmat	76
4.3.5	Problèmes posés par la modélisation Pareto	77
4.3.6	Combinaison avec la série des coûts moyens	78
4.3.6.1	Garantie Vol	78
4.3.6.2	Garantie Rccorp	79
4.3.6.3	Garantie Rcmat	80
4.3.7	Comparaison avec l'approche bootstrap sur les coûts moyens	81
4.4	Approche mixte	81
4.4.1	Intérêt potentiel d'une séparation entre "gros" sinistres et "petits" sinistres	81
4.4.2	Applications aux différentes garanties	81
4.4.2.1	Garantie Vol	82
4.4.2.2	Garantie Rccorp	85
4.4.2.3	Garantie Rcmat	88
4.5	Conclusion	91
5	Conséquences sur la mise en œuvre opérationnelle	93
5.1	Comparaison des prédictions de charge par rapport aux sinistres réalisés . .	94
5.1.1	Construction des prévisions de charge	94
5.1.1.1	Nombre de sinistres	94
5.1.1.2	Coûts moyen de sinistre	95

5.1.2	Comparaison des résultats sur l'historique	95
5.1.2.1	Garantie Vol	95
5.1.2.2	Garantie RC Corporels	97
5.1.2.3	Garantie RC Matériels	97
5.2	Prévision de l'exercice 2009	98
5.2.1	Estimation de la charge ultime par liquidation de triangles	99
5.2.1.1	Présentation des triangles et liquidation par la méthode Chain Ladder	99
5.2.1.2	Intervalle de confiance sur la charge ultime	99
5.2.2	Prévision de la charge 2009 à partir des modèles de détection de rupture	100
5.2.2.1	Garantie Vol	100
5.2.2.2	Garantie RC Corporels	101
5.2.2.3	Garantie RC Matériels	102
5.3	Mises en œuvre opérationnelles	102
5.3.1	Reportings et surveillance de portefeuille	103
5.3.2	Provisionnement en gestion et en inventaire	103
5.3.3	Construction de scénarii et projections de résultats techniques . . .	104
5.3.4	Conclusions	105
5.4	Conclusion	105
A	Statistiques descriptives	113
A.1	Répartition des sinistres par mois	113
A.1.1	Garantie Vol	113
A.1.2	Garantie RC Corporels	114
A.1.3	Garantie RC Matériels	114
A.2	Répartition des coûts moyens mensuels	114
A.2.1	Garantie Vol	115
A.2.2	Garantie RC Corporels	115
A.2.3	Garantie RC Matériels	116
	Annexe	113
B	Résumé détaillé de la preuve du Théorème 2	117
B.1	Contrôle des variations d'un processus	117
B.2	Application	118
C	Détection de ruptures multiples dans le nombre de sinistres	121
C.1	Résultats détaillés de la sélection multiple	121
C.2	Heuristique de pente	125
D	Exemples de détection dynamique de rupture et données complémen- taires	127
D.1	Détails de la détection dynamique de rupture	127
D.2	Modélisation des mois d'été pour la garantie Vol	130

Bibliographie

131

Résumé

Dans ce mémoire, nous mettons en place des techniques de détection de rupture qui nous permettent de tester la stabilité dans le temps d'un modèle fréquence-coût portant sur un portefeuille d'assurance non-vie (nous nous concentrons sur trois garanties, RC Corporels, RC Matériels, et Vol). Les modèles retenus sont utilisés pour la prédiction de la charge de l'exercice suivant. Nous modélisons le nombre de sinistres par un processus de Poisson homogène par morceaux, l'année étant découpée en différentes périodes de sinistralité comparable, découpage légitimé par nos méthodes statistiques. Les coûts de sinistre sont modélisés par des variables de Pareto. Les tests de stabilité mis en œuvre reposent sur l'estimation du maximum de vraisemblance, et sur des méthodes de type bootstrap. Nous confrontons les modèles retenus avec les données historiques de notre portefeuille.

Abstract

In this report, we introduce change-point detection techniques, in order to test the stability through time of a frequency-cost model, applied to a non-life insurance portfolio (we focus on three guarantees, bodily injuries, material damage, and theft). The selected models are used to predict the total claim amount of the next year. We modelize the number of claims by a piecewise homogeneous Poisson process, splitting the year into a fix number of periods which have been chosen accordingly to our statistical models. The claim amounts are modeled through Pareto random variables. Testing stability of the model is done using maximum likelihood techniques, and using bootstrap. We compare the prediction obtained through our models to historical data in our portfolio.

Remerciements

Nous remercions tout particulièrement Arnaud Cohen et Hervé Douard, associés au sein du Cabinet ALTIA, pour leur soutien et la confiance qu'ils nous ont témoignée.

Nous adressons également nos remerciement à Olivier Arles, Directeur de l'actuariat Groupe MACIF ainsi qu'à ses collaborateurs Céline Fèvre et Benoît Chaput pour nous avoir permis d'accéder à des données réelles de grande qualité.

Nous remercions Michel Delecroix, Directeur de l'ISUP, ainsi que l'ensemble des membres du LSTA (Université Pierre et Marie Curie) pour leur soutien apporté à ce projet. Nous remercions notamment Bertrand Michel pour son aide concernant l'heuristique de pente.

Introduction

La prévision de la charge de sinistre supportée par les exercices futurs est un enjeu important en assurance. Grâce à cette prévision, l'assureur est à même d'évaluer son besoin de fonds propres, d'ajuster son tarif, ou encore de prévoir son résultat d'activité. A toute approche prédictive préexiste une hypothèse de stabilité du risque sous-jacent. En se basant sur un certain nombre d'exercices passés, on suppose que le comportement du portefeuille reste peu ou prou le même d'une année sur l'autre. Même si l'on suppose une évolution, par exemple une dérive progressive des coûts de sinistres, cette évolution obéit à un modèle stable dans le temps. Cette hypothèse confortable possède l'intérêt - si tant est qu'elle soit vérifiée - de proposer un cadre statistique pertinent. Plus cette stabilité remonte loin dans le passé, plus l'actuaire dispose d'information pour estimer les paramètres de son modèle, et donc pour atteindre une meilleure acuité de la prédiction. Néanmoins, cette hypothèse ne permet pas de prendre en compte les changements de comportement du portefeuille, qu'ils soient dus à une évolution de comportement des assurés, ou à un écart entre le modèle et la réalité qui s'accroît avec le temps. Si c'est le cas, les modèles statistiques mis en œuvre risquent de ne plus être adaptés. Le but de ce mémoire est de proposer une méthode permettant de tester cette stabilité, et de détecter d'éventuelles anomalies dans les modèles de prédiction. Par ailleurs, cette méthode peut représenter un système d'alerte, capable de détecter au plus vite des erreurs de modélisation afin d'agir en conséquence. En effet, d'un point de vue prudentiel, il est important d'alerter au plus tôt sur une défaillance des outils statistiques qui sous-estimerait le risque et pourrait conduire à des conséquences néfastes. A l'inverse, d'un point de vue concurrentiel, il est nécessaire de répercuter au plus vite des éventuelles baisses de la sinistralité.

Le principe d'une telle approche consiste à trouver un compromis entre deux exigences contradictoires. D'un côté, il est nécessaire de mettre en place une approche qui détecte avec une grande précision des écarts, fussent-ils petits, entre l'hypothèse de stabilité et la réalité. A l'opposé, une procédure qui conclurait trop souvent à une instabilité se révélerait contre-productive. En effet, dans une telle situation, il devient impossible d'avoir recours à un modèle statistique pertinent, puisqu'on risque de ne plus disposer de suffisamment d'information. En d'autres termes, il est parfois préférable de se tromper de modèle - tout en restant dans les limites du raisonnable - plutôt que de se retrouver dans l'impossibilité de travailler.

Pour répondre à cet objectif, nous nous sommes orientés vers le problème statistique de la détection de rupture. Ce problème, issu historiquement de la théorie du signal, vise à détecter des instabilités dans un modèle, le plus rapidement possible après la survenance d'une rupture brutale. Le but consiste à essayer de mettre en évidence une modification dans les paramètres du modèle qui sont d'amplitude inconnue, et de date d'occurrence

inconnue. Le problème peut se transposer dans le domaine de l'assurance : jusqu'à une date inconnue, l'évolution de la sinistralité du portefeuille est décrite par un modèle, qui cesse d'être valide par la suite. Dans le cadre de notre approche, cette évolution se traduit par un changement brutal des paramètres qui gouvernent le modèle statistique considéré.

Cette "brutalité" de la rupture n'est pas forcément réaliste. Certains changements de comportements se traduisent par une évolution diffuse. Il ne faut donc pas croire que le modèle de détection de rupture n'est adapté qu'à ce cadre simpliste. Il doit être vu comme une approximation de la réalité, et l'instant de basculement qui est estimé par le modèle ne doit pas être forcément vu comme une date à laquelle la structure du risque est modifiée brusquement, mais plutôt comme une date à partir de laquelle un modèle cesse d'être stable. A partir de cette date, on a intérêt à ne plus tenir compte du passé pour calibrer les paramètres du modèle.

La procédure que nous nous proposons de mettre en œuvre est basée sur des résultats récents qui rendent l'adaptation possible de ces techniques au cadre de l'assurance non-vie. En effet, ils s'émancipent du cadre gaussien - souvent considéré comme le cadre asymptotique de la détection de rupture - en légitimant l'utilisation de simulations pour calibrer au mieux les statistiques de test utilisées. Nous confrontons ces techniques à un portefeuille d'assurance automobile fourni par la MACIF, composé d'un historique de sinistres (1994 à 2009) concernant les trois garanties Vol, Responsabilité Civile Corporels, et Responsabilité Civile Matériels.

Ces problématiques sont d'autant plus d'actualité dans le cadre de l'assurance automobile que l'année 2009 semble présenter les premiers signes d'un renversement d'une tendance à la baisse de la sinistralité. La question de statuer sur le caractère durable ou non de ce changement n'est bien entendu pas possible sans information supplémentaire sur les exercices futurs, mais il importe de pouvoir mettre en place des procédures qui pourront, avec le recul nécessaire, corroborer ou contester cette vision. Pouvoir trancher le plus rapidement une telle question peut avoir des conséquences importantes dans le cadre d'un marché fortement concurrentiel.

Notre approche est à la fois historique et prospective. Nous souhaitons tout d'abord confronter les nouvelles méthodes que nous proposons à des exercices passés (et donc parfaitement connus), afin de pouvoir conclure ou non à leur pertinence. En complément, nous souhaitons également fournir un modèle qui puisse être utile à la prédiction des charges de sinistres à venir, et nous souhaitons observer plus précisément l'année 2009 pour voir si les changements enregistrés par le marché entrent dans le cadre de nos prévisions.

Le mémoire est organisé de la façon suivante. Dans le Chapitre 1, nous présentons les principales caractéristiques du portefeuille que nous considérons. Le Chapitre 2 présente une introduction aux principaux concepts de la détection de rupture, et rappelle un certain nombre de résultats classiques et généraux. Nous développons ensuite une approche fréquence-coût pour modéliser la charge de sinistres du portefeuille. Ces deux aspects du risque sont traités séparément. Le nombre de sinistres enregistrés par l'assureur dans une année est modélisé par un processus de Poisson. En utilisant de nouveaux résultats sur la détection de rupture qui sont présentés au Chapitre 3, nous montrons comment ces techniques peuvent être utilisées d'une part pour identifier et modéliser des phénomènes de saisonnalité, d'autre part pour tester la stabilité du modèle de fréquence considéré. Le Chapitre 4 se concentre sur la modélisation des coûts de sinistres. Cette modélisation est

rendue ardue par la difficulté à trouver une loi qui modélise convenablement les coûts de sinistres. La modélisation classique par une loi de Pareto ne semble pas adaptée à notre portefeuille, un certain nombre d'alternatives sont proposées en conséquence. L'évolution au cours du temps de la loi des coûts de sinistres apparaît par ailleurs plus complexe que celle des nombres de sinistres, et plusieurs méthodes de détection de rupture sont mises en œuvre pour tester leur stabilité. Munis d'un modèle sur la fréquence et d'un modèle sur les coûts, nous les rassemblons au Chapitre 5 afin de commenter leur performance en terme de prévision des exercices futurs.

Chapitre 1

Présentation des données

Dans ce chapitre, nous effectuons une brève présentation des caractéristiques principales du portefeuille que nous étudions. La section 1.1 rappelle un certain nombre de caractéristiques de base du marché de l'assurance automobile français, et définit le périmètre des garanties que nous allons étudier. La section 1.2 présente un certain nombre de caractéristiques des portefeuilles considérés (évolution des cotisations, saisonnalités...). Des statistiques descriptives complémentaires peuvent être consultées dans l'Annexe A.

1.1 L'assurance automobile

1.1.1 Description des garanties

L'assurance automobile offre plusieurs types de garanties mais seule l'assurance Responsabilité Civile est obligatoire.

***L'assurance Responsabilité Civile :** elle permet de couvrir des dommages causés à un tiers (blessures, décès, dégâts matériels causés aux autres véhicules) par l'assuré qu'il soit conducteur, passager ou seulement propriétaire du véhicule. Les sinistres responsabilité civile sont ensuite segmentés en 2 catégories :*

- les sinistres RC Matériels pour lesquels la charge sinistre est généralement rapidement connue et les sinistres rapidement clôturés.*
- les sinistres RC Corporels pour lesquels le délai de clôtures de sinistres peut être long car il dépend de l'état de consolidation médicale du tiers.*

Le caractère obligatoire de cette garantie rend cette dernière particulièrement intéressante pour les besoins de l'étude car elle représente à fin 2009 37% des cotisations automobiles.

Les autres garanties dites "facultatives" se regroupent en 3 catégories : garanties couvrant les dommages subis par le véhicule (vol, incendie, vandalisme, bris de glace, dommage accident), les dommages subis par le conducteur, les garanties de service (protection juridique et assistance). Parmi ces garanties facultatives, notre périmètre d'étude ne concernera que la **garantie Vol**.

1.1.2 Marché de l'automobile

(Source FFSA)

Le marché de l'automobile en France a connu de fortes baisses tarifaires depuis 5 ans accompagnées d'une diminution de la fréquence des sinistres (impact de la prévention routière et de la baisse de la circulation liée l'augmentation du prix des carburants).

Toutefois cette baisse sensible de la fréquence des sinistres est partiellement compensée par la hausse de leurs coûts moyens (augmentation des coûts moyens des pare brises de 3%, hausse tendancielle du coût moyen des sinistres corporels de 6,5% par an depuis 10 ans - source FFSA).

En 2009, le chiffre d'affaires de l'assurance automobile français s'élève à 17,8 milliards d'euros, en légère baisse par rapport à 2008 (-0,2%), pour un parc automobile resté stable de 37 millions de véhicules. L'amélioration de la sinistralité ainsi que la forte concurrence sur ce marché ont permis de baisser les primes notamment en responsabilité civile de 3,4% en 2009 (-2,7% en 2008) à parc constant. Les cotisations d'assurance automobile correspondant à la couverture des garanties facultatives sont, elles, en hausse de 0,7% en 2009.

Depuis 2009, le marché semble connaître un renversement de la tendance baissière des sinistres matériels (cf figure 1.4), alors que le nombre de sinistres corporels enregistrés par les pouvoirs publics continue de diminuer.

Evolution des fréquences de sinistres matériels en 2009.

- Hausse de 2% de la fréquence RC matériels
- Hausse de 6% pour les dommages tous accidents.

Ces évolutions traduisent un renversement de la tendance à la baisse qui était présente sur le marché automobile depuis plusieurs années.

Cette reprise de la fréquence des sinistres s'explique par :

- ***Les événements climatiques*** : Klaus, Quinten et grêles au cours du premier semestre.
- ***Hausse de la circulation routière*** : +1% au second semestre, liée notamment à la baisse moyenne des prix de l'ensemble des carburants (super et Gazole) en 2009.

En revanche, le nombre de sinistres corporels enregistrés par les pouvoirs publics continue de diminuer, et connaît une baisse de 1% de la fréquence des sinistres corporels indemnisés au titre de la RC pour 2009.

Concernant les coûts, on remarque les caractéristiques suivantes.

Coûts moyens de sinistres

- Hausse de 6% en 2009 pour le vol. Explication : évolution technologique et hausse du coût des matières premières.
- Répartition des sinistres corporels : 75% correspondent à des blessés légers (11% des indemnités versées). Les blessés les plus graves (dont le taux d'AIPP est supérieur à 20%) ne représentent que 1% des cas (1500 cas par an environ, pour 42% de la charge d'indemnités versée).

1.1.3 Description des données

Notre étude portera sur les sinistres relatifs aux garanties suivantes :

- Les garanties obligatoires
 - Responsabilité Civile Matériel
 - Responsabilité Civile Corporel
- La garantie Vol afin de tester le comportement des techniques utilisées sur une garantie à volumétrie plus faible.

1.1.4 Présentation de la base sinistres

La base de données à disposition pour réaliser notre étude est constituée de la base de sinistres Automobiles du Groupe MACIF.

Avec 4,9 millions de véhicules "4 roues" assurés, la MACIF est leader sur le marché de l'assurance automobile en France avec un taux de pénétration de 13% du parc automobile nationale et 1.4M de cotisations en 2008.

La base sinistre est constituée de l'ensemble des sinistres survenus entre le 1/1/1994 et le 31/12/2009, **vue au 31/12/2009**. Elle présente un historique de 16 années de survenance avec des données détaillées par sinistre pouvant offrir une vision de la sinistralité journalière.

Toutefois, s'agissant d'une vision de la base sinistre arrêtée au 31/12/2009, les données portant sur les derniers exercices de survenance présentent un niveau d'information moins complet (non stabilisé) que les exercices anciens, ainsi les derniers exercices et notamment 2009 n'intègrent pas tous les IBNR. Ceci explique notamment le fait que l'exercice 2009 semble présenter une sinistralité plus faible avec un taux de sinistres sans suite pour les branches RC encore faible, le marquage des sinistres en sans suite nécessitant des délais liés aux expertises.

Tableaux de volumétrie par garantie.

Année de survenance	RC Matériel		RC Corporel		Vol	
	Nb sinistres avec suite	Nb sinistres sans suite	Nb sinistres avec suite	Nb sinistres sans suite	Nb sinistres avec suite	Nb sinistres sans suite
1 994	214 356	44 969	34 763	11 266	118 507	11 947
1 995	216 552	46 587	36 207	11 685	110 184	11 873
1 996	208 092	49 823	35 118	11 430	106 310	12 176
1 997	215 849	50 352	36 591	11 629	100 915	11 037
1 998	222 493	51 743	37 750	12 853	101 030	11 721
1 999	228 306	58 002	38 892	14 151	98 604	12 262
2 000	220 666	57 385	36 972	13 905	97 613	12 131
2 001	216 518	57 420	34 928	14 117	102 214	11 985
2 002	202 493	52 681	31 530	13 434	96 248	11 503
2 003	190 039	49 126	27 445	11 370	85 804	10 963
2 004	190 201	48 758	27 352	11 346	76 379	10 303
2 005	189 988	47 983	27 062	10 721	67 836	9 773
2 006	186 539	45 824	25 682	10 068	60 418	9 741
2 007	190 173	43 049	25 685	10 187	53 507	9 062
2 008	197 889	26 445	24 340	8 655	46 051	8 197
2 009	202 597	17 722	23 609	6 071	44 990	8 058
Total	3 292 751	747 869	503 926	182 888	1 366 610	172 732

1.1.5 Présentation des variables

Les variables disponibles dans la base de données sont les suivantes.

Variables disponibles

- *numéro de sinistre*
- *code de la garantie*
- *date de survenance du sinistre*
- *date de déclaration du sinistre*
- *date de clôture du sinistre*
- *indicateur corporel/matériel*
- *indicateur sans suite*
- *indicateur clôture*
- *code département du sinistre*
- *montant des règlements au 31/12/2009*
- *montant des frais d'experts et honoraires versés au 31/12/2009*
- *montant des frais d'épave au 31/12/2009*
- *montant des recours encaissés au 31/12/2009*
- *stock de provision de sinistres vu au 31/12/2009*
- *stock de prévisions de recours au 31/12/2009*
- *montant des abandons de recours enregistrés au 31/12/2009.*

Dans la base de données, nous avons assimilé les sinistres avec suite à charge nulle aux sinistres sans suite. Nos études seront réalisées sur les sinistres avec suite (ou assimilés) à partir d'une charge sinistre brute de recours, calculée de la manière suivante :

$$\text{Charge sinistre} = \text{Règlement} + \text{Frais d'expert} + \text{Frais d'épave} + \text{Stocks de provisions}$$

1.1.6 Autres données techniques disponibles

En complément de la base sinistre, nous avons obtenu auprès des services d'actuariat les informations suivantes :

- le montant total des cotisations vues au 31/12/2009 pour les 16 exercices concernés des branches Vol, RC Corporels et RC Matériels. L'évolution du nombre de sociétaires au cours des années ayant été obtenue très tardivement, cette information n'a pu être exploitée dans ce mémoire. Ainsi nous utiliserons l'information sur les cotisations pour normer l'évolution du nombre de sinistres par an (bien que cette approche soit critiquable car le stock de cotisations annuelles ne reflète pas totalement le portefeuille assuré, car il tient également compte de l'évolution de la politique tarifaire).
- les triangles de règlements et de provisions de 1993 à 2009.

Ces triangles seront utilisés dans la dernière partie du mémoire pour être confrontés aux prévisions de charges sur l'exercice courant qui sont obtenues par les différents modèles que nous considérons.

1.2 Analyse des garanties étudiées

1.2.1 Stabilité des cotisations

Les deux graphiques suivant présentent l'évolution des cotisations pour les garanties Responsabilité Civile et Vol.

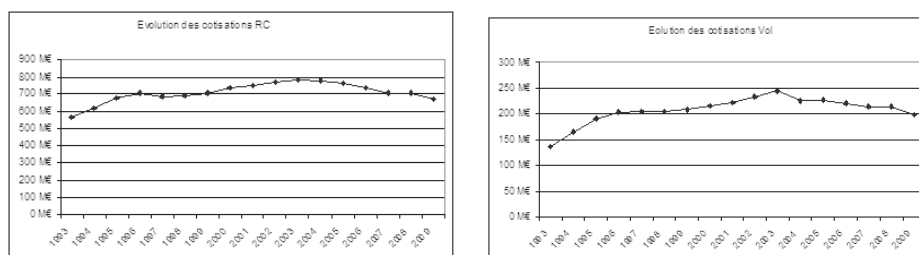


FIGURE 1.1 – Evolution des cotisations

Les cotisations responsabilité civile couvrent à la fois les garanties corporelles et matérielles. Le graphique est donc commun aux deux garanties. L'évolution constatée des cotisations semble cohérente avec les évolutions tarifaires connues sur le marché français, avec une baisse des tarifs constatée depuis 2004 (cf source FFSA). Ainsi, au regard des évolutions il semblerait que le portefeuille d'assurés soit resté globalement stable au cours de ces dernières années¹.

Concernant la garantie Vol, celle-ci étant facultative, il est difficile de juger si l'évolution des cotisations est liée davantage à l'évolution du nombre d'assurés ou de l'effet tarifaire. Toutefois, sur le marché français, le poids des souscriptions des garanties Vol est resté globalement stable depuis 2005 alors que les tarifs ont diminué ce qui semble être cohérent avec l'évolution constatée sur les cotisation MACIF. A titre indicatif, nous mentionnons ces observations issues de la FFSA.

Taux de souscription des garanties dans les contrats des véhicules de 1ère catégorie (1)
(en pourcentage du nombre de contrats)
Source : FFSA

	2005	2006	2007	2008
Vol et incendie	81	81	81	82
Bris de glaces	88	88	88	89
Dommages tous accidents	59	60	61	62
Poids des garanties facultatives dans le chiffre d'affaires automobile	60	61	61	62

(1) Véhicules 4 roues à moteur, de poids inférieur à 3,5 tonnes

FIGURE 1.2 – Taux de souscription garanties facultatives

1. En réalité, le portefeuille MACIF a connu une croissance globalement régulière, de l'ordre de 1,6% par an entre 1996 et 1999.

1.2.2 Responsabilité civile matériels

Nous présentons ici l'évolution du nombre et du coût des sinistres pour la garantie RC Matériels.

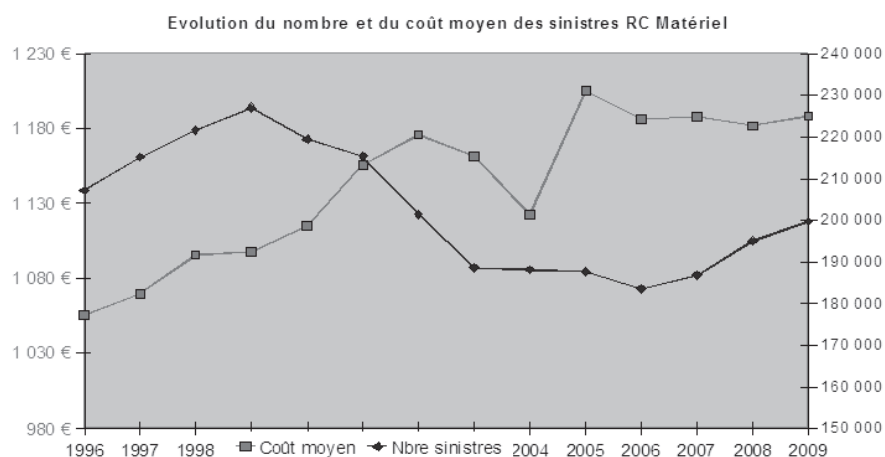


FIGURE 1.3 – Evolution du nombre et du coût des sinistres en RC Matériels.

Globalement, le portefeuille sinistres RC matériels est marqué par une baisse du nombre de sinistres entre 1996 et 2009 de l'ordre de -4%. A l'inverse, les sinistres RC Matériels ont connu une croissance cumulée de 14% du coût moyen des sinistres entre 1996 et 2005, légèrement inférieure à l'inflation constatée sur cette même période (15%, calculée à partir de l'indice des prix à la consommation).

A titre indicatif nous présentons les tableaux d'évolution de la sinistralité RC Matériels de la FFSA.

Evolution de la fréquence et du coût moyen des sinistres RC matériels avec suite
Source : FFSA

Année	Fréquence (en %)	Variation (en %)	Coût moyen (en euros)	Variation (en %)
2005	42,4	- 3,0	1 460	+ 1,0
2006	40,3	- 5,0	1 490	+ 2,0
2007	39,4	- 2,2	1 493	+ 0,2
2008	37,5	- 4,9	1 454	- 2,6

FIGURE 1.4 – Evolution de la fréquence et du coût moyen des sinistres RC Matériels avec suite.

Le graphique ci-dessous indique la présence d'une saisonnalité manifeste sur la série du nombre de sinistres.

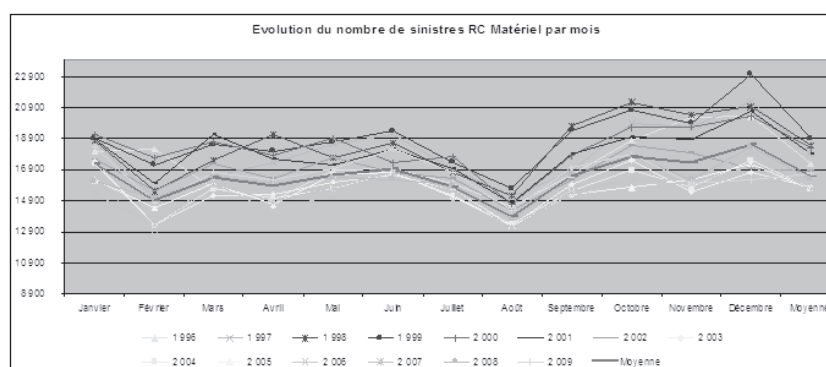


FIGURE 1.5 – Saisonnalité dans le nombre de sinistres en RC Matériels.

On note des écarts moyens de 26% entre août et décembre. La fréquence de sinistres semble donc marquée à la fois par des effets saisonnalités au sein d'une année et une évolution globalement à la baisse des sinistres sur les 16 années étudiées.

1.2.3 Responsabilité civile corporels

L'évolution de la sinistralité pour la garantie RC Corporels est présentée dans le graphique suivant.

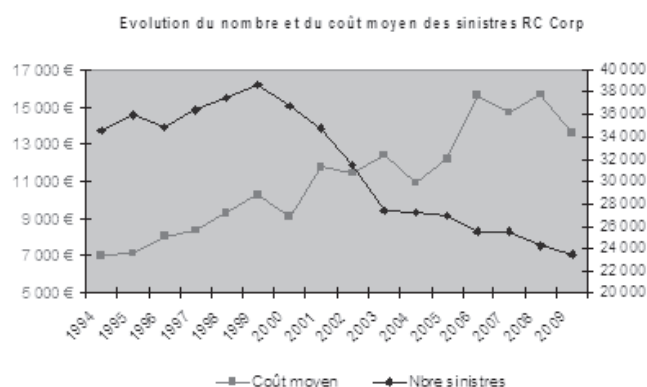


FIGURE 1.6 – Evolution de la fréquence et du coût des sinistres, garantie RC corporels.

Depuis 2000, le portefeuille est marqué par une baisse de la fréquence des sinistres de 36% avec une décroissance forte sur la période 2000-2003, période qui coïncide avec le développement des contrôles radar. Cette évolution est cohérente avec la baisse des accidents corporels enregistrés par les pouvoirs publics (figure 1.7).

Evolution du nombre des accidents corporels
 Source : Ministère de l'Équipement, des Transports, du Logement, du Tourisme et de la Mer – ONISR

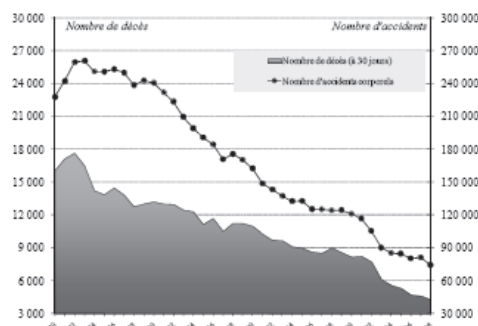


FIGURE 1.7 – Evolution du nombre d'accidents corporels.

En revanche, la garantie RC Corporels se caractérise par une augmentation régulière des coûts de l'ordre de 4,5% par an. Cette évolution est en ligne avec les données de marché qui selon les données FFSA s'explique essentiellement par les sinistres corporels "lourds". Si le nombre de blessés et de tués sur la route a baissé de 40% de 1997 à 2007, le nombre de sinistres corporels supérieurs à 3 millions d'euros enregistrés par les assureurs, a été multiplié par près de six sur la même période.

La figure 1.8 ci-dessous montre elle aussi la présence de fortes saisonnalités dans le nombre de sinistres.

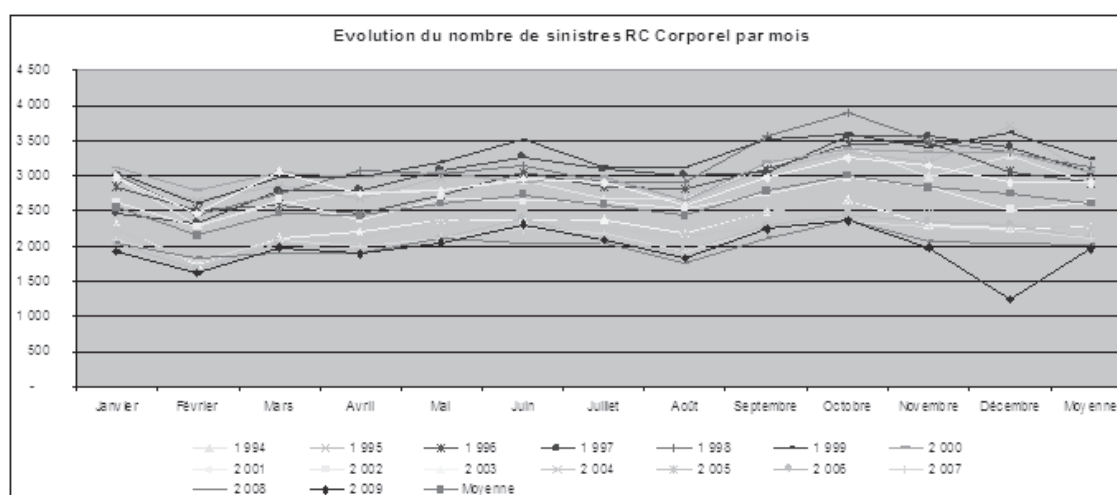


FIGURE 1.8 – Saisonnalité dans le nombre de sinistres en RC Corporels.

Ainsi, on a un écart moyen entre février et octobre de l'ordre de 40%.

1.2.4 Vol

Pour la garantie Vol, nous présentons l'évolution de la sinistralité du portefeuille dans la figure suivante.

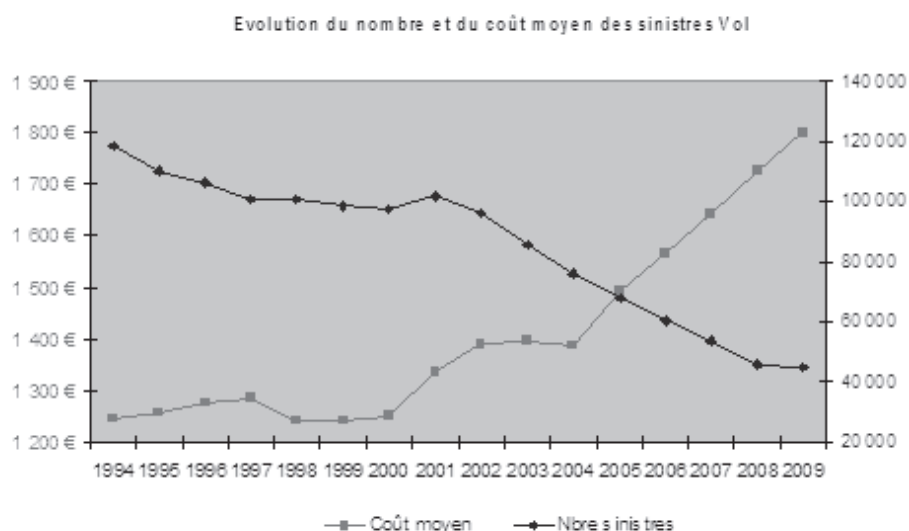


FIGURE 1.9 – Evolution de la fréquence et du coût moyen de sinistre pour la garantie Vol.

Le portefeuille vol se caractérise par une baisse du nombre de sinistres (6% par an) avec une accélération de la baisse à partir de 2001. En revanche, les coûts des sinistres connaissent depuis 2000 une forte progression de l'ordre de 4% par an. Ces évolutions sont globalement cohérentes avec les statistiques FFSA et s'explique notamment par les évolutions technologiques et le coûts des matières premières.

Evolution de la fréquence et du coût moyen des sinistres vol
Source : FFSA

Année	Fréquence (en %)	Variation (en %)	Coût moyen (en euros)	Variation (en %)
2005	12,0	- 12,0	2 060	+ 4,0
2006	10,6 *	- 11,0	2 060	--
2007	9,4 *	- 11,5	2 100	+ 2,0
2008	8,1 *	- 13,3	2 178	+ 3,7 %

* FFSA+GEMA

FIGURE 1.10 – Evolution pour la garantie Vol, statistiques FFSA.

La présence de saisonnalité dans le nombre de sinistres est une nouvelle fois constatée.

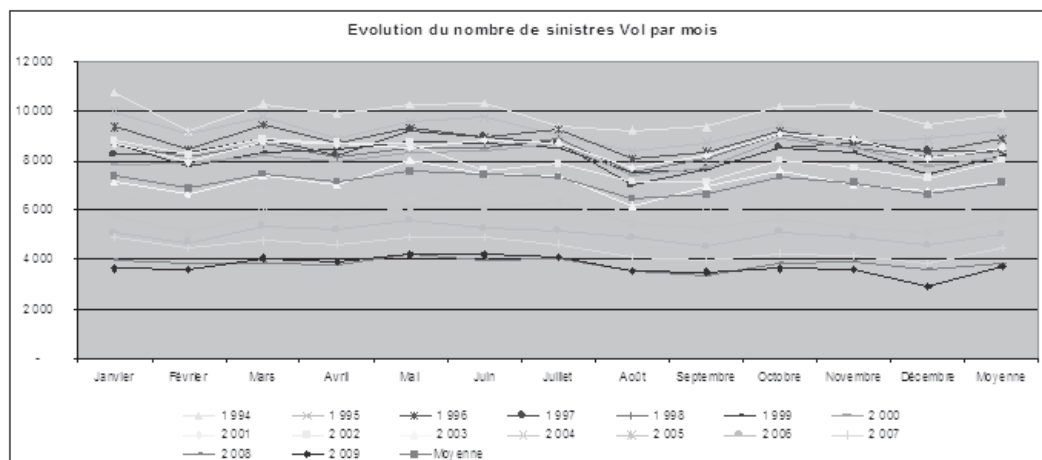


FIGURE 1.11 – Saisonnalité dans la fréquence de sinistres pour la garantie Vol.

Néanmoins, l'amplitude des effets saisonniers semblent a priori moins marquée, même si les mois d'été présentent une sinistralité généralement plus faible.

1.3 Conclusion

Le portefeuille que nous considérons semble être représentatif des évolutions constatées au sein du marché français. Au moment de notre étude, ne possédant pas d'information sur le nombre exact d'assurés présents dans le portefeuille au cours du temps, nous ne pouvons étudier la stabilité du parc assuré qu'à travers le volume de cotisations versé. Ce volume semble suivre les évolutions générales du marché et semble donc correspondre à un portefeuille relativement stable. Parmi les spécificités que nous pouvons constater sur nos données, nous remarquons une saisonnalité marquée dans le nombre de sinistres (la garantie Vol offre la spécificité de présenter une saisonnalité également dans les coûts moyens de sinistres d'un mois à l'autre, comme nous le verrons au Chapitre 4). Cette saisonnalité est attendue, puisque, par exemple, le nombre d'accidents automobiles apparaît naturellement lié aux conditions météorologiques. Nous verrons néanmoins que cette saisonnalité complique notre travail dans l'étude de la stabilité des modèles de prévision (voir Chapitre 3).

Chapitre 2

Détection de rupture : cadre théorique

Les méthodes de détection de rupture ont pour but d'étudier la stabilité de la modélisation d'un processus au cours du temps (ou à travers l'espace, mais ce cadre ne sera pas envisagé dans notre travail), en essayant de détecter la date à partir de laquelle un changement de comportement s'opère dans la loi du processus. Du point de vue des applications que nous avons en vue, on pense par exemple à un changement de la fréquence des accidents d'un assuré (dû par exemple à l'efficacité des politiques de prévention routière), ou un changement dans l'évolution temporelle des coûts de sinistres.

Dans ce chapitre, nous présentons succinctement le cadre théorique de la détection de rupture, en insistant sur la problématique, ainsi que sur le vocabulaire que nous emploierons dans tout le reste de ce mémoire. La section 2.1 présente la formulation générale du problème statistique, que nous développerons dans plusieurs cadres dans les chapitres suivants (détection de rupture dans des processus de Poisson pour l'étude du nombre de sinistres, dans des suites de lois de Pareto pour l'étude des coûts de sinistres). Du point de vue méthodologie, nous nous concentrons sur les approches basées sur la vraisemblance, puisque c'est celle que nous employerons par la suite. La section 2.2 liste un certain nombre de résultats théoriques existant, leurs avantages ainsi que leurs insuffisances pour le problème qui nous intéresse.

2.1 Problématique de la détection de rupture

Du point de vue statistique, deux cadres distincts (mais complémentaires) existent pour la détection de rupture. Le cadre statique (ou retrospectif) consiste à disposer de la totalité des observations d'un processus, et à chercher l'endroit où une (ou plusieurs) rupture a pu apparaître. Il s'agit d'un cadre d'estimation classique. Le cadre dynamique est sans doute le plus intéressant dans l'optique de définir un système d'alerte utilisable par l'assureur, puisqu'il s'agit de traiter les données au fur et à mesure qu'elles arrivent, la quantité d'observations évoluant donc au cours du temps. Le cadre dynamique nécessite l'emploi successif de techniques importées du cadre statique, mais pose des problèmes spécifiques qui seront décrits dans la section 2.1.3.

2.1.1 Formulation du problème

Considérons un processus $(Y_t)_{t \in T}$ à valeur dans $\mathbb{R}$. Dans tout ce qui suit, T sera soit un sous-ensemble de $\mathbb{R}$ (temps continu) ou $\mathbb{N}$ (temps discret). Soit $(\mathbb{P}_\theta)_{\theta \in \Theta}$ une famille de lois de probabilités sur $\mathbb{R}^T$, où $\Theta \subset \mathbb{R}^k$ est un espace paramétrique. Un modèle à une rupture consiste à supposer que

$$\exists t_0 \text{ t.q. } (Y_t)_{t \leq t_0} \sim \mathbb{P}_{\theta|_{\{t \leq t_0\}}}, (Y_t)_{t > t_0} \sim \mathbb{P}_{\theta'|_{\{t > t_0\}}}, \quad (2.1)$$

où $\theta' \neq \theta$, et $\mathbb{P}_{\theta|_A}$ désigne la restriction de $\mathbb{P}_\theta$ sur l'ensemble des $t \in A$.

En d'autres termes, le processus $(Y_t)_{t \in T}$ a deux comportements distincts au cours du temps : jusqu'à la date t_0 , sa loi est $\mathbb{P}_\theta$, après t_0 , sa loi change et devient $\mathbb{P}_{\theta'}$.

Vocabulaire :

- L'instant t_0 sera désigné par le terme **instant (ou date) de rupture**.
- Nous appellerons θ le **paramètre de base**, et θ' le **paramètre après rupture**.
- Si on définit une distance d sur l'espace Θ (dans le cas d'une paramétrisation unidimensionnelle, $d(\theta, \theta') = |\theta - \theta'|$), on parlera **d'amplitude de rupture** pour désigner $d(\theta, \theta')$.

D'un point de vue statistique, les paramètres θ' et t_0 sont inconnus et doivent être estimés. Le paramètre θ peut être soit connu, soit inconnu suivant les cas. Le modèle (2.1) nécessite généralement d'imposer des conditions d'identifiabilité sur les paramètres, qui seront discutées au cas par cas dans les cadres particuliers où nous considérerons ce modèle.

Un certain nombre de remarques s'impose concernant le modèle (2.1) :

- Ce modèle n'a pas vocation à capter des phénomènes exceptionnels et localisés dans le temps (points atypiques d'une série temporelle, par exemple, suivis d'un retour à la normale).
- Le modèle de rupture capte des changements de régime d'un processus qui persistent dans le temps.

Le modèle à une rupture peut être complexifié en modèle à k ruptures.

2.1.2 Cadre statique

Dans un cadre statique de détection de rupture, les observations sont constituées d'une réalisation d'un processus $(Y_t)_{0 \leq t \leq \tau}$, où τ est une date **fixée**. En supposant, pour simplifier, que le paramètre de base θ est connu, les deux quantités à estimer sont la date de rupture t_0 , et le paramètre après rupture θ' .

La manière la plus standard de procéder pour estimer (t_0, θ') consiste à exprimer la vraisemblance du modèle (2.1) et à la maximiser. Si le processus Y_t est à accroissements indépendants, la maximisation de cette vraisemblance peut se décomposer généralement en deux parties :

1. Pour chaque date candidate t , on traite un problème paramétrique classique en maximisant la vraisemblance $L(t, \theta')$ par rapport à θ' .
2. On maximise ensuite par rapport à t .

Cette dernière étape peut être délicate dans un cadre général si l'ensemble T est continu, mais par la suite, nous parviendrons toujours à nous ramener à un sous-ensemble fini de T .

Une fois obtenu un estimateur $(\hat{t}, \hat{\theta}')$ du paramètre (t_0, θ') , on pourra tester la stabilité du processus en effectuant le test statistique suivant,

$$\begin{aligned} H_0 &: t_0 = \tau \text{ (pas de rupture),} \\ H_1 &: t_0 < \tau \text{ (rupture).} \end{aligned}$$

Les tests que nous considérerons seront basés sur le rapport de vraisemblance, i.e. seront de la forme : "rejet de H_0 si $L(\hat{t}, \hat{\theta}') > s_\alpha$," où s_α est un seuil permettant d'atteindre le niveau α (on parlera par la suite de **seuil d'alerte**).

2.1.3 Cadre dynamique

Dans un cadre dynamique, la problématique est sensiblement différente, puisque l'information disponible pour estimer la date de rupture s'accroît progressivement à chaque instant. A chaque instant, on s'interroge sur l'occurrence ou l'absence d'une rupture dans le processus. Le cadre statique est ainsi un cadre rétrospectif : à une date donnée, on s'interroge sur la présence de rupture dans le passé. Dans le cadre dynamique, le processus de décision est plus complexe, puisqu'il s'agit d'une surveillance continue du processus : alors que le cadre statique donne un jugement définitif, le cadre statique se propose de réviser à chaque instant la décision de considérer que le modèle a connu ou non une rupture.

A chaque instant t , l'information disponible est constituée de l'observation du processus $(Y_s)_{s \leq t}$. A partir de cette observation, on voit si une rupture est présente ou non, et si oui à quelle date elle s'est produite, avec quelle amplitude. Une détection dynamique se résume donc par une suite de décisions (déclenchement ou non de l'alerte), et d'estimations.

Vocabulaire et problématiques spécifiques :

- On parlera de **fausse alarme** si, au cours d'une détection dynamique, on conclut à l'occurrence d'une rupture alors que celle-ci ne s'est pas produite.
- On s'autorise une proportion α de fausses alarmes sur la durée globale d'observation. Ce α s'interprète comme un niveau de test, l'hypothèse H_0 étant l'absence de rupture sur **toute** la durée d'observation.
- On parlera de **délai de détection** pour désigner la différence entre la date de rupture t_0 et la première date ultérieure à laquelle la décision est prise de déclencher l'alerte.
- A un instant t , on parlera de **date détectée** pour désigner l'estimateur $\hat{t}$ de l'instant de rupture t_0 .

La méthode du maximum de vraisemblance présentée dans le cadre statique est utilisée à chaque instant pour estimer la date potentielle de rupture ainsi que son amplitude. Néanmoins, du point de vue décisionnel, le cadre dynamique ne se résume pas à une succession d'étapes statiques. En effet, si l'on effectue à chaque instant un test de niveau α comme dans le cadre statique, on évacue deux problèmes importants :

1. D'un instant à l'autre, les statistiques de tests sont dépendantes l'une de l'autre (on a considéré les mêmes observations, en rajoutant une ou plusieurs d'un instant sur l'autre).
2. Même s'il y avait indépendance, la méthode aboutirait à une **sur-détection** (i.e. un pourcentage de fausses alarmes trop important), puisque même une trajectoire sans rupture finirait systématiquement par franchir le seuil d'alerte si le temps d'observation tend vers l'infini.

Une autre solution consisterait à baisser le niveau du test statique de α à $\alpha' < \alpha$, pour conserver la bonne proportion α de fausses alarmes. Mais si la durée d'observation totale est potentiellement infinie, α' devrait être pris égal à zéro.

La question cruciale est donc de déterminer des **courbes d'alertes** appropriées, différentes de la **courbe des seuils**, i.e. la courbe des $s(\alpha)$ du cadre statique.

***Définition d'une courbe d'alerte de niveau α :** - Soit L_t la valeur du maximum de vraisemblance obtenu à partir de toutes les observations jusqu'à la date t . On appellera courbe d'alerte au niveau α une fonction $t \rightarrow C(t)$ telle que*

$$\mathbb{P}(\exists t \in T : L_t > C(t) | t_0 = \infty) \leq \alpha.$$

En d'autres termes, un processus ne connaissant aucune rupture engendrera une courbe L_t qui ne dépassera jamais $C(t)$ sauf dans une proportion α des trajectoires. La détermination d'une courbe d'alerte de niveau approprié est en général difficile, et la question de son optimalité est ouverte. Une courbe $C(t)$ doit en effet répondre à deux exigences contradictoires :

1. *Ne pas détecter à tort trop de ruptures, ce qui engendrerait une instabilité trop grande dans la modélisation, et rendrait le modèle inutilisable. Exemple extrême : le modèle qui détecte un changement de loi à chaque instant est inutilisable du point de vue statistique.*
2. *Détecter suffisamment rapidement des ruptures d'amplitude même faible.*

2.2 Résultats théoriques généraux

2.2.1 Résultats asymptotiques

L'ouvrage de Csörgö et Horváth (1997) recense un nombre important de résultats généraux sur la détection de rupture. On pourra également consulter Basseville et Nikiforov (1993), et l'article de Haccou, Meelis, Van de Geer (1988) dans le cas particulier de la détection de rupture pour la loi exponentielle. Ils présentent tous l'inconvénient d'être asymptotiques et de reposer sur un comportement asymptotique gaussien.

A titre d'exemple, on pourra consulter le Théorème 1.6.3 de Csörgo et Horváth (1997) qui fournit la convergence en loi de l'estimateur $\hat{t}$, sous réserve que $(\tau - t_0)$ tende vers l'infini à une certaine vitesse (voir leur hypothèse F3), où t_0 désigne la date de rupture, et τ la durée totale d'observation. Si le raisonnement asymptotique fournit des approximations utiles des lois des estimateurs dans de nombreux cadres statistiques, il pose problème dans le cadre de la détection de rupture, où l'on voudrait précisément identifier les éventuelles ruptures le plus tôt possible après leur occurrence, c'est à dire précisément quand $(\tau - t_0)$ est petit.

2.2.2 Résultats à distance finie

Le problème principal de la plupart des résultats à distance finie existant sur le sujet vient du fait qu'ils se concentrent essentiellement sur le cas de variables gaussiennes, qui nous intéresse peu dans notre problème, où nous avons surtout en tête de l'appliquer à des lois de Poisson (modélisation du nombre de sinistres), et à des lois de Pareto (modélisation du coût de sinistre).

Citons notamment Lebarbier (2003) qui propose une technique de sélection du nombre de ruptures dans un processus gaussien, avec des résultats non-asymptotiques. Nous nous attardons sur le résultat proposé par Golubev et Spokoiny (2009), présenté comme un exemple de leur nouvelle technique d'obtention de bornes exponentielles. Même s'il ne s'applique pas dans notre cadre, son énoncé met en lumière un certain nombre de points qui seront utiles par la suite à la compréhension. Nous reformulons leur résultat dans le Théorème suivant.

Théorème 1. *Considérons une suite $(Y_i)_{1 \leq i \leq n}$ de variables indépendantes $\mathcal{N}(m_i, \sigma^2)$ avec σ^2 connu, et*

$$m_i = m \mathbf{1}_{i > t_0},$$

avec m et t_0 inconnus. Soit $\hat{t}$ l'estimateur du maximum de vraisemblance de t_0 . Alors

$$E \left[\frac{|\hat{t} - t_0|}{\max(\hat{t}, t_0)} \right] \leq \frac{C \sigma^2 \log \log n}{m^2 (n - t_0)}.$$

Une fois obtenu le résultat sur l'estimateur $\hat{t}$, il est aisé d'obtenir un résultat sur l'estimateur $\hat{m}$, qui est défini comme $(n - \hat{t})^{-1} (\sum_{i=\hat{t}+1}^n Y_i)$ (il suffit d'utiliser des résultats de convergence uniforme des moyennes de gaussiennes). Nous pouvons effectuer un certain nombre de remarques concernant le Théorème 1 :

– Plus σ^2 est important, plus la performance se dégrade.

- La "vitesse" de convergence dépend de la valeur de $m^2(n - t_0)$. Il faut donc que l'amplitude de rupture (m) soit suffisamment importante relativement au temps d'exposition à cette rupture ($n - t_0$). Une rupture d'amplitude importante sera détectée avec une grande précision même si on a peu de recul, c'est à dire si t_0 proche de n . Inversement, on aura une forte imprécision si la rupture est très petite, à moins qu'elle se soit produite très loin dans le passé.
- Un terme $\log \log n$ apparaît, qui vient augmenter l'incertitude quand la durée d'observation augmente (toutes choses égales par ailleurs). Ce terme a un impact assez faible pour les ordres de grandeur usuels. Il est intéressant de noter que, dans le cas particulier où m serait connu exactement, ce terme n'apparaîtrait pas, il est uniquement dû à l'estimation de m .

Le résultat du Théorème 1 peut être étendu relativement facilement à des lois de la famille exponentielle, c'est à dire qui possèdent une densité par rapport à une mesure dominante μ du type

$$f_\theta(u) = \exp(u\theta - d(\theta))l(u),$$

où l est une fonction positive, et d caractérise le modèle. Néanmoins, cette extension nécessite des conditions sur d trop fortes pour les cas que nous considérons (typiquement, un comportement quasi-gaussien). On a besoin d'utiliser le fait que la distance de Kullback entre deux lois $\mathbb{P}_\theta$ et $\mathbb{P}_{\theta'}$ du modèle est équivalente à la distance $\|\theta - \theta'\|^2$.

La loi de Poisson est une loi de type exponentiel, mais ne rentre pas dans ce cadre trop strict. Pour cette raison, nous présenterons au Chapitre 3 des résultats plus récents concernant les processus de Poisson qui légitimeront les techniques que nous serons amenés à considérer.

2.2.3 Illustration graphique de phénomènes de rupture

Dans cette section, nous illustrons sur un exemple (obtenu par simulation) comment se traduit une rupture dans une série. Le graphique suivant représente les valeurs d'une suite de loi de Poisson correspondant au nombre d'accidents par période de temps. Entre la période 19 et la période 20, une rupture se produit qui se traduit par une augmentation du nombre moyen de sinistres par unité de temps. La trajectoire est comparée avec une trajectoire typique qui aurait été obtenue s'il n'y avait pas eu de rupture.

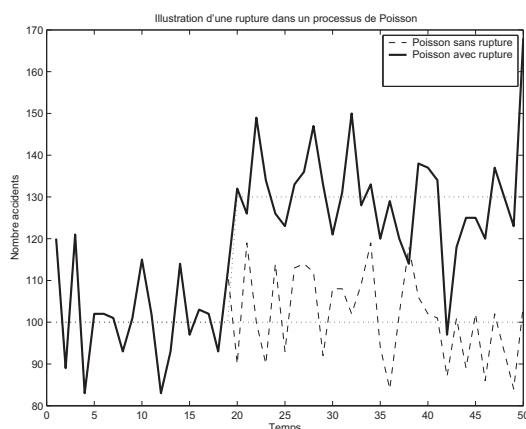


FIGURE 2.1 – Exemple de Processus de Poisson avec rupture.

On distingue le phénomène de rupture du phénomène de choc, représenté dans le graphique suivant. Dans le cas du choc, seule une observation est anormalement forte à la date 20, mais cette hausse ne perdure pas dans le temps.

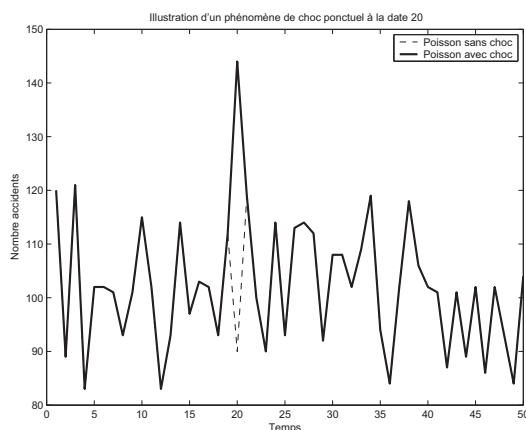


FIGURE 2.2 – Exemple de choc dans un processus de Poisson.

Sur un tel processus, une détection de rupture revient à effectuer une suite de tests statistiques permettant de conclure ou non à un changement dans la loi du processus.

Pour le processus de la figure 2.1, nous présentons de façon synthétique un extrait de la procédure de détection.

Date du test	Date détectée	Intensité avant	Intensité après	Log-vrais.	Seuil détection	Test
17	3	110	99.3	280	302	pas de rupt.
18	10	102.7	98.4	291	311	pas de rupt.
19	10	102.7	99.8	297	323	pas de rupt.
20	19	101.4	120	341	329	détection
21	19	101.4	132	724	334	détection
50	19	101.4	129.5	2250	773	détection

Le tableau se lit de la façon suivante : la date du test indique la date à laquelle on effectue la détection. Par exemple, lorsque la date du test est 17, on ne se base que sur l'information disponible à la date 17, c'est à dire les 17 premières observations. La colonne date détectée fournit l'estimateur de l'instant de rupture, i.e. l'instant où il est le plus vraisemblable qu'il y ait eu un changement dans le processus. Les colonnes intensité avant et intensité après fournissent les estimateurs de l'intensité du processus de Poisson avant et après rupture. La colonne log-vraisemblance fournit la valeur d'une statistique de test (qui sera définie dans les sections suivantes de façon plus précises) qui, si elle est supérieure au seuil de la colonne Seuil détection, amène à conclure à la présence d'une rupture à un certain niveau de confiance (ici 95%).

Ici, avant l'instant de la rupture (qui a lieu entre les observations 19 et 20), la procédure fournit un estimateur pour la date la plus vraisemblable de changement, mais la statistique de test reste inférieure au seuil critique, et on conclut à l'absence de rupture. Au fur et à mesure que le temps passe après la rupture, on conserve une estimation à la date 19, et la statistique de test franchit le seuil de détection de manière de plus en plus importante.

La même procédure, cette fois appliquée à la série de la figure 2.2 donne des résultats sensiblement différents. Elle détecte une rupture à la date 19 (en raison du choc), mais elle cesse rapidement de la détecter (à l'instant 27). La procédure intègre le fait que le changement n'était que momentané, et qu'il y a eu retour à la normale.

Chapitre 3

Etude du nombre de sinistres

Dans ce chapitre, nous étudions une modélisation possible du nombre de sinistres au cours du temps, puis nous présentons comment mettre en place un outil de détection de ruptures au sein de ce modèle. Nous modélisons le processus des sinistres par un processus de Poisson homogène par morceaux. La section 3.1 présente les difficultés posées par cette modélisation dans le cadre de nos données (parmi les difficultés, une forte présence de saisonnalité). Afin de résoudre ces problèmes, notre méthode consiste à découper l'année en périodes distinctes qui correspondent à différents types de sinistralité. Ce découpage est effectué en utilisant des techniques de détection de ruptures multiples, en utilisant un nouveau critère pénalisé qui permet d'éviter de spécifier a priori un nombre de périodes de temps. La description de cette technique est effectuée dans la section 3.2. Nous appliquons ces techniques à l'étude des trois garanties Vol, RC Corporels, et RC Matériels dans la section 3.3.

3.1 Objectifs et difficultés posées

Dans cette section, nous précisons succinctement les propriétés de l'outil de détection que nous voudrions mettre en place. Les difficultés posées par les données sont de deux types : manque d'information sur la stabilité du portefeuille et/ou des politiques tarifaires, et prise en compte d'une forte saisonnalité dans les observations.

3.1.1 Caractéristiques attendues de l'outil de détection

Nous souhaitons détecter des changements durables dans les risques que nous étudions, le but étant de mettre en lumière des éventuelles erreurs de modélisation préjudiciables à l'assureur. Les caractéristiques attendues sont donc les suivantes.

1. **Rapidité de détection** : nous souhaitons pouvoir détecter des changements en un nombre d'années le plus faible possible. L'ordre de grandeur de l'année provient d'une vision en terme de comptabilité annuelle, mais nous souhaitons si possible réduire ce délais de plusieurs mois pour que l'alerte puisse éventuellement être donnée en cours d'année.
2. **Fausses alarmes** : déclencher une "alarme" peut avoir des conséquences opérationnelles lourdes qui s'avéreraient contre-productives si cette alarme n'était pas associée à un changement significatif.
3. **Insensibilité à l'instabilité du portefeuille** : le nombre d'assurés peut augmenter d'une année sur l'autre et augmenter donc le nombre de sinistres, sans pour autant que le comportement des assurés ait subi de modification.

3.1.2 Stabilité du portefeuille

Les données qui nous ont été fournies ne permettent pas d'avoir une connaissance parfaite de la stabilité du portefeuille, puisque nous ne disposons pas du nombre d'assurés présents. Néanmoins, nous disposons du montant de primes acquises chaque année, qui donne une information partielle sur le nombre d'assurés (l'absence de données sur les changements de politiques tarifaires au cours du temps ne nous permet pas d'en déduire une connaissance parfaite du nombre d'assurés).

Nous choisissons donc de rapporter le nombre de sinistres au montant de primes de la façon suivante.

Hypothèse 1 : 1 euro acquis dans une année engendre un nombre aléatoire de sinistres N qui suit un processus de Poisson (non homogène) d'intensité $t \rightarrow \lambda(t)$.

Hypothèse 2 : On néglige les variations d'effectif au cours d'une année. On suppose l'effectif d'assurés de l'année en cours connu, à chaque instant de l'année.

Le nombre de sinistres rencontrés par l'assureur au cours d'une année est donc l'addition de π processus de Poisson de même intensité, où π est la quantité de primes acquises de l'année.

Hypothèse 3 : Les π processus de Poisson sont indépendants.

Cette troisième hypothèse n'est pas justifiée d'un point de vue micro-économique, mais devient justifiable dès lors que l'on s'intéresse à l'ensemble du portefeuille.

Précautions d'emploi : Notre hypothèse 1 peut être vue comme une hypothèse simplificatrice qui considère que l'évolution de la prime dépend uniquement de la fluctuation du nombre d'assurés. Si l'information disponible le permettait, il serait sans doute plus juste de raisonner en terme de sinistres engendrés par un assuré. A notre sens, le principal problème vient de l'impact d'un changement de tarif sur la méthode. Si, au cours d'une année n , le tarif a été augmenté pour faire face à une augmentation du nombre de sinistres, notre procédure devrait sans doute conclure à une stabilité du nombre de sinistres (le nouveau tarif répercute le changement de comportement des assurés et le compense, et aucune alerte n'est donnée). En revanche, si l'augmentation de tarif est due à une augmentation globale des coûts de sinistres mais que le nombre de sinistres par assuré reste globalement le même, notre outil déclenchera une alerte. Il convient donc d'être prudent dans l'interprétation des résultats : si l'on détecte une hausse de l'intensité λ , il convient d'analyser également les fluctuations tarifaires afin de comprendre d'où provient ce changement.

3.1.3 Présence de saisonnalité

La présence de saisonnalité dans le nombre de sinistres enregistrés au cours d'une année est une évidence. L'analyse spectrale (que nous ne présentons pas dans ce document pour ne pas l'alourdir) met clairement en évidence une saisonnalité d'ordre 7 si l'on raisonne en données journalières, et d'ordre 12 si l'on raisonne en données mensuelles.

Cette saisonnalité pose problème puisqu'elle polluerait considérablement notre procédure de détection : des changements de sinistralités risquent d'être systématiquement détectés aux changements de saisons. Afin de gommer cet effet saisonnier, on peut raisonner en données annuelles. Il en résulte néanmoins une importante perte d'information (au final, nous n'observons que 15 années), qui conduit fatalement à un retard de détection.

Nous avons donc décidé d'étudier avec plus de précision ces phénomènes saisonniers en essayant de découper l'année en plusieurs périodes correspondant à des profils similaires en terme de nombre de sinistres, et d'utiliser cette information pour mettre en place un outil de détection anticipée. Dans deux des garanties étudiées plus loin, nous verrons que cette approche nous permettra de détecter le phénomène avec un peu d'avance (6 mois environ).

Pour effectuer ce découpage de l'année, nous utilisons une technique présentée dans la section suivante.

3.2 Détection de ruptures multiples dans un processus de Poisson

Dans toute section nous désignons par $(N(t))_{t \in [0, T]}$ un processus de Poisson non homogène d'intensité $\lambda(t)$. Le processus de Poisson constitue la manière la plus simple de modéliser un nombre de sinistres au cours du temps. Il est notamment très utilisé en théorie de la ruine (voir notamment Assmussen, 2000). L'intensité $\lambda(t)$ est inconnue, et la seule observation disponible consiste en la trajectoire $t \rightarrow N(t)$ du processus.

Nous considérons la modélisation de λ comme une fonction constante par morceaux, sur des intervalles de temps de longueur inconnue. Le cas particulier où λ ne comporte que deux morceaux correspond au problème de détection de rupture classique. Dans le cas général (k morceaux), il s'agit d'estimer la meilleure décomposition de l'intervalle $[0, T]$

en k sous-intervalles sur lesquels N est homogène. Dans la section 3.2.1, nous supposons k connu et nous présentons la méthode d'estimation de λ ainsi que les résultats théoriques récents qui permettent de valider leur emploi. La section 3.2.2 explique comment choisir ce k à partir des données. Les sections 3.2.3 et 3.2.4 détaillent l'implémentation de la procédure, ainsi que son adaptation au cas où le processus n'est pas observé en temps réel mais simplement sur une grille discrète.

3.2.1 Estimation d'une rupture dans un processus de Poisson

Dans cette section, nous supposons que $\lambda(t)$ est de la forme suivante,

$$\lambda(t) = \lambda_0 + \sum_{i=1}^k \mu_{0,i} \mathbf{1}_{t_{0,i} < t \leq t_{0,i+1}}, \quad (3.1)$$

où $\mu_{0,i} > 0$ inconnu, $t_0 = (t_{0,i})_{1 \leq i \leq k}$ une suite croissante de $[0, T]$, avec $t_{0,k+1} = T$. On suppose λ_0 connu. Cette hypothèse n'est pas restrictive, puisque $t_{0,1}$ peut être égal à 0. La présence de λ_0 est due à un problème technique : la famille des lois de Poisson présente une singularité au voisinage de 0, et l'introduction de λ_0 permet de ne pas s'approcher trop près de 0. Ce problème est fréquent, et on peut trouver des restrictions similaires dans Haccou, Meelis, Van de Geer (1988) pour le cas de lois exponentielles.

On note $\mu = (\mu_i)_{1 \leq i \leq n}$. Dans ce contexte, la vraisemblance du modèle s'écrit de la façon suivante,

$$L(t, \mu) = N(t_1) \log \lambda_0 - \lambda_0 t_1 + \sum_{i=1}^k \{N(t_{i+1}) - N(t_i)\} \log(\mu_i + \lambda_0) - (\mu_i + \lambda_0)(t_{i+1} - t_i),$$

et on définit $(\hat{t}, \hat{\mu}) = \arg \max_{\mu > 0, t \in [0, T]} L(t, \mu)$. Notons

$$d(t, \mu, t_0, \mu_0) = -E[L(t, \mu) - L(t_0, \mu_0)].$$

Les vrais paramètres (t_0, μ_0) réalisent le maximum de $E[L(t, \mu)]$, donc $d(t, \mu, t_0, \mu_0)$ est positif pour tout (t, μ) . Le théorème suivant permet de contrôler $d(\hat{t}, \hat{\mu}, t_0, \mu_0)$, ce qui peut être vu comme une majoration d'une certaine distance entre $(\hat{t}, \hat{\mu})$.

La preuve du théorème suivant est due à Faugeras et Lopez (2010). Certains éléments de preuves sont retranscrits en annexe, l'intégralité pouvant être consultée dans l'article cité.

Théorème 2. *Si le modèle (3.1) est vérifié, alors, si on suppose que $\mu \in \mathcal{M}$ où $\mathcal{M}$ est un compact connu, sur l'ensemble $\{\hat{t} \leq \tau < T\}$, où τ est fixé,*

$$E \left[\sup_{(t, \mu)} \exp(1/2 \{L(t, \mu) - L(t_0, \mu_0) + d(t, \mu, t_0, \mu_0)\} - C_0 k \log \log T) \right] \leq C_1,$$

où C_0 et C_1 sont deux constantes positives ne dépendant pas de T ni de k .

Nous discutons et détaillons ce résultat au travers de quelques remarques.

Remarque 1 : Le résultat est non-asymptotique. Cette propriété est particulièrement importante dans notre contexte, puisque le temps d'observation est fixé, et éventuellement relativement faible (soit 15 unités de temps dans le cas de données annuelles, soit 52 dans le cas où l'on raisonne avec des données hebdomadaires sur une année).

Remarque 2 : Le résultat est énoncé dans sa forme la plus générale, car il permet d'obtenir à la fois des résultats sur la qualité d'estimation et sur des régions de confiance. En effet,

$$E \left[\sup_{(t, \mu)} \exp(1/2 \{L(t, \mu) - L(t_0, \mu_0) + d(t, \mu, t_0, \mu_0)\}) \right] \geq E [\exp(d(\hat{t}, \hat{\mu}, t_0, \mu_0))],$$

en utilisant le fait que $L(\hat{t}, \hat{\mu}) - L(t_0, \mu_0) > 0$, par définition du maximum de vraisemblance. En utilisant l'inégalité de Jensen, on en déduit donc

$$E [d(\hat{t}, \hat{\mu}, t_0, \mu_0)] \leq C'_1 k \log \log T. \quad (3.2)$$

Ceci montre une dégradation de la qualité d'estimation de l'ordre de $k \log \log T$. On peut montrer que le terme $\log \log T$ (généralement négligeable vu les ordres de grandeur considérés) disparaîtrait si les $(\mu_{0,i})_{1 \leq i \leq k}$ étaient exactement connus. Par ailleurs, en utilisant le fait que $d(\hat{t}, \hat{\mu}, t_0, \mu_0) > 0$ et l'inégalité de Chernoff, on déduit du Théorème 2 que

$$\mathbb{P} (L(\hat{t}, \hat{\mu}) - L(t_0, \mu_0) > z + C_0 k \log \log T) \leq C_1 \exp(-z/2). \quad (3.3)$$

On en déduit que les régions $\mathcal{R}(z) = \{(t', \mu') : L(\hat{t}, \hat{\mu}) - L(t', \mu') \leq z + C_0 k \log \log T\}$ peuvent servir de régions de confiance raisonnables, puisque la probabilité que (t_0, μ_0) n'appartiennent pas à $\mathcal{R}(z)$ décroît exponentiellement en fonction de z .

Remarque 3 : Dans le cas du modèle à une rupture, les régions $\mathcal{R}(z)$ peuvent servir à construire un test du rapport de vraisemblance pour l'hypothèse H_0 : absence de rupture $((t_0, \mu_0) = (n, 0))$ contre H_1 : il existe une rupture, on accepte H_0 si $(n, 0) \in \mathcal{R}(z)$ pour z suffisamment grand. La relation (3.3) montre que la probabilité des régions de rejet décroît exponentiellement en fonction de z , donc qu'on peut s'attendre à de bonnes performances de ce test en terme de puissance.

Remarque 4 : Le terme en $C'_1 k \log \log T$ dans (3.2) fait apparaître une dégradation en fonction de k et T . On peut montrer (voir Faugeras et Lopez, 2010, pour plus de détails) que ce terme est optimal à une constante près, au sens du risque minimax. Si, pour un estimateur $(\tilde{t}, \tilde{\mu})$ on définit $R(\tilde{t}, \tilde{\mu}) = \sup_{t_0, \mu_0} E_{t_0, \mu_0} [d(\tilde{t}, \tilde{\mu}, t_0, \mu_0)]$, où E_{t_0, μ_0} désigne l'espérance lorsque les vrais paramètres sont (t_0, μ_0) , alors, pour tout estimateur, $R(\tilde{t}, \tilde{\mu}) \geq C''_1 k \log \log T$. Cette remarque prend tout son sens dans l'optique de la section 3.2.2 où k est inconnu : le terme $k \log \log T$ est à la base d'un critère pénalisé permettant de choisir k à partir des données. Le résultat concernant le risque minimax nous assure une certaine forme d'optimalité pour notre critère.

Remarque 5 : On peut expliciter la fonction d . Nous présentons le cas $k = 1$, le cas général s'étudiant de la même manière. On a, lorsque $t > t_0$

$$d(t, \mu, t_0, \mu_0) = (t - t_0) \left[\mu_0 - (\lambda_0 + \mu_0) \log \left(\frac{\lambda_0 + \mu_0}{\lambda_0} \right) \right] + (T - t) \left[(\mu - \mu_0) - \log \left(\frac{\lambda_0 + \mu}{\lambda_0 + \mu_0} \right) \right].$$

Lorsque $t < t_0$,

$$d(t, \mu, t_0, \mu_0) = (t_0 - t) \left[\mu - \lambda_0 \log \left(\frac{\lambda_0 + \mu}{\lambda_0} \right) \right] + (T - t_0) \left[(\mu - \mu_0) - \log \left(\frac{\lambda_0 + \mu}{\lambda_0 + \mu_0} \right) \right].$$

On remarque que cette distance d se décompose en deux parties, une partie qui correspond à une différence entre t , date estimée, et t_0 date réelle de rupture, et une seconde partie qui ne concerne que la différence entre l'amplitude de rupture estimée, et la vraie amplitude μ_0 . Si on se focalise sur le terme qui concerne la date de rupture, on remarque que

$$E [|\hat{t} - t_0| \mathbf{1}_{\hat{t} > t_0}] \leq \frac{C'_1 \log \log T}{\left[\mu_0 - (\lambda_0 + \mu_0) \log \left(\frac{\lambda_0 + \mu_0}{\lambda_0} \right) \right]},$$

où le dénominateur du membre de droite est une fonction croissante de l'amplitude de rupture μ_0 . On retrouve le même phénomène que dans le Théorème 1 pour la rupture gaussienne, plus la rupture est d'amplitude forte, mieux elle est détectée. La façon dont joue cette amplitude μ_0 est néanmoins différente du cas gaussien.

Remarque 6 : L'énoncé reste valable si le modèle (3.1) est mal spécifié. Le paramètre (t_0, μ_0) n'est alors plus lié à la vraie fonction $\lambda(t)$, mais correspond à une fonction $\lambda^*(t)$ constante par morceaux qui peut être vue comme la meilleure approximation de λ par une fonction constante par morceaux.

Remarque 7 : Le fait d'imposer $\hat{t} \leq \tau < T$ est une limitation technique de faible importance. En particulier elle disparaît lorsque le problème est discrétisé.

3.2.2 Sélection du nombre de ruptures

En pratique, le nombre de rupture est inconnu. Le problème d'identifier le nombre de rupture peut être vu comme un problème de classification non supervisée. Du point de vue de l'application que nous avons à l'esprit, nous cherchons à identifier le nombre de périodes nécessaire pour découper l'année en périodes similaires de façon satisfaisante. Définissons par $L_k(t^{(k)}, \mu^{(k)})$ la vraisemblance dans le modèle à k ruptures. Même si les modèles sont imbriqués, l'estimateur du maximum de vraisemblance choisit toujours un modèle à k ruptures, donc on risque de surestimer leur nombre.

Le Théorème 2 suggère d'utiliser un critère de vraisemblance pénalisée de la forme suivante :

$$(\hat{t}, \hat{\mu}, \hat{k}) = \arg \max L_k(t^{(k)}, \mu^{(k)}) - \text{pen}(k),$$

où le terme de pénalité est de la forme

$$\text{pen}(k) = C_0 k \log \log T.$$

Ainsi, on introduit un terme de pénalité qui défavorise les modèles présentant beaucoup de ruptures. On peut montrer (voir Faugeras et Lopez, 2010) que l'estimateur obtenu vérifie

$$E [d(\hat{t}, \hat{\mu}, t_0, \mu_0)] \leq \tilde{C}_1 k_0 \log \log T,$$

où k_0 est le nombre optimal de ruptures.

Choix de la constante. Le choix de la constante C_0 est bien entendu crucial. Si C_0 est trop petite, le modèle sélectionné sera trop complexe (trop de ruptures, donc mauvaise précision), et si C_0 est très grand, on sélectionnera des modèles pas suffisamment complexes pour bien décrire les données. En suivant la preuve du Théorème 2, on peut obtenir une valeur explicite pour la constante C_0 , mais cette valeur apparaît en pratique trop grande. Pour sélectionner un C_0 adapté, nous avons donc procédé en deux phases :

Méthode de choix de C_0 .

1. Utilisation de l'heuristique de pente (Birgé, Massart, 2006).
2. Amélioration par simulation.

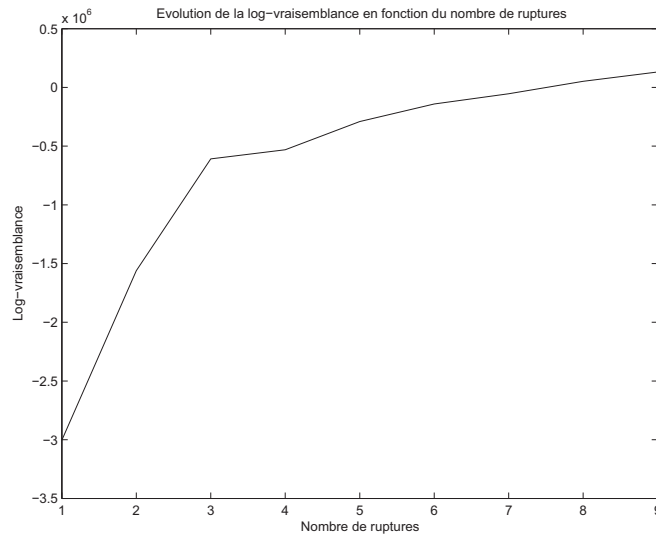


FIGURE 3.1 – Illustration de la croissance de la log-vraisemblance, Garantie Vol, année 2000.

L'heuristique de pente (qui n'est pas validée d'un point de vue théorique) consiste à observer la croissance de la vraisemblance en fonction de k . Sur la figure 3.2.2, on a tracé $L_k(\hat{t}^{(k)}, \hat{\mu}^{(k)})$ en fonction de k . L'estimation a été effectuée sur la garantie Vol pour l'année 2000. On observe une croissance de la vraisemblance qui se ralentit à partir de $k = 4$ ou $k = 5$, et qui devient quasiment linéaire. Puisque la forme attendue de la pénalité est $k \log \log T$ et que $\log \log T$ est fixé (on raisonne sur une année fixée), la pente de la droite doit nous informer sur la valeur de C_0 . Birgé et Massart proposent d'estimer la pente p à partir du moment où la vraisemblance devient quasiment linéaire (ici $k = 4$ ou $k = 5$), et de retenir $C_0 = 2p$.

Cette première méthode permet d'obtenir un ordre de grandeur de C_0 . Nous recherchons ensuite à améliorer cette constante en utilisant des simulations, de la façon suivante.

Obtention de C_0 par simulations

1. On se donne un vecteur C de valeurs proches de C_0 .
2. Utiliser le C_0 de l'heuristique de pente pour obtenir un modèle $(\hat{t}, \hat{\mu}, \hat{k})$.
3. Simuler n trajectoires d'un processus de Poisson associé au modèle $(\hat{t}, \hat{\mu}, \hat{k})$.
4. Pour chaque élément c du vecteur C , et chaque trajectoire, utiliser le critère pénalisé basé sur la constante c .
5. Choisir la constante c qui fournit le meilleur résultat (qui détecte le plus grand nombre de fois le bon modèle).

3.2.3 Implémentation

Dans la définition de l'estimateur du maximum de vraisemblance, le maximum est pris sur un ensemble infini, et peut donc sembler difficile à calculer. En réalité, on peut remarquer que, pour t fixé, le maximum de $L(t, \mu)$ est atteint pour la suite

$$\mu_i(t) = \frac{N(t_{i+1}) - N(t_i)}{t_{i+1} - t_i} - \lambda_0.$$

L'optimisation du critère revient donc uniquement à l'optimisation de $L(t, \mu(t))$ par rapport à t .

En réalité, cette optimisation se ramène à l'étude de cette vraisemblance en un nombre fini de points (les points où se produisent des sauts du processus N). Pour plus de simplicité, nous expliquons ce phénomène uniquement dans le cas où $k = 1$, le cas général s'étudiant avec des arguments similaires. En effet, soit $t' > t$ tel que aucun saut du processus N ne se produise entre t et t' , et soit μ fixé. L'estimateur du maximum de vraisemblance maximise

$$L(t, \mu) - L(0, n) = \{N(T) - N(t)\} \log \left(\frac{\lambda_0 + \mu}{\lambda_0} \right) - \mu(T - t), \quad (3.4)$$

où l'on a retranché $L(0, n)$ pour plus de lisibilité. Si aucun saut ne se produit entre t' et t , alors $N(T) - N(t') = N(T) - N(t)$, et le premier terme de (3.4) ne change pas quand on passe de t à t' . En revanche, le second terme diminue puisque $\mu > 0$ et $T - t > T - t'$. On en déduit donc que le maximum de $L(t, \mu)$ ne peut être atteint que pour des t_i correspondant à des instants de sauts du processus.

Cette conclusion a peu d'impact dans notre cadre, puisque nous travaillons dans une version discrétisée du problème. Mais elle est peut être utile dans le cas où le processus de Poisson est observé complètement.

3.2.4 Discrétisation du problème

Les données que nous considérons sont des données journalières, mensuelles, ou annuelles. Par ailleurs, le nombre de sinistres enregistré est toujours supérieur à 1 pour chaque jour de l'année. Nous ne pouvons donc pas considérer que l'observation du processus est effectuée en temps continu. On dispose donc d'observations $(N_i)_{1 \leq i \leq n}$ définies

comme $N_i = N(i) - N(i - 1)$, où chaque N_i suit une loi de Poisson de paramètre $\lambda(i)$. La log-vraisemblance reste la même, en remplaçant $N(t)$ par $\sum_{i=1}^n N_i \mathbf{1}_{i \leq t}$.

Le résultat du Théorème 2 reste valide (puisqu'il suffit de prendre le supremum sur i et non sur l'ensemble $[0, T]$), en remplaçant T par n , nombre de périodes de temps observées.

3.3 Application aux données

Nous présentons ici les résultats obtenus sur les garanties Vol, RC Corporels, et RC Matériels. Dans un premier temps, nous montrons comment la détection de ruptures multiples nous permet de décomposer l'année en période de sinistralité comparable (section 3.3.1). Dans un second temps, nous mettons en place l'outil de détection de rupture, d'abord dans une vision purement annuelle, puis en utilisant l'information supplémentaire apportée par des observations intermédiaires effectuées en cours d'année (section 3.3.4).

3.3.1 Décomposition de l'année

Nous nous basons sur des données hebdomadaires, N_i désigne le nombre de sinistres en semaine i , avec la convention suivante : si le 1er janvier est un lundi, les semaines commencent un lundi, si c'est un mardi, les semaines commencent un mardi etc. Il n'est pas possible de raisonner en journalier, puisque cette approche introduirait une saisonnalité importante d'ordre 7, qui est gommée par la vision journalière. Nous avons donc $n = 52$. Nous ne tenons pas compte des derniers jours de l'année (1 à 2 jours suivant que l'année est bissextile ou non), qui apportent peu d'information.

Nous considérons des modèles de ruptures pouvant aller jusqu'à $k = 10$. Nous n'envisageons pas des modèles plus complexes pour deux raisons : la première tient au temps de calcul qui augmente fortement avec k , la seconde tient au fait qu'au-delà de 10, on s'attend à trouver des décompositions en période de moins d'un mois. Ce découpage semble trop fin, d'abord parce qu'il existe en pratique un délai entre la survenance du sinistre et sa connaissance par l'assureur : de ce fait, on pourrait avoir plusieurs périodes de sinistralité de retard lors de l'arrivée des données. L'autre point concerne le nombre de paramètres à estimer, qui serait très important et donc sujet à caution.

Les profils que nous avons dégagés sur les différents risques sont assez similaires : la courbe de vraisemblance en fonction de k qui est utilisée pour l'heuristique de pente et la calibration de la pénalité, on retrouve un profil similaire à celui de la figure 3.2.2 présentée dans la section 3.2.2, c'est à dire un tassement de la croissance autour de $k = 4$ ou $k = 5$. D'autres graphiques sont présentés en Annexe C. Nous avons donc construit deux pénalités : l'une se basant sur la pente estimée entre $k = 4$ et $k = 10$, la seconde entre $k = 5$ et $k = 10$. A quelques rares exceptions, la première pente conduit à une constante c_1 plus grande que la constante c_2 obtenue sur la deuxième pente. La constante c_1 correspond donc à une pénalisation plus forte, mais l'usage des deux constantes donne des résultats souvent similaires.

Nous présentons ci-dessous, à titre d'illustration, le résultat d'une détection de ruptures multiples sur la garantie Vol et l'année 1996.

Nb ruptures	Dates de ruptures	Log-vraisemblance	Pen. forte (c_1)		Pen. faible (c_2)	
			valeur	choix	valeur	choix
1	44	-3.286e+6	-3.469e+6	non	-3.468e+6	non
2	3 30	-1.676e+6	-2.043e+6	non	-2.040e+6	non
3	9 27 44	-7.386e+5	-1.289e+6	oui	-1.285e+6	oui
4	9 27 30 38	-5.737e+5	-1.308e+6	non	-1.302e+6	non
5	9 27 30 32 44	-4.684e+5	-1.386e+6	non	-1.379e+6	non
6	9 27 31 33 42 44	-3.719e+5	-1.473e+6	non	-1.465e+6	non
7	3 8 9 25 27 30 38	-3.365e+5	-1.622e+6	non	-1.611e+6	non
8	3 8 9 22 27 28 30 38	-2.029e+5	-1.672e+6	non	-1.660e+6	non
9	3 8 9 22 27 28 30 42 44	-9.756e+4	-1.750e+6	non	-1.737e+6	non

Les valeurs de c_1 et c_2 sont :

$$\begin{aligned} c_1 &= 91791 \\ c_2 &= 91067. \end{aligned}$$

Le modèle sélectionné par la pénalité forte comme la pénalité faible est celui qui comporte $\hat{k} = 3$ ruptures, ce qui correspond à un découpage de l'année en 4 périodes. Les dates de ruptures estimées sont donc au niveau des semaines 9, 27 et 44. Pour les autres années et les autres garanties, on retrouve généralement un résultat similaire : c'est généralement le modèle à 3 ruptures qui est sélectionné, plus rarement le modèle à 4 ruptures, et exceptionnellement le modèle à 5 ruptures. Les autres modèles ne sont jamais sélectionnés. Dans le tableau suivant, nous résumons le nombre de cas où le modèle à k ruptures est choisi en séparant chaque garantie, pour $k = 3$ à 5.

Garantie	$k = 3$		$k = 4$		$k = 5$	
	Pen. forte	Pen. faible	Pen. forte	Pen. faible	Pen. forte	Pen. faible
Vol	13	10	2	3	0	2
Rccorp	10	7	2	3	3	5
Rcmat	11	11	2	1	2	3

L'ensemble des résultats est détaillé dans l'annexe C. Nous retenons donc un modèle à 3 ruptures pour chaque année. On remarque également que les instants de ruptures estimés sont toujours situés à peu près au même endroit d'une année sur l'autre, ce qui dénote une certaine stabilité du découpage (voir l'annexe C pour les résultats détaillés). Nous avons essayé d'interpréter les fluctuations dans ces résultats quand cela nous a semblé possible. Pour les garanties Vol et Rccorp, nous avons notamment remarqué que ces fluctuations suivaient à peu près celles des dates de début de vacances scolaires. Quand nous n'avons pas pu interpréter cette fluctuation, nous avons simplement effectué une moyenne pour estimer la date de rupture. Les résultats définitifs sont présentés dans le tableau suivant.

Garantie	Période 1	Période 2
Vol	01/01 à fin vacances février	fin vacances février à 30/06
Rccorp	01/01 à début vacances Pâques	début vacances Pâques au 30/06
Rcmat	semaines 1 à 10	semaine 11 au 30/06
Garantie	Période 3	Période 4
Vol	01/07 au 31/10	du 01/11 au 31/12
Rccorp	01/07 au 25/10	26/10 au 31/12
Rcmat	01/07 à 30/09	01/10 à 31/12

Pour chaque garantie, l'année est donc découpée en 4 périodes sur lesquelles le nombre de sinistres est à peu près homogène. Pour une garantie donnée g , une année i , et une période j (j compris entre 1 et 4), on note $\lambda_{i,j}(g)$ l'intensité du processus de Poisson sur cette période. Nous effectuons à présent une hypothèse de stabilité d'une année sur l'autre.

Hypothèse de stabilité : Pour toute période $j > 1$, et pour toute année i , $\lambda_{i,j}(g) = \alpha_{i,g} \lambda_{i,1}(g)$, où le paramètre $\alpha_{i,g}$ ne varie pas d'une année sur l'autre.

La période 1 sert donc de période de référence (donc $\alpha_{1,g} = 1$), le paramètre $\alpha_{i,g}$ permettant de comparer une période donnée avec la période 1. Cette hypothèse de stabilité est contestable, mais il s'agit de l'hypothèse la plus simple qui nous permette de travailler par la suite. Nous présentons dans les trois tableaux suivants les résultats de l'estimation des différents $\alpha_{i,g}$. Le tableau récapitule, année par année, les valeurs de $\lambda_{i,j}(g)/\hat{\lambda}_{i,1}$, où $\hat{\lambda}_{i,j}(g)$ est le nombre moyen de sinistres par semaines sur la période i , l'année j , pour la garantie g . La dernière ligne de chaque tableau présente les estimateurs des $\alpha_{i,g}$ obtenus en moyennant chaque colonne.

Garantie Vol.

Année	$\alpha_{2,g}$	$\alpha_{3,g}$	$\alpha_{4,g}$
1994	0.9894	0.9190	0.9569
1995	0.9582	0.8945	0.8932
1996	0.9914	0.9465	0.9411
1997	1.0203	0.9164	0.9885
1998	0.9772	0.9593	0.9596
1999	1.0067	0.9376	0.9261
2000	1.0288	1.0101	1.0188
2001	1.0143	0.9744	0.9910
2002	0.9560	0.8474	0.8466
2003	1.0468	0.9763	0.9618
2004	0.9857	0.9099	0.8949
2005	1.0792	1.0136	0.9327
2006	1.0606	0.9666	0.9368
2007	0.9901	0.8698	0.8255
2008	1.0153	0.9376	0.9550
Valeur estimée	1.0080	0.9386	0.9352

Garantie Rccorp.

Année	$\alpha_{2,g}$	$\alpha_{3,g}$	$\alpha_{4,g}$
1994	1.0150	1.0679	1.1696
1995	1.0191	1.0347	1.2105
1996	1.0489	1.1384	1.2314
1997	1.1922	1.2165	1.3499
1998	1.1097	1.1684	1.2366
1999	1.1037	1.1274	1.1955
2000	0.9979	1.0037	1.1078
2001	0,9861	1.0080	1.0396
2002	0.9973	1.0539	1.0318
2003	0.9255	1.2553	0.7021
2004	1.1110	1.1425	1.1538
2005	1.1251	1.0969	1.1426
2006	1.0904	1.0925	1.1819
2007	1.0532	1.0696	1.1037
2008	1.022	1.0478	1.0575
Valeur estimée	1.0531	1.1016	1.1276

Garantie Rcmat.

Année	$\alpha_{2,g}$	$\alpha_{3,g}$	$\alpha_{4,g}$
1994	0.9618	0.9272	1.0617
1995	0.9270	0.8568	1.0602
1996	0.8963	0.8993	1.1123
1997	1.0480	0.9818	1.1856
1998	1.030	1.0079	1.1771
1999	0.9910	0.9638	1.1384
2000	0.9653	0.9180	1.0654
2001	0.9848	0.9357	1.0703
2002	0.9982	0.9520	1.0138
2003	1.0128	0.9719	1.0441
2004	1.0210	0.9762	1.0700
2005	0.9914	0.9204	1.0340
2006	0.9823	0.9540	1.0640
2007	1.0546	1.0022	1.1358
2008	1.0350	1.0010	1.0956
Valeur estimée	0.9933	0.9512	1.0885

On note une certaine stabilité d'une année sur l'autre, ce qui semble montrer que le découpage effectué en 4 périodes n'est pas aberrant. Par exemple, pour la garantie Rccorp, on remarque que c'est presque toujours la période 4 qui correspond à la plus grande proportion journalière de sinistres. Par ailleurs, nous avons effectué un test de Kolmogorov-Smirnov pour tester l'adéquation à une loi de Poisson sur les différentes périodes (en regroupant les observations sous forme hebdomadaire pour éviter la saisonnalité), et l'hypothèse de loi de Poisson n'est pas rejetée. Il faut néanmoins nuancer ce résultat, car la puissance du test de Kolmogorov-Smirnov est relativement faible pour le nombre d'observations que nous considérons (environ une quinzaine de semaines observées par période).

3.3.2 Détection de rupture

Nous présentons à présent les résultats liés à la détection de rupture proprement dite. Nous considérons tout d'abord des données annuelles. On rappelle la modélisation du nombre annuel de sinistres. On suppose que N_i (nombre de sinistres l'année i) suit une loi de Poisson de paramètre $\pi_i \lambda_i$, où π_i est la quantité de primes acquises l'année i (supposée connue), et λ_i suit un modèle de rupture, c'est à dire

$$\lambda_i = \lambda(\mu_1, \mu_2, i_0) = \lambda_0 + \mu_1 + (\mu_2 - \mu_1) \mathbf{1}_{i > i_0}, \quad (3.5)$$

les inconnues étant (μ_1, μ_2) positifs, et i_0 . Comme le nombre d'observations est faible en vision annuelle (16 observations), nous considérons ici un modèle où l'intensité avant rupture n'est pas connue (nous n'avons pas assez de recul pour l'estimer suffisamment fidèlement pour la considérer connue). On peut donc voir le modèle (3.5) comme un modèle à deux ruptures, la première rupture se produisant à la date 0.

La vraisemblance est modifiée par l'introduction des primes π_i de la façon suivante :

$$L(\mu_1, \mu_2, i) = \sum_{j=1}^n N_j \log(\pi_j \lambda(\mu_1, \mu_2, i_0)) - \pi_j \lambda(\mu_1, \mu_2, i_0).$$

A i fixé, les valeurs de μ_1 et μ_2 qui maximisent L sont respectivement

$$\begin{aligned} \mu_1(i) &= \frac{\sum_{j=1}^i N_j}{\sum_{j=1}^i \pi_j} - \lambda_0, \\ \mu_2(i) &= \frac{\sum_{j=i+1}^n N_j}{\sum_{j=i+1}^n \pi_j} - \lambda_0. \end{aligned}$$

On effectue le test de détection de rupture de la façon suivante. On note $L_n = \max_{\mu_1, \mu_2, i} L(\mu_1, \mu_2, i)$. Dans un cadre de détection statique de rupture, on teste

$$H_0 : i_0 = n,$$

contre

$$H_1 : i_0 < n.$$

Le test du rapport de vraisemblance consiste alors à rejeter H_0 dès lors que $L(\mu_1, \mu_2, i) > s_n(\alpha)$, où α est le niveau du test. Pour calibrer ce seuil de rejet, on utilise une procédure de simulation décrite ci-dessous.

Evaluation de $s_n(\alpha)$ par simulations.

1. Les $(\pi_j)_{1 \leq j \leq n}$ sont fixés.
2. On calcule $\mu = \mu_1(n)$. μ est l'estimation du maximum de vraisemblance de la μ_1 si l'on suppose qu'il n'y a pas de rupture.
3. On simule m vecteurs de tailles n , $(N_i^{(k)})_{1 \leq i \leq n}$, $k = 1, \dots, m$, où N_i suit une loi de Poisson de paramètre $\pi_i(\lambda_0 + \mu)$.
4. Pour chaque trajectoire, on calcule le maximum de vraisemblance $L_n^{(k)}$, puis on trie le vecteur des $L_n^{(k)}$.
5. On prend $s_n(\alpha)$ tel que seule une proportion α des $L_n^{(k)}$ soit supérieure à $s_n(\alpha)$.

Ici, nous réalisons une détection dynamique, puisque nous réeffectuons un test chaque année. Si l'on ne veut pas avoir trop de fausses alertes, il faut donc modifier sensiblement le seuil de rejet. Néanmoins, dans le cas présent, le nombre d'observations est suffisamment faible pour que cette modification ait peu d'impact. Nous renvoyons le lecteur à la section 4.2.2 du chapitre 4 pour avoir une idée de la façon de passer du test statique au test dynamique, où une technique rigoureusement similaire est présentée dans le cadre de l'étude de données mensuelles de coûts de sinistres.

Nous présentons dans le tableau ci-dessous les résultats de la procédure de détection, garantie par garantie.

Résultat de la détection

Garantie	Année	Date détectée	Intensité avant	Intensité après	Log-vrais.	Seuil détection	Test
Vol	1997	1994	7.163e-4	5.301e-4	4.619e+6	4.631e+6	H_0
	1998	1994	7.163e-4	5.212e-4	5.682e+6	5.695e+6	H_0
	1999	1994	7.163e-4	5.118e-4	6.717e+6	6.730e+6	H_0
	2000	1994	7.163e-4	5.013e-4	7.740e+6	7.756e+6	H_0
	2001	1995	6.423e-4	4.822e-4	8.816e+6	8.833e+6	H_0
	2002	1995	6.423e-4	4.712e-4	9.823e+6	9.842e+6	H_0
	2003	1995	6.423e-4	4.545e-4	1.070e+7	1.073e+7	H_0
	2004	2001	5.175e-4	3.675e-4	1.148e+7	1.151e+7	H_0
	2005	2001	5.175e-4	3.507e-4	1.217e+7	1.219e+7	H_0
	2006	2002	5.042e-4	3.166e-4	1.281e+7	1.279e+7	H₁
Rccorp	2007	2002	5.042e-4	3.041e-4	1.331e+7	1.330e+7	H₁
	2008	2002	5.042e-4	2.899e-4	1.375e+7	1.374e+7	H₁
	1997	1995	5.6062e-5	5.190e-5	1.3443e+6	1.3497e+6	H_0
	1998	1995	5.606e-5	5.255e-5	1.702e+6	1.708e+6	H_0
	1999	1995	5.6062e-5	5.299e-5	2.072e+6	2.078e+6	H_0
	2000	1999	5.346e-5	4.995e-5	2.4220e+6	2.429e+6	H_0
	2001	1999	5.346e-5	4.821e-5	2.750e+6	2.758e+6	H_0
	2002	2000	5.292e-5	4.371e-5	3.045e+6	3.052e+6	H_0
	2003	2001	5.206e-5	3.795e-5	3.297e+6	3.304e+6	H_0
	2004	2001	5.206e-5	3.696e-5	3.548e+6	3.555e+6	H_0
Rcmat	2005	2001	5.206e-5	3.653e-5	3.795e+6	3.803e+6	H_0
	2006	2001	5.206e-5	3.621e-5	4.029e+6	4.036e+6	H_0
	2007	2001	5.206e-5	3.594e-5	4.483e+6	4.245e+6	H₁
	2008	2001	5.206e-5	3.583e-5	4.696e+6	4.488e+6	H₁
	1997	1995	2.535e-4	3.036e-4	8.365e+6	8.373e+6	H_0
	1998	1995	2.535e-4	3.097e-4	1.087e+7	1.088e+7	H_0
	1999	1995	2.535e-4	3.125e-4	1.344e+7	1.345e+7	H_0
	2000	1995	2.535e-4	3.095e-4	1.592e+7	1.593e+7	H_0
	2001	1995	2.535e-4	3.057e-4	1.834e+7	1.836e+7	H_0
	2002	1995	2.535e-4	2.990e-4	2.060e+7	2.062e+7	H_0
	2003	2001	2.936e-4	2.514e-4	2.270e+7	2.272e+7	H_0
	2004	2001	2.936e-4	2.480e-4	2.479e+7	2.482e+7	H_0
	2005	2001	2.936e-4	2.474e-4	2.689e+7	2.691e+7	H_0
	2006	2001	2.936e-4	2.480e-4	2.892e+7	2.895e+7	H_0
	2007	2001	2.936e-4	2.505e-4	3.100e+7	3.103e+7	H_0
	2008	2001	2.936e-4	2.540e-4	3.318e+7	3.321e+7	H_0

Pour chaque garantie, nous n'avons commencé la détection qu'à partir de l'année 1997 afin d'avoir suffisamment de recul. Dans le tableau, pour chaque ligne, on se place à la fin

de l'année x (colonne Année), et on présente le résultat de la détection de rupture basée sur les observations de 1994 jusqu'à l'année x .

La colonne "Date détectée" fournit l'estimation de la date de rupture.

La colonne "Intensité avant" fournit l'estimation de $\lambda_0 + \mu_1$ (intensité avant rupture).

La colonne "Intensité après" fournit l'estimation de $\lambda_0 + \mu_2$. La colonne "Log-vraisemblance" fournit la valeur du maximum de vraisemblance, qui doit être comparée au seuil $s_n(0.05)$ qui est reporté dans l'avant dernière colonne.

La dernière colonne donne le résultat du test de détection (H_0 si pas de rupture, H_1 sinon). On conclut à une détection au niveau 5% si la log-vraisemblance est supérieure au seuil, ce qui ne se produit que pour les garanties Vol et Rccorp. Le délai de détection est de quatre ans pour le vol, six ans pour la Rccorp. Ce seuil de 5% est arbitraire. Les seuils de test "raisonnables" (jusqu'à 15%) donnent des résultats similaires.

Pour la garantie Vol, on identifie une rupture à la date 2002, qui se caractérise par une diminution moyenne de sinistres engendrés par un euro de prime (passage de $5e - 4$ à $2.9e - 4$). On remarque également une diminution pour la garantie Rccorp. La date détectée étant 2002 pour la garantie Vol, on peut examiner le comportement de la procédure de détection dynamique juste après cette rupture supposée. Au début de la procédure (1997), la date de rupture détectée est 1994, et reste dans ce voisinage (1995 à partir de 2001). Puis on a un saut de la date détectée peu après 2002, avec une stabilisation de cette date détectée aux alentours de 2001-2002. Un phénomène similaire se produit pour la garantie Rccorp. Ce type de comportement aurait peut-être pu alerter sur la présence d'une rupture, et de ce fait réduire le délai de détection. On remarque qu'aucune rupture n'est détectée par le test pour la garantie RC Matériels, malgré une stabilisation de la date détectée vers 2001, ce qui laisse supposer une rupture. Un regard plus attentif sur nos données montre qu'il y a diminution des sinistres à partir de 2001, mais que cette diminution est peu à peu compensée par une diminution des primes à partir de 2004, ce qui empêche la détection.

Suite à cette détection, on en déduit la prévision suivante. La colonne λ_i du tableau ci-dessous présente la valeur que l'on estime pour le λ_i de l'année indiquée dans la première colonne. Ce λ_i est estimé à partir des informations jusqu'au 31 décembre de l'année précédente. Pour prédire le nombre de sinistres, on a supposé connu la quantité de primes qui sera acquise dans l'année étudiée (en pratique, cette quantité ne peut être que prédite). Ce nombre de sinistres prédit est comparé au nombre de sinistres effectivement réalisés.

	Vol			Rccorp			Rcmat		
Année	λ_i	Prédit	Réalisé	λ_i	Prédit	Réalisé	λ_i	Prédit	Réalisé
1998	5.705e-4	116 380	101 023	5.295e-5	36 463	37 543	2.789e-4	192 075	221 615
1999	5.546e-4	115 005	98 581	5.327e-5	37 675	38 692	2.875e-4	203 359	226 964
2000	5.406e-4	116 658	97 589	5.351e-5	39 406	36 782	2.931e-4	215 838	219 369
2001	5.269e-4	117 326	102 191	5.300e-5	39 620	34 764	2.938e-4	219 611	215 228
2002	5.175e-4	120 792	96 208	5.219e-5	40 073	31 478	2.931e-4	225 020	201 337
2003	5.042e-4	122 862	85 752	5.094e-5	39 889	27 380	2.896e-4	226 767	188 632
2004	4.865e-4	109 804	76 326	4.935e-5	38 466	27 267	2.847e-4	221 962	188 054
2005	4.720e-4	107 053	67 779	4.804e-5	36 700	26 912	2.808e-4	214 507	187 593
2006	4.565e-4	100 613	60 374	4.697e-5	34 412	25 534	2.778e-4	203 559	183 403
2007	3.166e-4	67 621	53 455	4.604e-5	32 540	25 525	2.757e-4	194 881	186 812
2008	3.041e-4	65 088	45 996	3.594e-5	25 378	24 225	2.749e-4	194 127	195 050

Pour les deux garanties pour lesquelles une rupture a été identifiée, il est intéressant de comparer ces prévisions à celles qui auraient été obtenues si l'on n'utilise pas la méthode de détection de rupture, et qu'on se base sur tout l'historique depuis 1994 pour prédire le nombre de sinistres. Seules les années après rupture diffèrent dans ces deux méthodes.

	Vol			Rccorp		
Année	λ_i	Prédit	Réalisé	λ_i	Prédit	Réalisé
2007	4.420e-4	94 390	53 455	4.604e-5	32 540	25 525
2008	4.282e-4	91 657	45 996	4.553e-5	32 009	24 225

L'utilisation de la méthode de détection de rupture donne des résultats plus satisfaisants. Si l'on utilise l'ensemble des données historiques pour prédire le nombre de sinistres l'année suivante, on obtient une surestimation forte du nombre de sinistres dans le cas de la garantie Vol (par exemple, pour 2008, la méthode avec détection de rupture prédit 65 088 sinistres, la méthode sans détection prédit 91 657 sinistres, soit près du double des 45 996 qui seront effectivement enregistrés). Le même type de comportement est observé sur la garantie RC corporels, avec un écart certes moins important.

Pour la garantie Vol, l'écart sur le nombre de sinistres prévus est important sur les années 2002-2006 (c'est à dire après la rupture enregistrée). Néanmoins, comme on l'a mentionné, on pourrait vraisemblablement anticiper la détection soit en jouant sur le niveau du test, soit en tenant compte du fait que l'estimateur de la date de rupture se stabilise rapidement autour de 2002. De ce fait, la performance pourrait sans doute être améliorée. On peut cependant remarquer que la méthode de détection de rupture utilisée est prudente : l'évolution du nombre de sinistres est à la baisse, mais la méthode répercute progressivement cette baisse. Le réel saut n'a lieu que lorsque le test de détection conclut à une rupture significative.

Concernant la garantie RC matériels, les prédictions sont en-dessous de ce qui est effectivement réalisé pour les premières années (c'est également le cas pour l'année 2008, mais

les deux chiffres sont du même ordre de grandeur). Il ne faut pas nécessairement en tirer de conclusions négatives : pour prédire le nombre de sinistres des premières années, nous utilisons vraisemblablement moins d'information que ce que nous aurions pu effectivement utiliser à cette date (nos données ne remontent qu'à 1994).

3.3.3 Etude par simulations

Dans cette section, nous essayons d'évaluer par simulations le comportement des procédures de détection de rupture. Nous nous concentrons sur deux aspects : la puissance du test, c'est à dire sa capacité à détecter une rupture quand celle-ci est présente, et la qualité d'estimation.

Pour évaluer la puissance par simulations, nous nous fixons une taille d'échantillon de 15 pour correspondre au nombre d'années que nous considérons. Nous fixons une date de rupture en $t_0 = 7$, $t_0 = 10$ et $t_0 = 12$ suivant les cas que nous considérons, afin de cerner l'influence d'une rupture tardive sur la qualité de détection. Pour chacun des trois cas considérés, nous simulons 1000 trajectoires de la façon suivante.

1. Les observations de 1 à t_0 sont générées suivant une loi de Poisson de paramètre $\pi_i \lambda_0$, où λ_0 est une intensité de base, et les π_i sont les primes enregistrées l'année i .
2. Les observations de t_0 à 15 sont générées suivant une loi de Poisson de paramètre $\pi_i \lambda_i$, où $\lambda_i = \lambda_0 + \mu$.
3. On fait varier μ pour faire varier l'amplitude de rupture
4. Pour chaque trajectoire, et chaque valeur de μ , on compte le nombre moyen de fois où une rupture est détectée au niveau 1%.

Les graphiques suivants présentent le nombre moyen de détections en fonction de μ . Deux cas ont été considérés, le premier correspondant à un $\lambda_0 = 5e^{-4}$ proche de ce qui est observé sur la garantie Vol et la garantie RC Matériels, le deuxième cas pour un λ_0 proche de ce qui est observé sur RC Corporels ($\lambda_0 = 5e^{-5}$). Dans chacun des cas, le montant des primes annuelles est pris constant d'une année à l'autre (égal au montant moyen de prime reçues). Ces courbes de puissance s'interprètent de la façon suivante. Une bonne procédure de test a une courbe aux alentours de 0.01 à l'origine (correspondant au niveau du test), puis croît rapidement vers 1 (de plus en plus à mesure qu'on s'éloigne de l'hypothèse nulle).

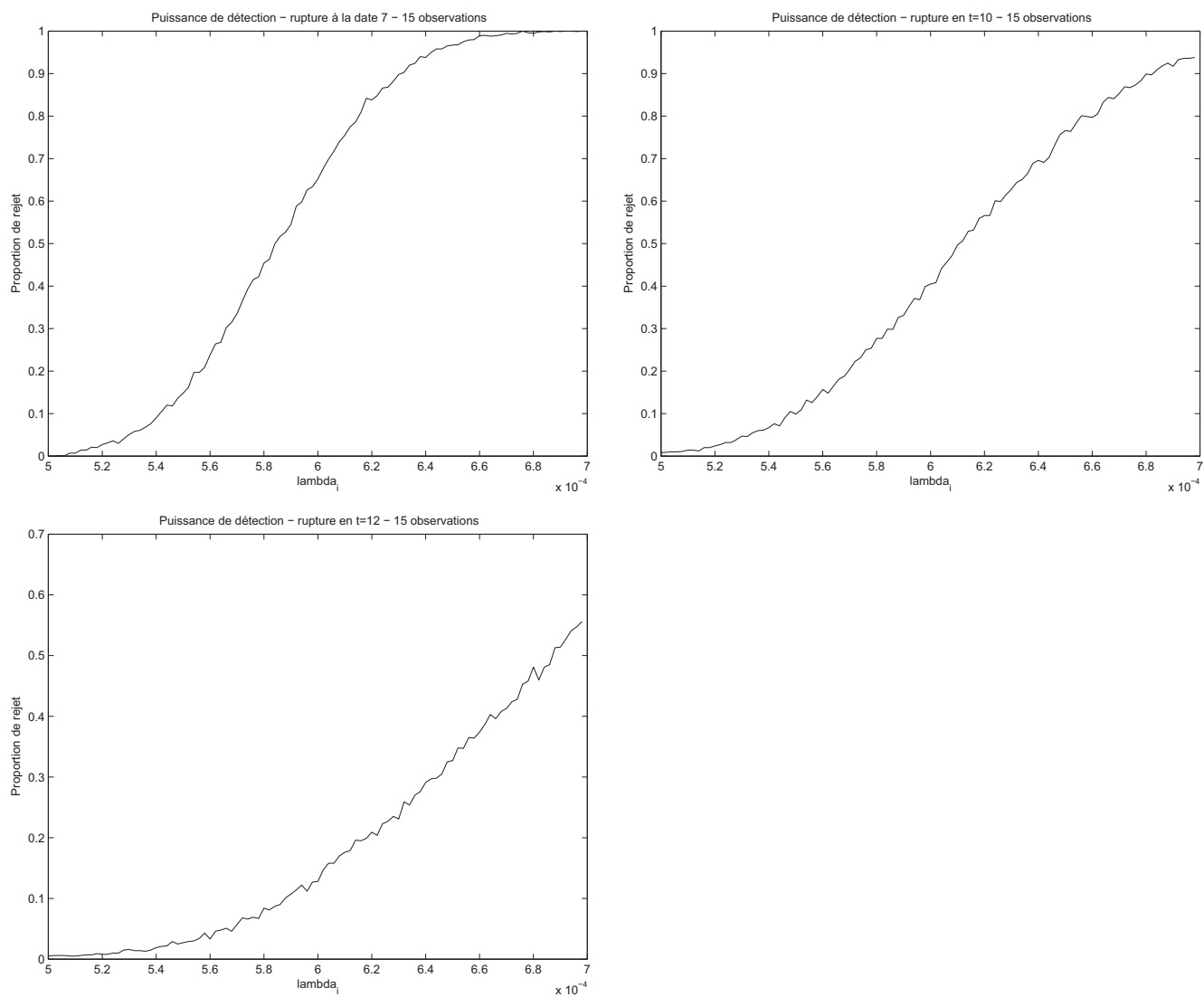


FIGURE 3.2 – Puissance de détection en fonction de l'intensité de la rupture, et de la date d'occurrence de la rupture, $\lambda_0 = 5e^{-4}$.

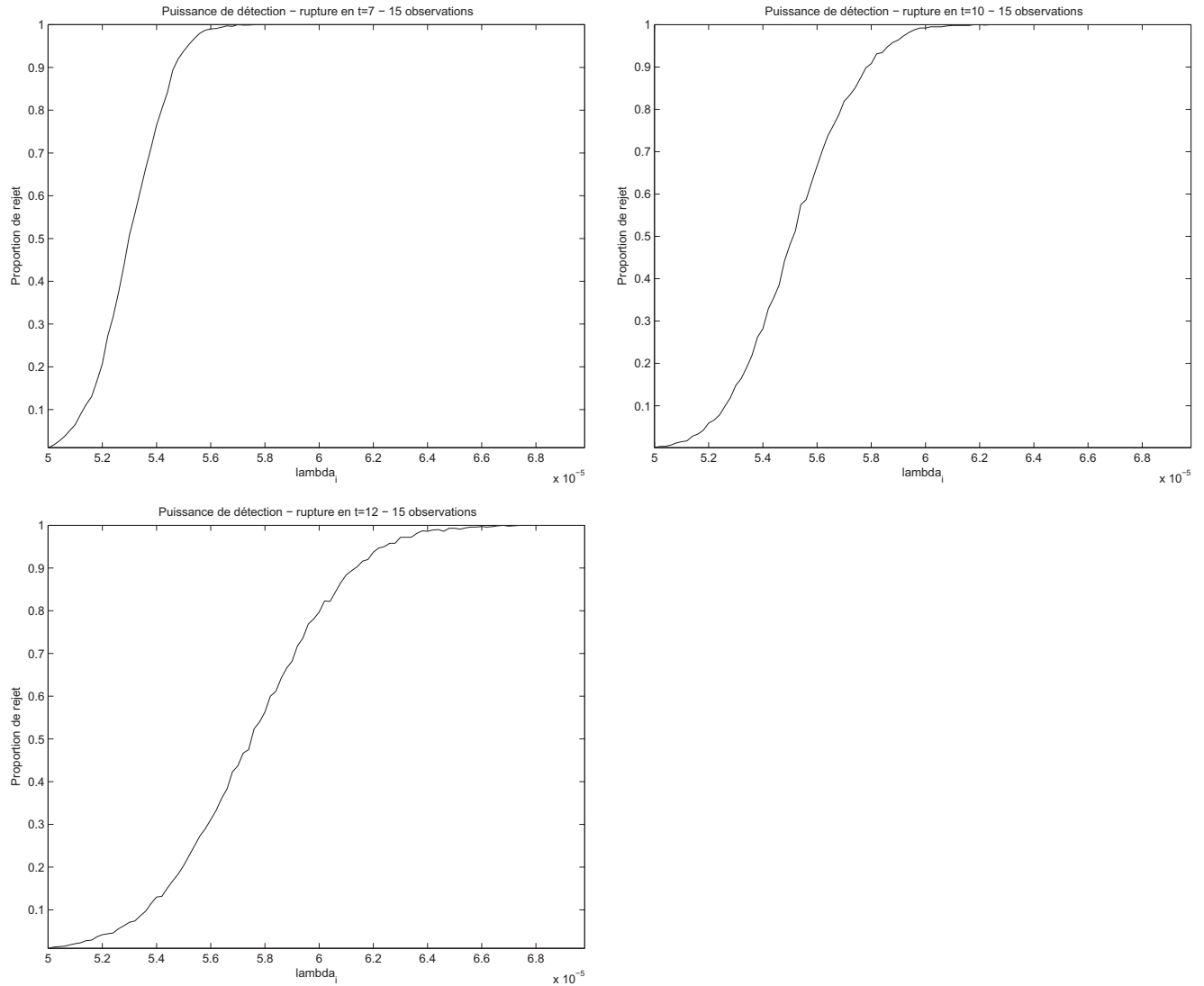


FIGURE 3.3 – Puissance de détection en fonction de l’intensité de la rupture, et de la date d’occurrence de la rupture, $\lambda_0 = 5e^{-5}$.

Comme attendu, plus la rupture est tardive, plus la puissance est dégradée. Pour la garantie Vol, une rupture est détectée pour la garantie vol à une date située à 7 années de la fin d’observation, et elle est de l’ordre de 2×10^{-4} . Si l’on compare avec les résultats de la figure 3.2, on note qu’une rupture présentant ces caractéristiques sera détectée avec une probabilité très proche de 1. Si elle n’avait lieu que 5 observations avant la fin d’observations ($t_0 = 10$), on aurait environ 95% de chances de la détecter. En revanche, on tombe à moins de 60% pour une rupture trop récente ($t_0 = 12$).

Dans le cas où $\lambda_0 = 5e^{-5}$ (figure 3.3), la puissance de la procédure relativement à l’ordre de grandeur du phénomène est meilleure, une rupture de l’ordre 2×10^{-5} est détectée de façon quasiment sûre dans les trois cas de figure.

Le deuxième point que nous présentons concerne la qualité d’estimation. Le but est

de s'interroger sur la qualité d'estimation de la date de rupture. Sur 1000 simulations, on présente la répartition des estimateurs de la date de rupture. Seuls les échantillons pour lesquels une rupture a été détectée ont été considérés. Les deux graphiques ci-dessous présentent le cas d'une rupture à la date 6, dont l'amplitude correspond à 1% et 0.5% du paramètre de Poisson avant rupture.

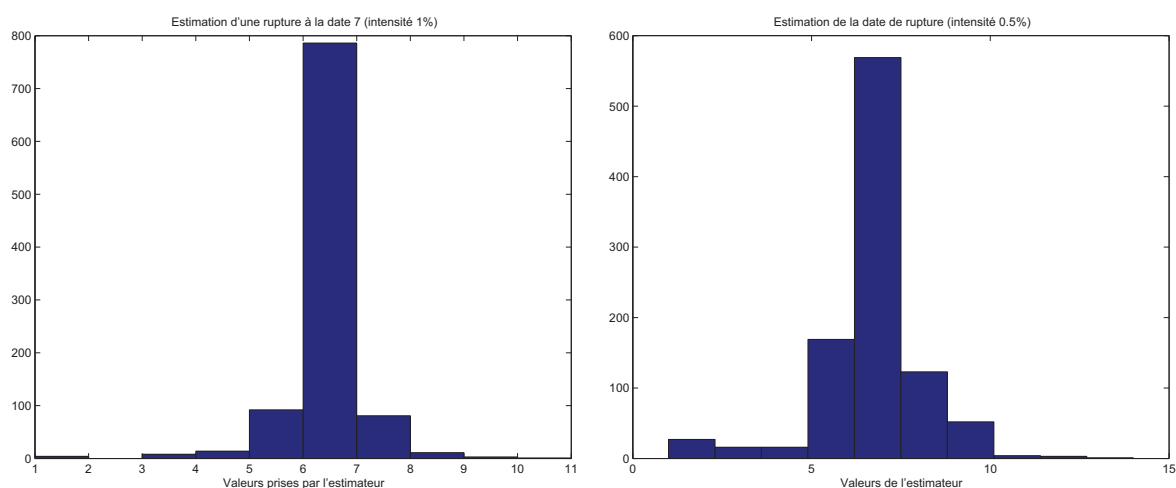


FIGURE 3.4 – Répartition des estimateurs de la date de rupture.

Dans chacun des cas, les estimateurs restent généralement concentrés autour de la véritable date de rupture (sauf cas rares). Les résultats concernant une rupture plus tardive ($t_0 = 11$) sont présentés dans le graphique suivant, et fournissent également une répartition satisfaisante.

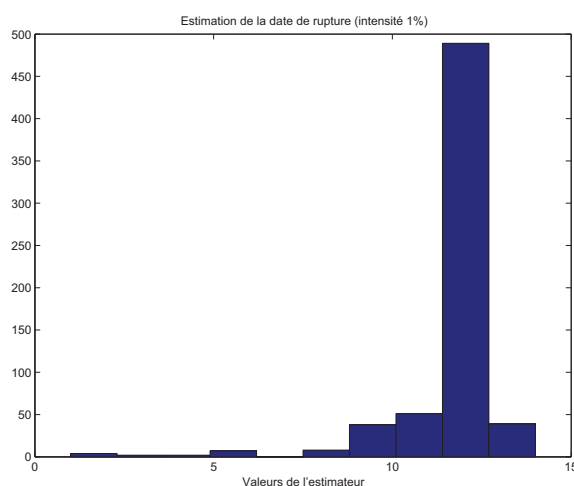


FIGURE 3.5 – Répartition des estimateurs de la date de rupture (rupture tardive).

3.3.4 Apport d'information

Pour tenter d'anticiper la détection, nous avons incorporé l'information supplémentaire qui était à notre disposition, grâce à la décomposition de l'année en périodes homogènes (section 3.3.1). Les observations sont alors constituées de $(\tilde{N}_k)_{1 \leq k \leq n}$, où n est le nombre de périodes de temps. Nous ne considérons que des ruptures pouvant se produire au changement d'année.

Nous détaillons le modèle considéré. Soit N_i le nombre de sinistres l'année i , qui suit une loi de Poisson de paramètre $\lambda_i \pi_i$, où λ_i suit le modèle (3.5). On a la relation

$$\lambda_i \pi_i = \tilde{\lambda}_i \sum_{k=1}^4 \alpha_k n_k,$$

où $\tilde{\lambda}_i$ correspond à l'intensité (journalière) du processus de Poisson sur la période de référence (période 1), n_k le nombre de jours de la période k , et α_k le coefficient estimé dans la section 3.3.1 (on omet le g qui traduisait la dépendance du coefficient suivant la garantie pour ne pas alourdir). On en déduit que

$$\tilde{\lambda}_i = \frac{\lambda_i \pi_i}{\sum_{k=1}^4 \alpha_k n_k}.$$

La variable $\tilde{N}_k$ suit donc une loi de Poisson de paramètre $\tilde{\lambda}_i n_k$, où i est l'année dans laquelle se trouve la période k .

La méthode de détection se transpose de façon identique dans ce cadre. L'utilisation de cette méthode ne change pas la date de rupture détectée sur les jeux de données que nous avons considérés. En revanche, la date à laquelle elle est détectée est avancée. Pour la garantie Vol, la détection a lieu à l'issue de la période 2 de l'année 2006 (au lieu de se produire à la fin de l'année). Pour la garantie Rccorp, cette détection a lieu à l'issue de la période 2 de l'année 2007 (au lieu de la fin de l'année).

Même si l'anticipation de détection est faible (on gagne moins d'une année), cette information permet de réviser en cours d'année la prévision du nombre de sinistres pour l'année en cours.

3.4 Conclusion

En mettant en œuvre les techniques de détection de ruptures, nous avons identifié, pour chacune des garanties étudiées, qu'on pouvait effectuer un découpage de l'année en 4 périodes sur lesquelles le nombre de sinistres possède un comportement suffisamment homogène. Par ailleurs, les méthodes que nous utilisons permettent d'améliorer la prédiction du nombre de sinistres d'une année sur l'autre. En effet, ces méthodes de prévision supposent une hypothèse de stabilité sur le risque considéré. En effectuant une détection de rupture, on s'autorise à écarter les années qui "polluent" la prévision puisqu'elles ne sont plus représentatives (un changement dans la structure du risque s'est opéré entre-temps).

Afin de détecter ou non la présence de rupture, nous avons proposé une approche basée sur le test du rapport de vraisemblance. Cette approche a pour but de ne pas conclure trop hâtivement à la présence d'une rupture. Néanmoins, nous remarquons que

cette approche (qui peut s'avérer presque trop prudente étant donnés les résultats de la section 3.3.2) peut sans doute être complétée par une lecture plus avisée des estimations successives d'une année sur l'autre. En effet, on remarque, dans les années suivant une rupture, que l'estimateur du maximum de vraisemblance se stabilise rapidement autour de la date effective de rupture. Ce phénomène mériterait de plus amples investigations.

Enfin, grâce au découpage de l'année en période de sinistralité similaire, nous parvenons à avancer sensiblement la détection, en permettant une réévaluation en cours d'année. Dans les cas des garanties Vol et RC corporelle, cette approche permet d'avancer la détection de 6 mois, et de réviser en conséquence les prévisions effectuées.

Chapitre 4

Modélisation des coûts de sinistre

Après avoir considéré l'étude du nombre de sinistres au cours du temps au Chapitre 3, l'étape suivante de notre démarche consiste à étudier le coût de ces sinistres, ainsi que la stabilité du modèle que nous considérons. Dans l'approche que nous présentons, nous nous concentrons principalement l'étude de l'espérance de la variable coût de sinistre. Les techniques de détection de rupture pourraient être modifiées pour s'appliquer à l'étude de la stabilité d'autres quantités telles que la variance ou des quantiles, mais cette approche n'est pas développée ici.

Dans cette partie, nous effectuons l'hypothèse que les coûts exacts des sinistres (par exacts, nous entendons tels qu'ils sont connus aujourd'hui) sont connus dès la survenance du sinistre. Cette hypothèse n'est pas réaliste, sauf éventuellement pour la branche vol qui est caractérisée par des vitesses de liquidation plus rapides. Elle nous permet néanmoins de simplifier notre étude, et de juger les méthodes de détection de rupture pour elles-mêmes. En effet, une incertitude supplémentaire vient se greffer sur notre procédure due à la qualité d'évaluation de la charge finale d'un sinistre. La question de cette qualité d'évaluation dépasse le cadre de notre étude.

La première approche que nous considérons pour l'étude de la stabilité du modèle de coûts de sinistres peut être qualifiée de non paramétrique puisqu'elle ne présuppose pas d'hypothèse sur la loi des coûts. Elle se base sur le contraste défini par la méthode du maximum de vraisemblance gaussien, présenté dans la section 4.1. Néanmoins, la méthode purement paramétrique basée sur la loi gaussienne échoue dans le cadre des données que nous considérons (voir section 4.1). La méthode non paramétrique que nous proposons s'impose comme une correction de cette méthode purement paramétrique. Elle repose sur la construction de seuils d'alertes à partir d'une méthode de bootstrap décrite dans la section 4.2.

La seconde approche, paramétrique, consiste à modéliser les coûts de sinistres par une variable de Pareto, et à étudier la stabilité de cette estimation à travers la série des paramètres estimés. La loi de Pareto, si elle est souvent utilisée pour modéliser des coûts de sinistres, paraît en revanche peu adaptée à nos données. Néanmoins, la technique que nous présentons dans la section 4.3 peut être utilisée, même en cas de défaut d'adéquation, pour obtenir des informations pertinentes sur l'évolution des coûts de sinistre.

La troisième approche combine les deux précédentes, en effectuant une séparation entre les sinistres les plus importants, qui sont modélisés par une loi de Pareto, et les sinistres les plus faibles, qui sont étudiés à travers la technique de la section 4.1. Cette dernière

approche est l'objet de la section 4.4. Les résultats des différentes méthodes considérées sont comparés les uns aux autres.

4.1 Maximum de vraisemblance gaussien

4.1.1 Expression de la vraisemblance

Dans le modèle de détection de rupture gaussien, les observations sont composées de variables $(X_i)_{1 \leq i \leq n}$ indépendantes telles que

$$X_i \sim \mathcal{N}(\mu(\theta, i), \sigma_i^2),$$

où $\mu(\theta, i) = f(\theta_1, i)\mathbf{1}_{i \leq i_0} + f(\theta_2, i)\mathbf{1}_{i > i_0}$, f étant une fonction fixée, θ_1 et θ_2 des paramètres de dimension finie. Nous supposons que la variance est **connue**, à un paramètre multiplicatif près (estimable à partir des données), de la forme

$$\sigma_i^2 = n_i \sigma_2,$$

où n_i est connu (dans la suite, il s'agira du nombre de sinistres enregistrés dans la période i). Le cadre du modèle de rupture classique consiste à considérer la fonction $f(\theta_j, i) \equiv \theta_j$, mais tout autre type de fonction paramétrique convient. Dans la suite, nous considérerons une fonction affine, avec $\theta_j = (\alpha_j, \beta_j)$ de dimension 2, puisque les données que nous considérons comportent une tendance linéaire manifeste.

Le modèle que nous considérons est donc défini par la fonction

$$\mu(\theta, i) = (\alpha_1 + \beta_1 i)\mathbf{1}_{i \leq i_0} + (\alpha_2 + \beta_2(i - i_0))\mathbf{1}_{i > i_0}.$$

Nous supposons que les paramètres α_1 et β_1 sont **connus**. Ceci est motivé d'une part par un souci de simplification, mais également notre problématique qui consiste à valider la stabilité d'un modèle. Or, dans le cadre d'une modélisation, les paramètres de base (α_1 et β_1) sont supposés connus exactement. Néanmoins, la procédure que nous allons décrire s'étend sans difficulté au cadre où α_1 et β_1 sont inconnus sans imposer de modifications majeures. Notons également que, pour obtenir un modèle identifiable, il est nécessaire d'imposer $\alpha_2 = \beta_2 = 0$ si $i_0 = n - 1$.

Soit $Z_i = Y_i - \alpha_1 - \beta_1 i$. La log-vraisemblance du modèle s'écrit (à une constante près),

$$L(\alpha, \beta, i_0) = \sum_{k=i_0+1}^n Z_k \sigma_k^{-2} \{ \alpha + \beta(k - i_0) - \alpha_1 - \beta_1 k \} - \{ \alpha + \beta(k - i_0) - \alpha_1 - \beta_1 k \}^2 / [2\sigma_k^2].$$

A i_0 fixé, les valeurs de α et β qui maximisent $L(\alpha, \beta, i_0)$ sont donc les estimateurs des moindres carrés usuels, basés sur la sous-population des observations de $i_0 + 1$ à n . On en déduit que la maximisation de la vraisemblance peut se faire de façon exacte et avec un temps de calcul faible de la façon suivante :

Implémentation :

- Pour chaque i plus petit que $n - 3$, calculer $\hat{\alpha}(i)$ et $\hat{\beta}(i)$, estimateurs des moindres carrés basés sur les observations $(Y_k)_{i+1 \leq k \leq n}$.
- Calculer $L_i = L(\hat{\alpha}(i), \hat{\beta}(i), i)$.
- Trier le vecteur des L_i et définir $\hat{i} = \arg \max_{0 \leq i \leq n-3} L_i$.
- Estimateurs finaux : $\hat{i}, \hat{\alpha} = \hat{\alpha}(\hat{i}), \hat{\beta} = \hat{\beta}(\hat{i})$.

La restriction à $i \leq n - 3$ est rendue nécessaire par le fait qu'on ne peut définir d'estimateur des moindres carrés de α_2 et β_2 si l'on considère une population de moins de 3 observations. Cette remarque n'est pas totalement neutre : si une rupture de grande amplitude se produit à l'observation $i_0 = n - 1$, cette rupture sera détectée de façon quasi-systématique, mais la date estimée ne pourra pas être égale à $n - 1$. En conclusion, une valeur élevée de $\hat{i}$, si elle n'est pas reliée à un phénomène évident (grande rupture), doit être prise avec précaution, et nécessite un peu plus de recul pour être complètement validée.

4.1.2 Rupture gaussienne et observations non gaussiennes

Si les observations ne sont pas gaussiennes, la vraisemblance gaussienne fournit une technique d'estimation "convergente" (dans le sens où, quand le recul augmente, on va toujours finir par détecter exactement le point de rupture). En effet, la quantité $L(\alpha, \beta, i)$ peut être vue comme un estimateur de

$$\sum_{k=i+1}^n [\mu(\theta, k) - \alpha_1 - \beta_1 k] \{ \alpha + \beta(k - i_0) - \alpha_1 - \beta_1 k \} - \{ \alpha + \beta(k - i_0) - \alpha_1 - \beta_1 k \}^2 / 2.$$

Supposons que $i \geq i_0$. Dans ce cas, $\hat{\alpha}(i)$ et $\hat{\beta}(i)$ sont des estimateurs consistants de α_2 et β_2 . Donc L_i peut être vu comme un estimateur de

$$\sum_{k=i+1}^n [\mu(\theta, k) - \alpha_1 - \beta_1 k]^2 / 2.$$

Comme cette quantité est une somme de terme positifs, le i qui la maximise est bien i_0 . Un raisonnement analogue dans le cas $i < i_0$ montre que i_0 maximise bien la cible du critère L_i , que les variables soient gaussiennes ou non, du moment que la moyenne de cette variable suit bien le modèle de rupture.

4.1.3 Impossibilité de la mise en oeuvre des tests gaussiens dans notre cadre

Si l'on se place dans un cadre statique, le cadre gaussien nous permet de fixer des seuils de rejet pour le test de

$$\begin{aligned} H_0 &: i_0 = n - 1 \text{ (pas de rupture)} \\ H_1 &: i_0 < n - 1 \text{ (rupture)}. \end{aligned}$$

Le test du rapport de vraisemblance consiste à rejeter H_0 si $\max_{1 \leq i \leq n-3} L_i > s(\alpha)$, où $s(\alpha)$ est un seuil dépendant du niveau de la procédure. Afin de construire un test de détection de rupture, on peut soit utiliser des bornes asymptotiques (voir Csörgö, Horváth, 2004), soit, de façon plus précise, utiliser des simulations. Nous détaillons cette procédure dans le cadre statique, le cadre dynamique sera étudié plus en détail dans les sections suivantes, lorsque nous en viendrons aux méthodes effectivement utilisées.

Calcul du seuil de rejet

- Pour $1 \leq k \leq N$ (N suffisamment grand), on simule $Y_i^{(k)} \sim \mathcal{N}(\alpha_1 + \beta_1 i, \sigma^2)$, pour $1 \leq i \leq n$ (correspondant à la loi sous H_0).
- Pour chaque k , évaluation de $L^{(k)} = \max_i L_i^{(k)}$.
- Triage du vecteur des $L^{(k)}$, on prend pour $s(\alpha)$ le $(1 - \alpha)N$ -ème élément du vecteur des $L^{(k)}$ trié.

Problème posé par cette approche dans notre cadre d'application : Appliquée à notre problème, cette technique aboutit à un phénomène de **sur-détection**. Mise en oeuvre (après modification) dans un cadre dynamique, la simulation de seuils à partir de la loi gaussienne aboutit à nous faire détecter une rupture à chaque instant (ou quasiment) sur notre série des coûts moyens mensuels de sinistres. Au moins deux phénomènes peuvent expliquer ce phénomène :

1. *L'hétéroscédasticité éventuelle des résidus gaussiens : ici nous avons supposé que les résidus avaient une variance du type $\sigma^2 n_i^{-1}$, où n_i est le nombre de sinistres enregistrés dans la période i . La constance du paramètre σ^2 peut être remise en doute, mais aucune modélisation évidente ne semble appropriée.*
2. *Ecart par rapport à la loi gaussienne : les variables considérées ont une loi bien différente de la loi gaussienne.*

Les sections suivantes présentent les approches que nous avons proposées pour dépasser cet écueil.

4.2 Méthode bootstrap basée sur la vraisemblance gaussienne

4.2.1 Bootstrap dans les résidus pour la détection de rupture

La loi des observations étant inconnue, les méthodes de type bootstrap forment une classe générale de procédure qui permet de palier à ce manque d'informations, en simulant artificiellement des échantillons à partir d'une distribution empirique. On pourra notamment consulter Davison et Hinkley (1997) pour une présentation générale des différentes techniques de bootstrap communément utilisées.

En l'absence de rupture, le modèle que nous considérons est un modèle de régression linéaire. Rappelons également que notre but est de simuler des trajectoires sous l'hypothèse nulle, pour en déduire un seuil de rejet. Afin de détailler notre approche, nous décomposons cette simulation en trois phases :

1. *Récupération des résidus estimés à partir de la série observée. Cette série se compose d'une variable Y_i (coût mensuel de sinistre), et d'un nombre de sinistres observés dans le mois, n_i .*
2. *Simulation d'une trajectoire $\hat{Y}_i$ par méthode bootstrap.*
3. *Obtention du seuil de rejet du test.*

Remarquons que, du fait de l'hypothèse d'indépendance entre les coûts de sinistres et le nombre de sinistres (hypothèse du processus de Poisson composé), nous n'avons pas besoin de simuler des réalisations des n_i . Ces nombres de sinistres peuvent être considérés comme fixes, c'est à dire qu'on travaille conditionnellement au nombre de sinistres effectivement réalisés.

1. Récupération des résidus

- Sélectionner une région de la courbe des Y_i où les observations sont sur une pente homogène. Soit m le nombre d'observations retenues.
- Estimer la droite des moindres carrés sur cette portion de courbe. On obtient des estimateurs $\hat{\alpha}_0$ et $\hat{\beta}_0$ de l'ordonnée à l'origine et de la pente.
- Récupérer les résidus estimés, $\hat{\varepsilon}_i = Y_i - \hat{\alpha}_0 - \hat{\beta}_0 i$.
- Afin de prendre en compte l'hétérogénéité due aux différents n_i , normalisation des $\hat{\varepsilon}_i$ en définissant $\varepsilon_i^* = \hat{\varepsilon}_i n_i^{-1/2}$.

2. Simulation d'une trajectoire

- Soit n la longueur de trajectoire qu'on souhaite simuler. Pour $1 \leq i \leq n$, tirer de façon uniforme un $\tilde{\varepsilon}_i$ parmi l'ensemble des $\{\varepsilon_j^* : 1 \leq j \leq m\}$ construits à l'étape 1.
- Le i -ème point simulé est défini par $\hat{Y}_i = \hat{\alpha}_0 + \hat{\beta}_0 i + n_i^{1/2} \tilde{\varepsilon}_i$.

3. Obtention des seuils de rejet (non dynamiques) pour des trajectoires de taille n

- Simulation de N trajectoires en suivant la méthode 2, N suffisamment grand.
- Pour chaque trajectoire k , calcul de $L^{(k)} = \max_i L_i^{(k)}$.
- Triage du vecteur des $L^{(k)}$, on prend pour $s(\alpha)$ le $(1 - \alpha)N$ -ème élément du vecteur des $L^{(k)}$ trié.

4.2.2 Détection dynamique

L'extension à la détection dynamique est plus difficile du fait de l'hétérogénéité des observations (les n_i sont différents). Rappelons que les seuils de détection obtenus en détection statique ne peuvent être utilisés directement, notamment du fait de la dépendance entre deux statistiques de test successives. Une difficulté supplémentaire est introduite ici du fait que les observations ne sont pas homogènes de part et d'autres de la rupture : le nombre de sinistres étant à chaque fois différent d'un mois sur l'autre, ceci modifie notamment la variance des observations. Dans un cadre homogène, un résultat comme

celui du Théorème 1 fournit une réponse au problème de la détection dynamique : si n désigne le nombre d'observations, la log-vraisemblance croît artificiellement en $\log \log n$. On en conclut qu'on peut définir une courbe d'alerte du type $i \rightarrow C \log \log i$, où C est une constante qui ne dépend pas de i . On peut ensuite facilement calibrer la constante C par simulations, de sorte que seule une proportion α des trajectoires dépasse le seuil d'alerte sous l'hypothèse nulle.

En revanche, dans le problème que nous considérons, à cause de l'hétérogénéité, la constante C doit dépendre de i (à moins de prendre une constante trop importante qui réduise considérablement la puissance de notre procédure). Nous avons donc développé une méthode heuristique afin de construire une courbe d'alerte qui soit raisonnable. Notre idée de base consiste à supposer que cette courbe est proche de la courbe des seuils statiques, puis de relever cette courbe suffisamment pour atteindre le niveau souhaité.

Obtention de la courbe d'alerte

- Soit n_{max} la durée maximale sur laquelle on souhaite effectuer le test.
- Pour chaque $n \leq n_{max}$, calcul de $s_\alpha(n)$, le seuil de rejet statique pour des trajectoires de taille n .
- Recherche d'une courbe du type $n \rightarrow C s_\alpha(n)$, où C est une constante. Initialisation de C à 1.
- Simulation de N trajectoires de taille n_{max} en suivant la procédure du paragraphe précédent.
- Tant que plus d'une proportion α de trajectoires dépassent la courbe $C s_\alpha(n)$, incrémenter $C = C + p$, où p est un pas suffisamment petit.
- La valeur de C obtenue à l'arrêt de la boucle définit une courbe d'alerte appropriée.

4.2.3 Application aux coûts moyens mensuels

Nous présentons à présent les résultats obtenus sur la courbe des coûts moyens mensuels pour les différentes garanties considérées. Pour étudier ces séries en utilisant la méthode bootstrap, nous avons utilisé un nombre d'itération $N = 10000$ dans la phase de génération bootstrap, après avoir récupéré des résidus ε_i sur les $m = 20$ premières observations. Après chaque détection de rupture, la procédure se réinitialise. Il est nécessaire d'attendre un certain laps de temps (ici $m = 20$) pour relancer une phase de tests, puisque la méthode bootstrap nécessite de récupérer de nouveaux résidus. Dans cette période de re-calibration, les résidus de la période précédente peuvent être réutilisés pour vérifier que la période est bien homogène. Dans certains cas, si la détection intervient plus de m observations après la rupture, les m résidus peuvent être calculés directement, et la re-calibration est donc instantanée. Mentionnons que ce changement de résidus servant de base au bootstrap après chaque rupture détectée est important, puisqu'on peut constater sur certaines garanties que la loi de ces résidus semble évoluer fortement avec le temps. En résumé, en cas de détection de rupture :

1. Soit la date détectée (d) est séparée de la date de détection (D) par plus de m observations. On utilise alors les m premières observations pour évaluer une nouvelle pente, et une nouvelle ordonnée à l'origine.
2. Soit $D - d = k$ avec $k < m$. On calcule une nouvelle pente et une nouvelle ordonnée à l'origine à partir des k premières observations, et on effectue un test de détection de rupture statique sur les observations comprises entre d et D , avec deux possibilités :
 - Si une rupture est détectée à la date d' , on recommence l'opération à partir de la date d' .
 - Si aucune rupture n'est détectée, à la date $D + 1$, on dispose de $k + 1$ observations pour améliorer l'estimation de la pente et de l'ordonnée à l'origine. Si $k + 1 < m$, on recommence l'étape précédente. Sinon, on garde la nouvelle droite estimée comme nouveau modèle.

4.2.3.1 Garantie Vol

Nous présentons ici la série des coûts moyens mensuels pour la Garantie Vol entre janvier 1994 et décembre 2008.

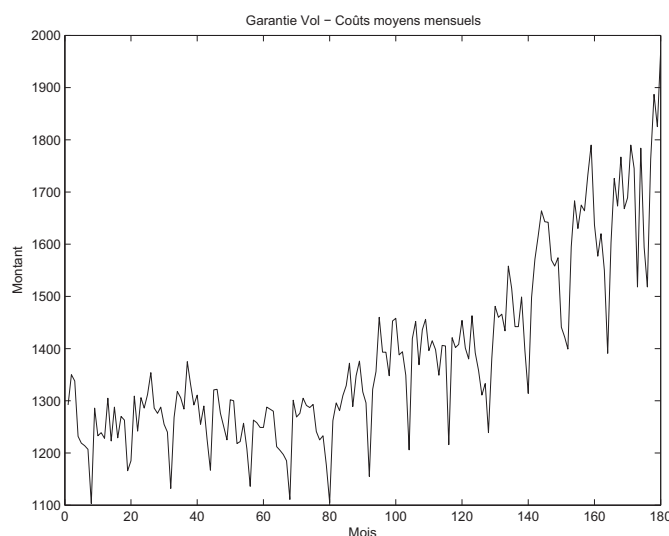


FIGURE 4.1 – Série des coûts moyens mensuels pour la Garantie Vol.

Cette série présente une saisonnalité manifeste de période 12. On remarque notamment trois mois (juin, juillet, août), correspondant à des coûts systématiquement plus faibles. Ces trois mois seront donc étudiés à part. Nous considérons donc à présent la série expurgée des mois d'été. Dans tout ce qui suit, les années sont donc composées de 9 mois, l'observation $9k + 1$ désignant le mois de janvier de l'année $k + 1$ etc.

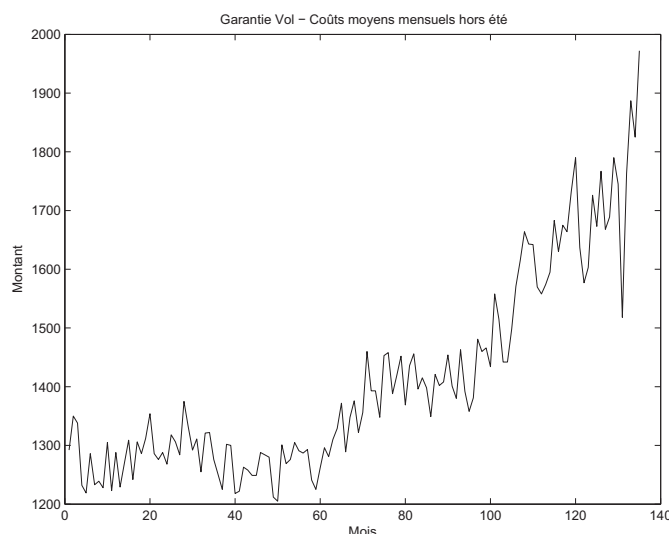


FIGURE 4.2 – Série des coûts moyens mensuels pour la Garantie Vol, hors période d'été.

Cette série ne présente plus de saisonnalité notable. Pour la détection dynamique, rappelons que l'intégralité de la courbe n'est pas utilisée, mais que nous introduisons une nouvelle observation au fur et à mesure. Nous reproduisons sur la page suivante (Figure 4.2.3.1) les différentes étapes de la détection dynamique de rupture sur cette série. Les graphiques ci-dessous représentent les courbes d'alertes aux niveaux 10%, 5 % et 1%. L'irrégularité de ces courbes de rejet est normale, et ne reflète pas seulement la fluctuation d'échantillonnage, mais vient surtout du fait que les n_i sont hétérogènes. La courbe en gras représente la valeur du maximum de vraisemblance basé sur les informations disponibles à la date courante. Dès que cette courbe franchit la courbe d'alerte, on conclut à une rupture et on estime cette rupture. Une nouvelle étape commence, avec une nouvelle calibration. La sixième rupture enregistrée n'est pas représentée graphiquement car elle intervient au cours de la période de calibration.

De façon plus synthétique, le résumé de la détection peut être présenté de la façon suivante sous forme de tableau. Nous nous focalisons essentiellement sur la détection de l'instant de rupture, plus que sur l'estimation des nouveaux paramètres de pente après rupture, cette estimation étant ensuite un problème classique une fois que la rupture a été fixée.

Rupture	Seuil 10%		Seuil 5%		Seuil 1%	
	Date détection	Date détectée	Date détection	Date détectée	Date détection	Date détectée
1	41	27	42	27	44	27
2	54	49	54	49	54	49
3	65	58	65	58	65	58
4	84	70	84	70	84	70
5	100	94	101	94	101	94
6	122	119	122	119	123	119

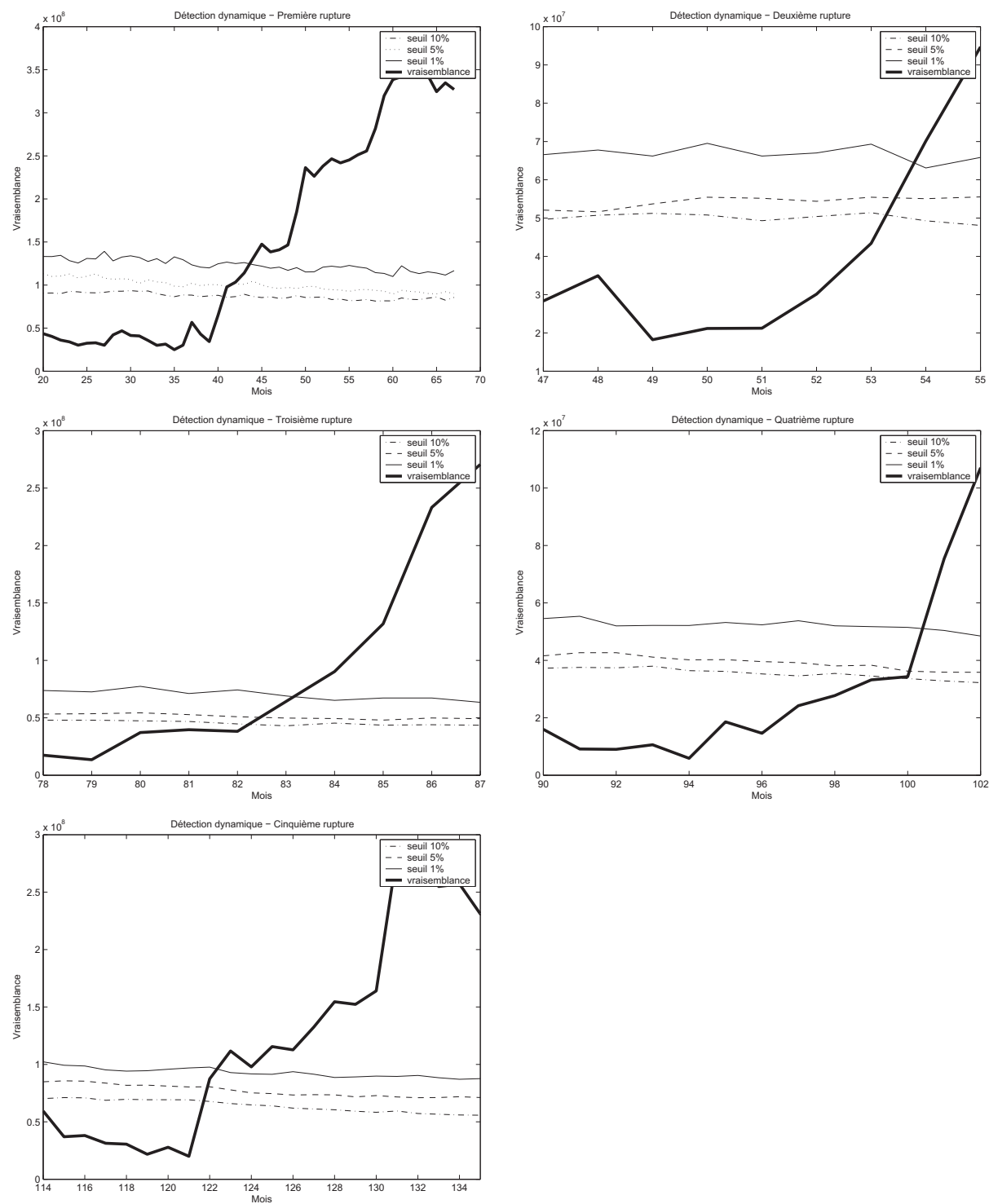


FIGURE 4.3 – Représentation graphique des différentes étapes de la détection dynamique de rupture pour la garantie vol.

On remarque que si les dates de détection changent (ce qui est logique puisque le seuil 1% nécessite de prendre plus de recul), les dates détectées sont toutes identiques, ce qui montre une bonne stabilité de la procédure d'estimation de la date de rupture.

Nous présentons également un résumé graphique de la détection sur la courbe complète. En plus de la série des coûts moyens, nous superposons deux courbes. La courbe finale représente le modèle qu'on retiendrait vu de la date 135 (les ruptures sont reportées sans délai). La courbe intitulée "ajustement dynamique" tient compte du délai de détection, et montre le retard pris avant la prise de décision de réajuster le modèle. Nous nous basons sur le seuil 5% pour la construction de ces courbes. Pour plus de lisibilité, les résultats sont présentés sans tenir compte du temps éventuel de réajustement du modèle (on rappelle que si le délai entre détection et date détectée est court, on se donne du temps pour réévaluer la pente et l'ordonnée à l'origine).

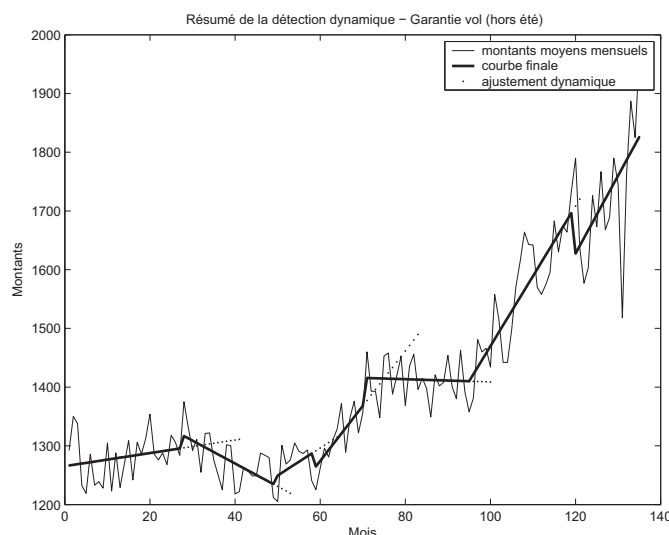


FIGURE 4.4 – Résumé de la détection dynamique de rupture pour la garantie vol (hors période d'été). La courbe finale est obtenue à partir du modèle estimé de façon dynamique, vu de la date 135. L'ajustement dynamique matérialise le retard de détection.

Voici le modèle retenu vu de la date 135. Pour chaque ligne, l'équation du tronçon de courbe correspondant est obtenue par la formule $y = \alpha + \beta(i - i_0)$, où i_0 désigne la dernière date de rupture enregistrée ($i_0 = 0$ pour la première ligne).

Tronçon de courbe	α	β
1 à 27	1265.4	1.1135
28 à 49	1320.6	-3.897
50 à 58	1244.7	4.6872
59 à 70	1255.5	9.3684
71 à 94	1415.8	-0.23008
95 à 119	1398	11.933
120 à 135	1614.2	13.262

Conclusions : La durée de détection varie entre 14 mois pour les ruptures les plus faibles, et est quasi-instantanée pour les grandes amplitudes. Le nombre de ruptures détectées est relativement important, une rupture au bout de 25 observations environ, rendant la modélisation assez instable.

Etude des trois mois d'été : Nous avons étudié chaque mois séparément. En effet, ces trois mois ne peuvent être regroupés du fait de comportement assez différents (voir annexe D).

4.2.3.2 Garantie Rccorp

Nous présentons ici la série des coûts moyens mensuels pour la Garantie Rccorp entre janvier 1994 et décembre 2008.

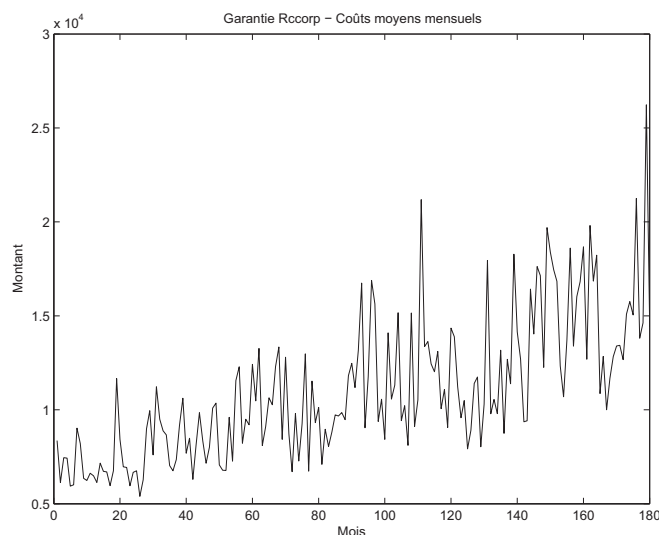


FIGURE 4.5 – Série des coûts moyens mensuels pour la Garantie Rccorp.

La série ne présente pas de saisonnalité manifeste (voir tests de stationnarité en annexe). Deux ruptures sont détectées au cours du temps.

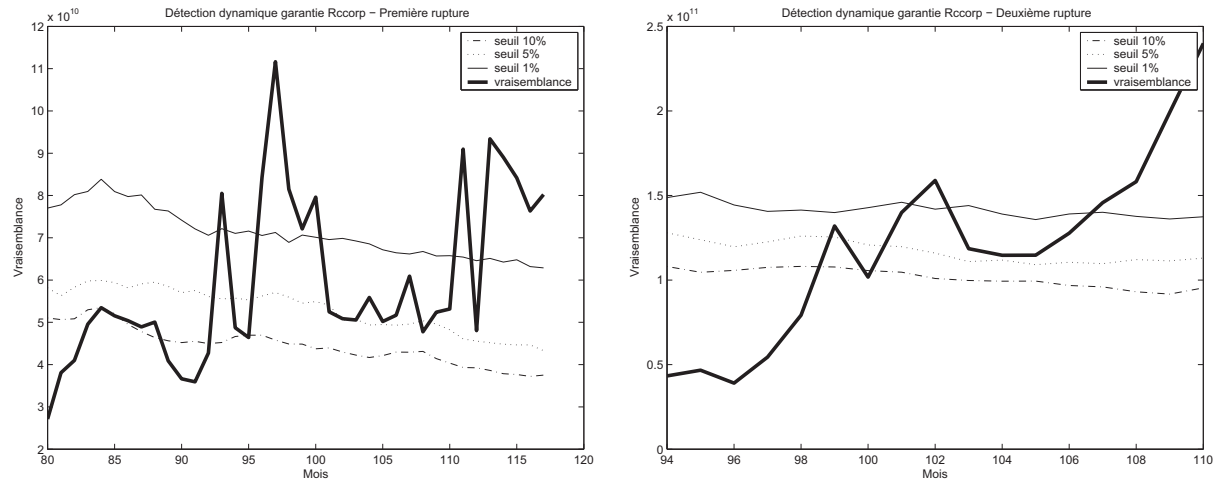


FIGURE 4.6 – Représentation graphique des différentes étapes de la détection dynamique de rupture pour la garantie Rccorp.

Nous présentons le tableau des dates de détection, et dates détectées, ainsi que le modèle retenu.

Rupture	Seuil 10%		Seuil 5%		Seuil 1%	
	Date détection	Date détectée	Date détection	Date détectée	Date détection	Date détectée
1	86	26	93	80	93	80
2	99	95	99	95	102	95

On observe une grande différence entre les dates détectées à l'occasion de la première rupture suivant qu'on se place au seuil 1% ou aux seuils 5 et 10%. Néanmoins, on pourra se rapporter à la figure 4.2.3.2 pour remarquer que le seuil 10% n'est dépassé que très légèrement à la date 26, par ailleurs la courbe redescend assez rapidement en-dessous du seuil. Nous ne retiendrons donc pas la date 26 comme date de rupture. Néanmoins, mentionnons le fait qu'écarter ou non la date 26 a peu d'incidence, puisque, si elle est retenue et que le modèle est recalibré sur elle, on détecte tout de même une rupture à la date 93, identifiée à la date 80.

Modèle retenu :

Tronçon de courbe	α	β
1 à 92	6728.9	43.726
93 à 99	8633	219.43
100 à 180	11960	38.12

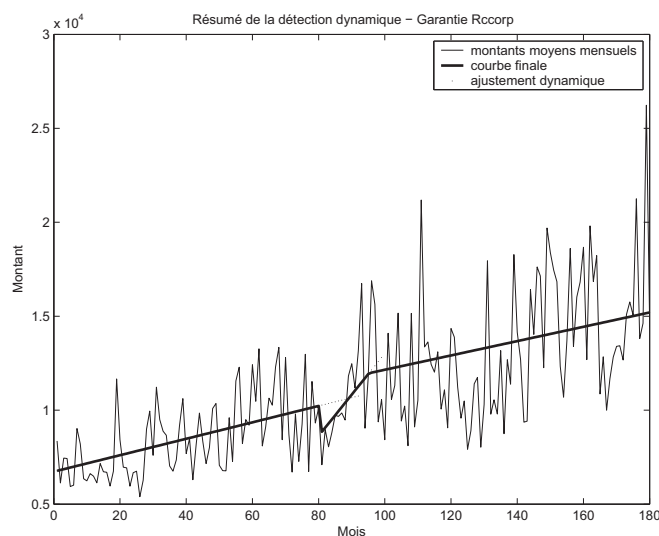


FIGURE 4.7 – Résumé de la détection dynamique de rupture pour la garantie Rccorp. La courbe finale est obtenue à partir du modèle estimé de façon dynamique, vu de la date 180. L'ajustement dynamique matérialise le retard de détection.

Conclusions : L'évolution de cette branche semble assez stable (2 ruptures détectées seulement). La première rupture est détectée avec un délai de plus d'un an, mais on peut remarquer que ce retard conduit à une évaluation plus prudente de l'évolution (on sur-estime la tendance entre).

4.2.3.3 Garantie Rcmat

Nous présentons ici la série des coûts moyens mensuels pour la Garantie Rcmat entre janvier 1994 et décembre 2008.

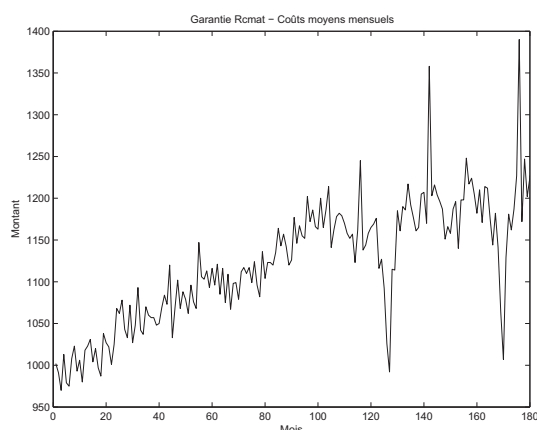


FIGURE 4.8 – Série des coûts moyens mensuels pour la Garantie Rcmat.

La présence de pics dans la dernière partie de la courbe rend la méthode de détection instable.

Résumé de la détection dynamique :

Rupture	Seuil 10%		Seuil 5%		Seuil 1%	
	Date détection	Date détectée	Date détection	Date détectée	Date détection	Date détectée
1	25	21	26	23	26	23
2	55	52	56	53	57	54
3	65	58	65	58	65	58
4	85	79	85	79	86	79
5	106	96	106	96	106	96
6	116	115	116	115	116	115
7	117*	104	117*	104	117*	104
8	125	124	125	124	125	124
9	132*	129	132*	129	132*	129
10	142	141	142	141	142	141
11	170	169	170	169	170	169
12	176	175	176	175	176	175

Les lignes en gras correspondent aux pics (positifs ou négatifs) que connaît la série dans sa deuxième partie. Les modèles de rupture ne sont pas adaptés à la prise en compte de ces pics, même s'ils les détectent instantanément. Pour prendre en compte ces pics, une modélisation de l'occurrence de ces pics par un processus à saut serait sans doute plus adaptée. Dans une première approche, nous écartons les pics en les remplaçant par leur valeur prévue à l'instant les précédant. Les lignes avec des étoiles mentionnent une détection faite une fois les pics ôtés de la série.

Modèle retenu.

Tronçon de courbe	α	β
1 à 23	986.54	1.6865
24 à 53	1053.4	0.48947
54 à 58	1100.8	-0.091729
59 à 79	1098.7	0.17218
80 à 96	1114.4	3.2932
97 à 104	1163.9	3.9048
105 à 129	1193.3	-3.9762
130 à 180	1191.9	0.0073394

Conclusions : La plupart des délais de détection sont faibles (deux seulement sont supérieurs à une année, et aboutissent à des évaluations plus prudentes). En revanche, le modèle n'est pas adapté à la bonne compréhension de la fin de la courbe, où l'on a manifestement un changement en variance qui se superpose au changement en moyenne, qui se traduit par la présence de pics. Une modélisation plus approfondie de ces pics serait nécessaire.

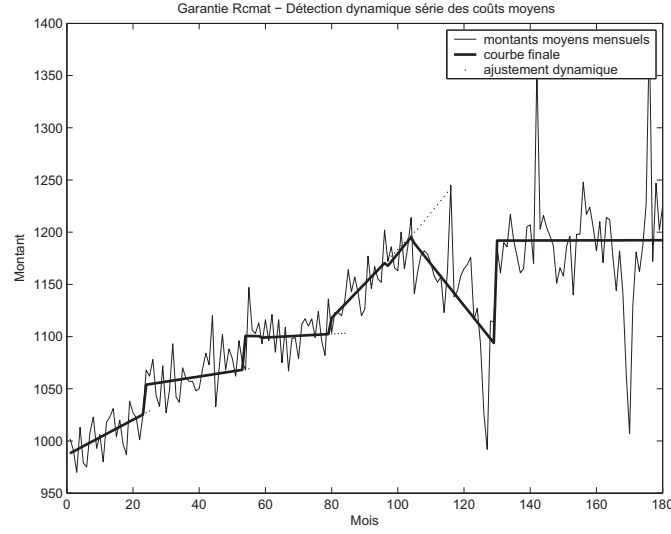


FIGURE 4.9 – Résumé de la détection dynamique de rupture pour la garantie Rccorp. La courbe finale est obtenue à partir du modèle estimé de façon dynamique, vu de la date 180. L'ajustement dynamique matérialise le retard de détection.

4.3 Modèle de Pareto

4.3.1 Modélisation du coût de sinistre par une variable de Pareto

La loi de Pareto, de paramètres (x_0, α) , est la loi d'une variable aléatoire X à valeurs dans $[x_0, \infty]$, de fonction de survie

$$\mathbb{P}(X > x) = \left(\frac{x_0}{x}\right)^\alpha,$$

pour $x > x_0$. L'intérêt de la modélisation d'un coût de sinistre par une loi de Pareto tient principalement à deux raisons, sa facilité d'application dans l'étude de sinistres franchisés, et son importance dans la théorie des extrêmes, lui permettant ainsi de rendre compte du comportement des sinistres importants. En effet, on peut remarquer que

$$\mathbb{P}(X > x | X > y) = \left(\frac{y}{x}\right)^\alpha,$$

donc que la loi de X conditionnellement à $X > y$ (y pouvant être interprété comme un changement de franchise) est une loi de Pareto de même paramètre d'échelle α . Par ailleurs, la loi de Pareto est une loi des extrêmes, à queue de distribution suffisamment lourde pour bien décrire un certain nombre de phénomènes.

Si l'on modélise la loi du coût de sinistre par une loi de Pareto de paramètres $(1, \alpha)$, et qu'on dispose des montants de sinistres $X_1, \dots, X_{n_i}$ au cours du mois i , l'estimateur du maximum de vraisemblance du paramètre α pour le mois i a l'expression suivante,

$$\hat{\alpha}(i) = \frac{n_i}{\sum_{j=1}^{n_i} \log X_j}. \quad (4.1)$$

Cette formule se redémontre aisément en remarquant que $\log X_j$ suit une loi exponentielle de paramètre α .

Reformulé dans le domaine de la modélisation par Pareto, le problème de détection de rupture s'exprime de la façon suivante :

- Jusqu'au mois i_0 , les sinistres suivent une loi de Pareto de paramètres $(1, \alpha_0)$ (α_0 connu).
- A partir du mois $i_0 + 1$, les sinistres suivent une loi de Pareto de paramètres $(1, \alpha)$, avec $\alpha \neq \alpha_0$.

Il s'agit d'estimer i_0 et α . Dans ce modèle, on suppose donc que tous les sinistres d'un même mois suivent la même loi, et qu'une rupture ne se produit donc pas en cours de mois.

4.3.2 Méthode de détection de rupture pour le modèle de Pareto

L'estimateur $\hat{\alpha}(i)$ est une statistique exhaustive pour le paramètre α à partir de l'échantillon $(X_1, \dots, X_{n_i})$ (cette propriété se montre facilement en remarquant que l'observation de l'échantillon des X_i est équivalente à l'observation de leur logarithme qui suit une loi exponentielle). De cette propriété vient l'idée de regarder directement l'évolution de l'estimateur $\hat{\alpha}(i)$ au cours du temps.

Les variables $(\hat{\alpha}(i))_{1 \leq i \leq n}$, sont supposées indépendantes puisque les sinistres des différents mois sont indépendants les uns des autres. Par ailleurs, si les X_i suivent une Pareto de paramètres $(1, \alpha)$, la loi des $\hat{\alpha}(i)$ est connue. En effet,

$$\frac{n_i}{\hat{\alpha}(i)} \sim \Gamma(n_i, \alpha),$$

où la loi $\Gamma(n_i, \alpha)$ a pour densité

$$f_{n_i, \alpha}(x) = \frac{\lambda^{n_i} x^{n_i-1} \exp(-\lambda x)}{\Gamma(n_i)},$$

où Γ désigne la fonction Gamma. Cette propriété se démontre aisément en remarquant que $\log X_i \sim \mathcal{E}(\alpha)$, où $\mathcal{E}$ désigne la loi exponentielle. Alternativement, la loi asymptotique des $\hat{\alpha}(i)$ peut être utilisée, si le nombre d'observations n_i est suffisant. Dans les applications que nous considérons, les deux approches donnent des résultats identiques. Nous préférons cependant présenter la procédure plus générale basée sur les lois Gamma.

Le modèle de rupture s'écrit donc :

$$n_i \hat{\alpha}(i)^{-1} \sim \Gamma(n_i, \alpha(i)),$$

où n_i est une suite fixée, et

$$\alpha(i) = (\alpha_0 + \beta_0 i) \mathbf{1}_{i \leq i_0} + (\alpha + \beta(i - i_0)) \mathbf{1}_{i > i_0},$$

avec α_0 connu, i_0 et α inconnus, n_i étant une suite fixée.

Si on considère une séquence de n mois, la log-vraisemblance s'écrit alors de la façon suivante,

$$L_n(i, \alpha, \beta) = \sum_{j=1}^i \{n_j \log(\alpha_0 + \beta_0 i) - (\alpha_0 - \beta_0 j) Z_j\} \\ + \sum_{j=i+1}^n \{n_j \log(\alpha + \beta(j - i)) - (\alpha - \beta(j - i)) Z_j\}.$$

La détermination des seuils de rejet pour les tests de rupture se fait de la même manière que dans le cas gaussien (section 4.2) en remplaçant les simulations de gaussiennes par des simulations de lois Gamma, à une différence près : l'estimateur du maximum de vraisemblance de la droite de régression après rupture n'est plus l'estimateur des moindres carrés dans le modèle Gamma. L'estimateur du maximum de vraisemblance n'a d'ailleurs plus de formule explicite. Une méthode numérique peut être employée. Nous avons choisi de conserver l'estimateur des moindres carrés, en raison de sa simplicité et pour des raisons de temps de calcul. L'estimateur que nous considérons pour la date de rupture est donc

$$\hat{i} = \arg \max_i L_n(i, \hat{\alpha}(i), \hat{\beta}(i)),$$

où $\hat{\alpha}(i)$ et $\hat{\beta}(i)$ désignent les estimateurs des moindres carrés de l'ordonnée à l'origine et de la pente, basés sur les observations de $i + 1$ à n . En toute rigueur, cet estimateur est sensiblement différent de l'estimateur du maximum de vraisemblance.

4.3.3 Simulations

Dans cette section, nous étudions par simulations le comportement de cette procédure en se focalisant sur le délai de détection. Nous générons 1000 échantillons de $n = 70$ observations de loi de Pareto de paramètres $\alpha_i = 2.6 - 0.005i + d_i$, où d_i est un paramètre de déviation. Ce paramètre est fixé à $d_i = k \mathbf{1}_{i > 30}$, où 30 est l'instant où se produit la rupture, et k un paramètre qui représente l'intensité de la déviation par rapport à l'hypothèse nulle d'absence de rupture ($k = 0$ correspond à une absence de rupture, k grand correspond au contraire à une rupture importante).

Pour chacune des itérations, nous considérons plusieurs trajectoires correspondant à différentes déviations d'intensité k , et nous mettons en œuvre la procédure de détection dynamique définies plus haut. Pour chaque itération et chaque k , nous stockons la première date pour laquelle une rupture est détectée. Le graphique suivant présente la date moyenne de détection suivant l'amplitude de la déviation, et pour les procédures basées sur le niveau de test 10%, 5% et 1%.

Pour les ruptures de valeur faible, le délai de détection est important (environ 40 observations de délai en moyenne). En revanche, dès qu'on augmente suffisamment la déviation, on remarque une rapide stabilisation de la courbe légèrement au-dessus de 31, qui est la date de rupture effective.

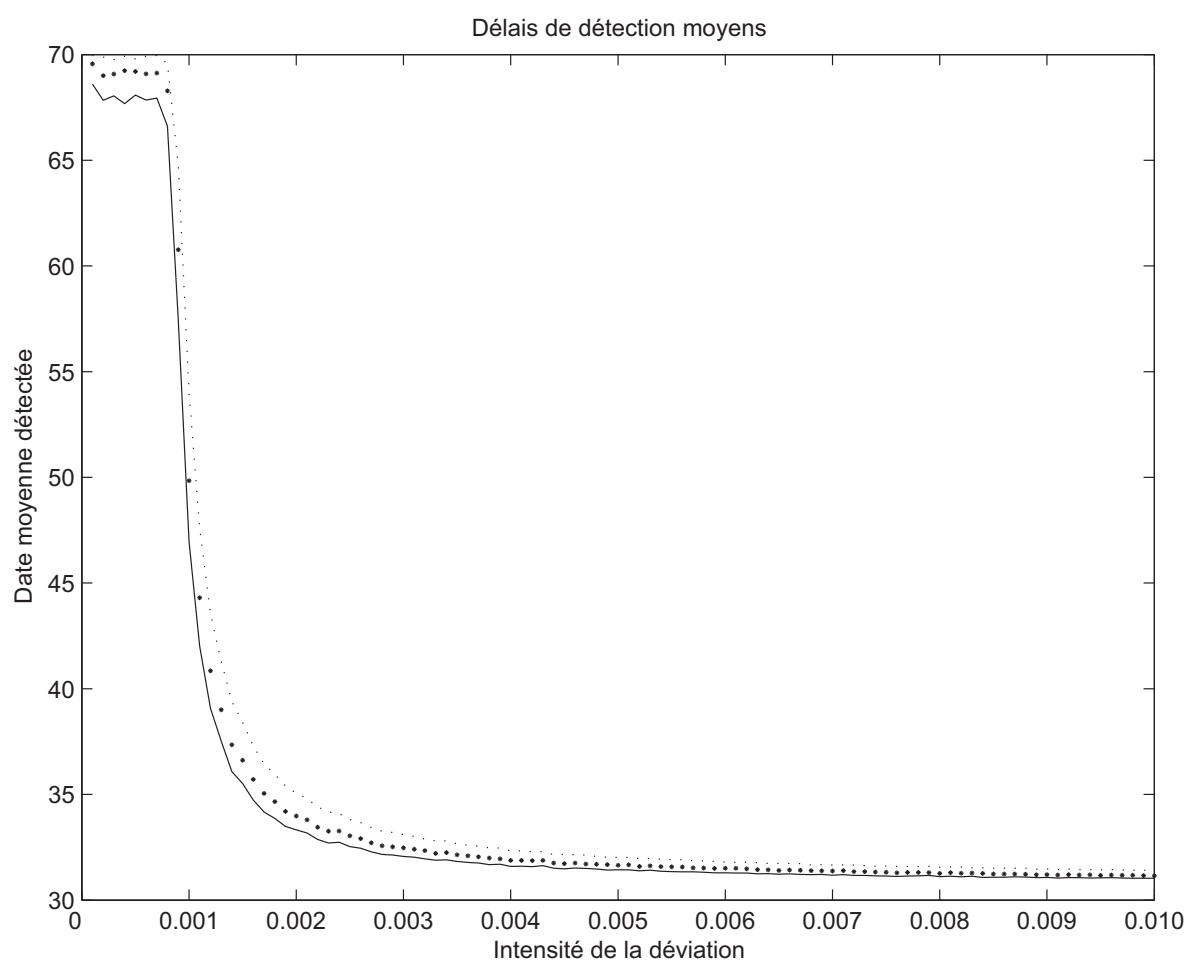


FIGURE 4.10 – Délai de détection de rupture dans le modèle de Pareto.

4.3.4 Application sur les données

Nous appliquons les mêmes principes pour la méthode de détection dynamique que dans la section 4.2.

4.3.4.1 Garantie Vol

La série des paramètres de Pareto estimés chaque mois présente la même saisonnalité que la série des coûts moyens. Nous présentons directement la série de ces paramètres pour la période hors été. Rappelons que plus le paramètre de Pareto diminue, plus la queue de distribution de la variable aléatoire associée est lourde. La diminution globale de la valeur des paramètres de Pareto est donc cohérente avec l'augmentation des coûts moyens observée dans la section précédente.

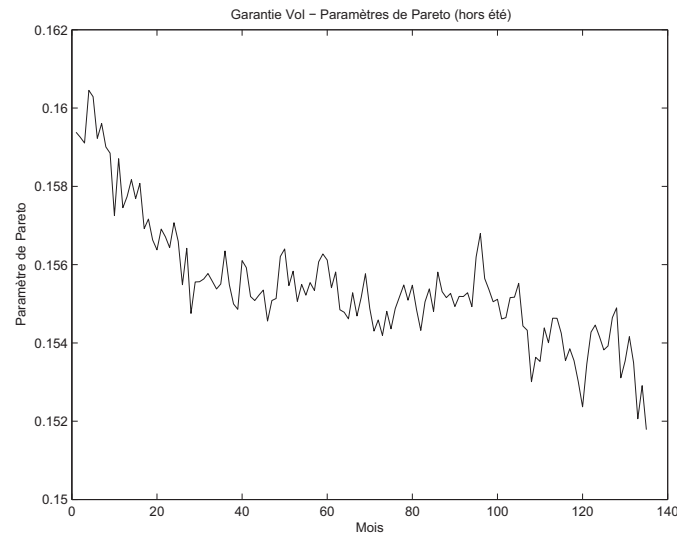


FIGURE 4.11 – Série des paramètres de Pareto estimés pour la garantie Vol (hors été).

La représentation graphique des différentes étapes de la détection est présentée en annexe, le résultat de cette détection est présenté dans le tableau suivant.

Rupture	Seuil 10%		Seuil 5%		Seuil 1%	
	Date détection	Date détectée	Date détection	Date détectée	Date détection	Date détectée
1	41	28	43	28	46	28
2	74	49	77	49	82	49
3	96	87	99	87	107	87

Le modèle retenu est présenté dans le tableau suivant, et sur la figure 4.3.4.1 ci-dessous.

Tronçon de courbe	α	β
1 à 28	0.16021	-0.00017592
29 à 49	0.15579	-3.5367e-5
50 à 87	0.15604	-5.2857e-5
88 à 135	0.15552	-3.0414e-5

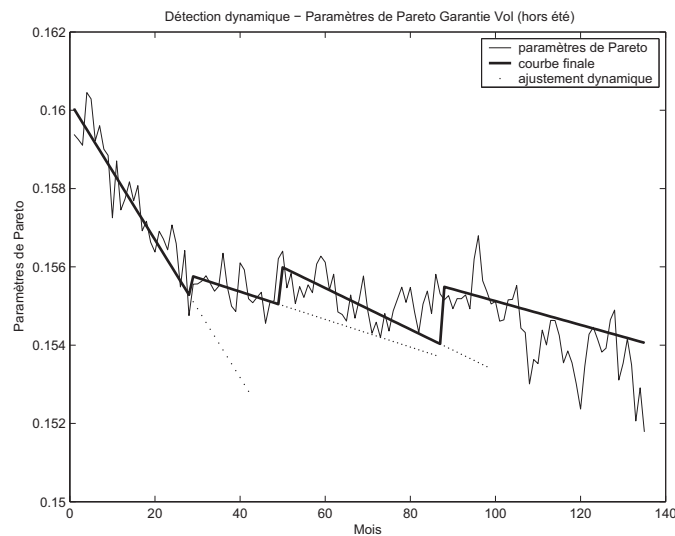


FIGURE 4.12 – Résumé de la détection dynamique de rupture pour la série des paramètres de Pareto garantie Vol (hors été). La courbe finale est obtenue à partir du modèle estimé de façon dynamique, vu de la date 135. L'ajustement dynamique matérialise le retard de détection.

Conclusions : Le délai de détection est relativement important (jusqu'à 28 périodes de temps pour la dernière rupture). Sur la dernière partie de la courbe, la tendance estimée semble être globalement supérieure à la courbe, ce qui laisse penser qu'on surévalue le paramètre de Pareto sur cette dernière partie de courbe.

4.3.4.2 Garantie Rccorp

Nous présentons la série des paramètres de Pareto estimés pour la garantie Rccorp entre janvier 1994 et décembre 2008.

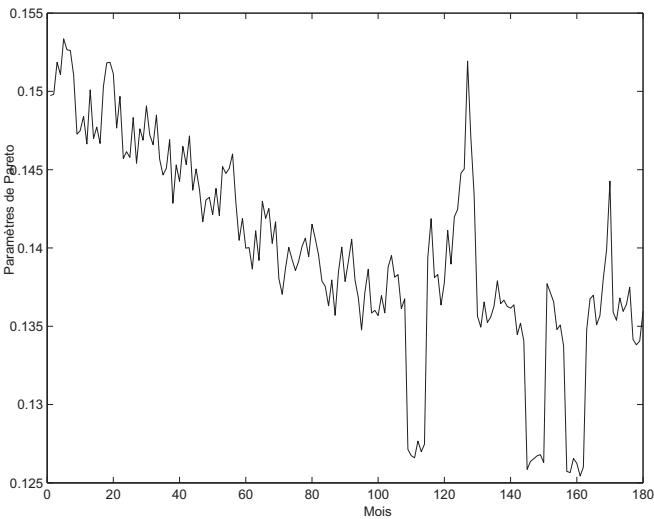


FIGURE 4.13 – Série des paramètres de Pareto estimés pour la garantie Rccorp.

La série présente des pics dans sa seconde moitié qui rendent la méthode instable. Nous présentons le résumé de la détection dynamique dans le tableau suivant en ne tenant pas compte de ces pics. En tout état de cause, la modélisation utilisée ici est peu adaptée à ce profil de séries.

Rupture	Seuil 10%		Seuil 5%		Seuil 1%	
	Date détection	Date détectée	Date détection	Date détectée	Date détection	Date détectée
1	114	109	114	109	114	109

Modèle retenu.

Tronçon de courbe	α	β
1 à 109	0.15129	-0.00014396
110 à 180	0.15185	-0.00010985

Conclusions : Le résultat ne paraît pas vraiment exploitable en raison des pics importants qui font leur apparition dans les 80 dernières observations. S’il semble y avoir rupture de comportement dans cette deuxième partie, cette rupture semble être moins une rupture en moyenne (celle que nous essayons de détecter) qu’en variance. Cette étude nécessiterait une modélisation plus poussée.

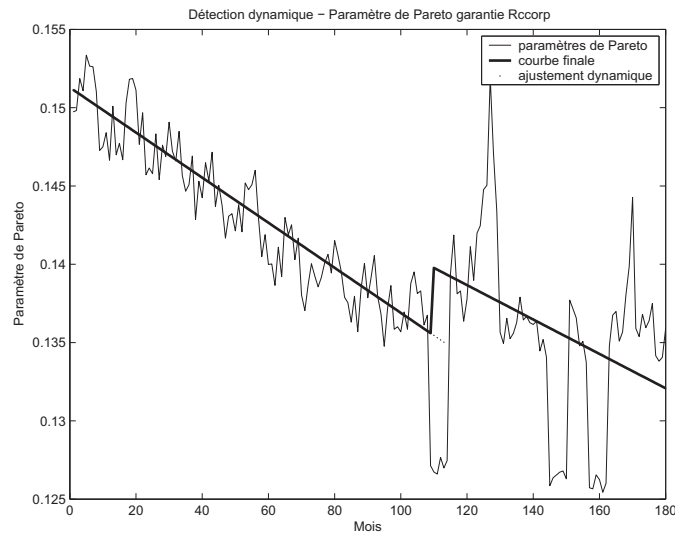


FIGURE 4.14 – Résumé de la détection dynamique de rupture pour la série des paramètres de Pareto garantie Rccorp. La courbe finale est obtenue à partir du modèle estimé de façon dynamique, vu de la date 180. L'ajustement dynamique matérialise le retard de détection.

4.3.4.3 Garantie Rcmat

Nous présentons la série des paramètres de Pareto estimés pour la garantie Rcmat entre janvier 1994 et décembre 2008.

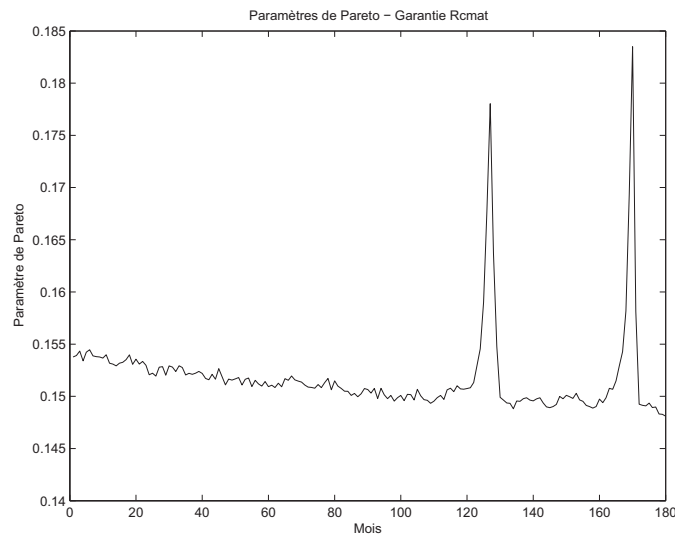


FIGURE 4.15 – Série des paramètres de Pareto estimés pour la garantie Rcmat.

Les deux pics présents sur la fin de la série sont détectés automatiquement et sont exclus de l'étude pour les mêmes raisons que précédemment.

Rupture	Seuil 10%		Seuil 5%		Seuil 1%	
	Date détection	Date détectée	Date détection	Date détectée	Date détection	Date détectée
1	91	64	91	64	92	64

Modèle retenu :

Tronçon de courbe	α	β
1 à 64	0.15418	-5.2948e-5
65 à 180	0.1517	-5.0791e-5

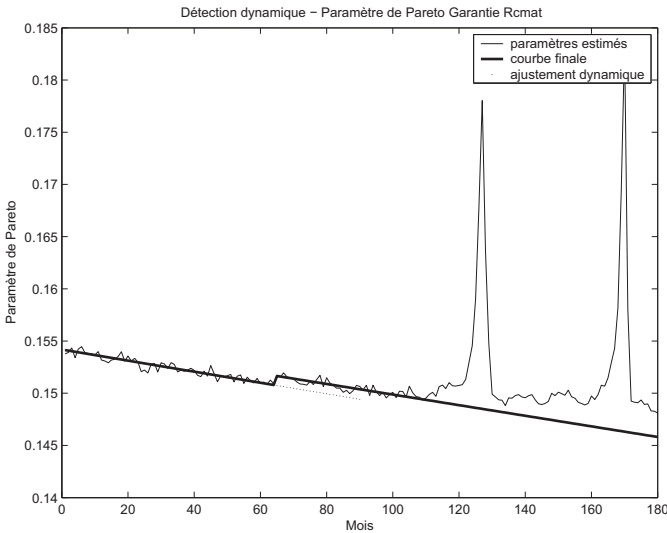


FIGURE 4.16 – Résumé de la détection dynamique de rupture pour la série des paramètres de Pareto garantie Rcmat. La courbe finale est obtenue à partir du modèle estimé de façon dynamique, vu de la date 180. L'ajustement dynamique matérialise le retard de détection.

Conclusions : Les deux pics perturbent la détection, mais de façon moins significative que dans l'exemple précédent, puisqu'il s'agit de pics peu nombreux et relativement espacés (pour la garantie Rccorp, les pics se succédaient presque sans discontinuer). En revanche, la tendance paraît sous-évaluée dans la dernière partie de la courbe. En terme de risque, ceci correspond à une sur-estimation du risque (paramètre de Pareto plus faible).

4.3.5 Problèmes posés par la modélisation Pareto

Au regard de nos données, la modélisation par une loi de Pareto est contestable essentiellement pour deux raisons.

Insuffisances de la modélisation par Pareto.

- *Mauvaise adéquation. Le test de Kolmogorov-Smirnov rejette systématiquement l'hypothèse d'adéquation à une Pareto pour n'importe quel niveau "raisonnable".*
- *Problèmes concernant l'exploitation du résultat : les paramètres estimés sont tous plus petits que 1. Si la modélisation Pareto était retenue, elle poserait un problème dans l'approche fréquence-coût, puisqu'elle aboutirait à des variables ayant une espérance infinie.*

Une façon de contourner le problème serait de proposer une modélisation par un autre type de loi paramétrique. Néanmoins, cette approche présente au moins deux inconvénients :

- Difficulté de trouver spontanément une loi permettant de modéliser convenablement les coûts de sinistres.
- En tout état de cause, même si la première difficulté était surmontée, elle déboucherait sur une loi plus compliquée (vraisemblablement avec plus de paramètres) qui serait sans doute difficilement exploitable d'un point de vue technique pour la détection de rupture.

Pour cette raison, nous proposons dans la section 4.3.6 ci-dessous une nouvelle interprétation de la technique de détection de rupture dans le modèle de Pareto, qui la rend légitime même lorsque la modélisation par une variable de Pareto n'est pas adéquate.

4.3.6 Combinaison avec la série des coûts moyens

Si l'on revient à l'expression (4.1), on voit que l'estimateur $\hat{\alpha}(i)$ possède une interprétation même dans le cas où les X_j ne suivent pas une Pareto. En effet, $\hat{\alpha}(i)^{-1}$ peut être vu comme un estimateur de $E[\log X_j]$. On peut donc interpréter cette recherche de rupture dans les paramètres $\hat{\alpha}(i)$ comme un changement d'échelle du problème : au lieu de détecter un changement de moyenne pour X_j , on le transforme en son logarithme. De ce fait, on espère écraser les queues de distributions de la variable X_j et obtenir de meilleurs résultats.

A la lumière de cette remarque, notre idée consiste alors à combiner l'approche Pareto avec l'étude de la série des coûts moyens : nous utilisons la série des paramètres de Pareto pour détecter des dates de rupture dans la loi des $\log X_j$, puis nous effectuons, sur la série des coûts moyens, une régression linéaire par morceaux, en utilisant les subdivisions déterminées par les dates de ruptures détectées.

4.3.6.1 Garantie Vol

En utilisant les dates de ruptures détectées à partir de l'étude de la série des paramètres de Pareto, le modèle retenu devient le suivant.

Tronçon de courbe	α	β
1 à 28	1265.4	1.1135
29 à 49	1303.6	-2.8779
50 à 87	1244.7	4.6872
88 à 135	1369.7	8.2233

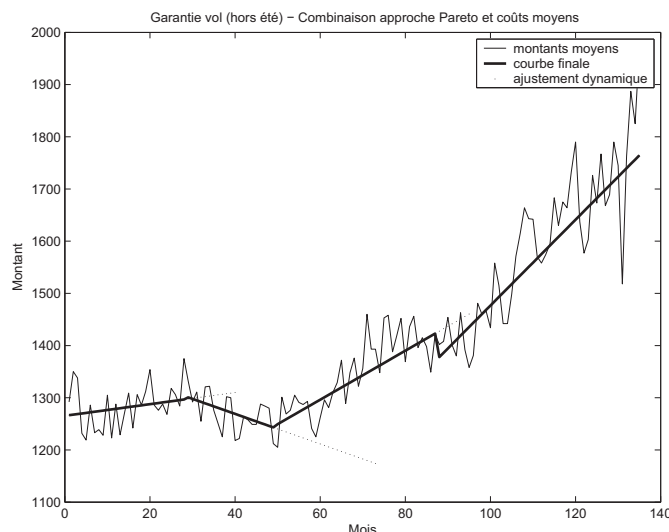


FIGURE 4.17 – Résumé de la détection dynamique de rupture pour la garantie Vol (hors été) en combinant les deux approches. La courbe finale est obtenue à partir du modèle estimé de façon dynamique, vu de la date 135. L’ajustement dynamique matérialise le retard de détection.

Conclusions : Le modèle final obtenu est plus confortable à utiliser, puisqu’il comporte moins de rupture que celui obtenu en utilisant la méthode bootstrap sur les coûts moyens (modèle plus stable). Par ailleurs, la courbe finale obtenue paraît convenablement ajustée à la courbe. En revanche, le délai de détection pose problème. Pour les ruptures 1 et 3, il aboutit à une évaluation plus prudente du risque. En revanche, pour la rupture 2, ce délai est particulièrement important et aboutit à une sous-évaluation importante du risque, ce qui n’était pas le cas avec l’approche coûts moyens et l’usage du bootstrap. Une utilisation conjointe des deux méthodes, permettrait sans doute, en confrontant les résultats obtenus, de palier à ce problème.

4.3.6.2 Garantie Rccorp

Nous procédons de façon similaire avec la garantie Rccorp.

Tronçon de courbe	α	β
1 à 109	6728.9	43.726
110 à 180	10634	19.584

Conclusions : L’ajustement obtenu paraît moins adapté que celui obtenu par la méthode bootstrap. Le nombre de rupture est certes inférieur (une au lieu de deux), mais la tendance estimée dans la deuxième partie de la courbe sous-évalue nettement le risque par rapport à l’approche bootstrap sur la série des coûts moyens.

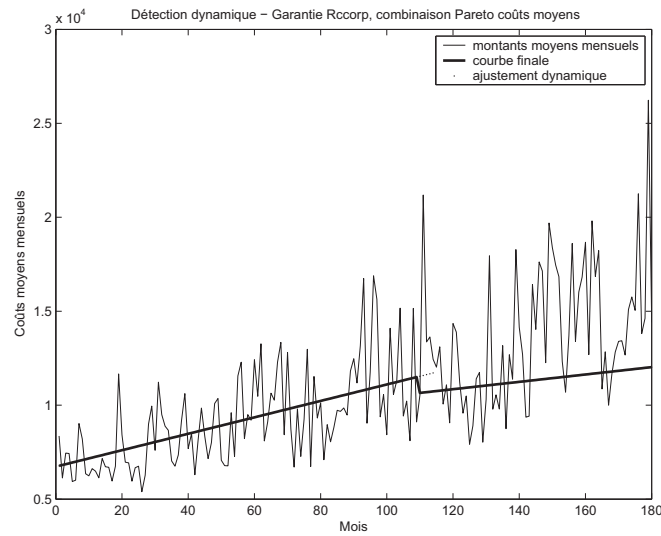


FIGURE 4.18 – Résumé de la détection dynamique de rupture pour la garantie Rccorp en combinant les deux approches. La courbe finale est obtenue à partir du modèle estimé de façon dynamique, vu de la date 180. L'ajustement dynamique matérialise le retard de détection.

4.3.6.3 Garantie Rcmat

Nous procédons de façon similaire avec la garantie Rcmat.

Tronçon de courbe	α	β
1 à 64	986.54	1.6865
65 à 180	1084.4	2.082

Conclusions : L'ajustement paraît satisfaisant dans la première partie de courbe, avec un retard de détection assez faible. Le nombre de ruptures détecté est bien moins important que pour la première méthode. En revanche, l'ajustement est mauvais sur la dernière partie de courbe, avec une tendance à la surestimation du risque. Ce mauvais comportement était attendu, puisqu'il avait déjà été identifié sur la série des paramètres de Pareto.

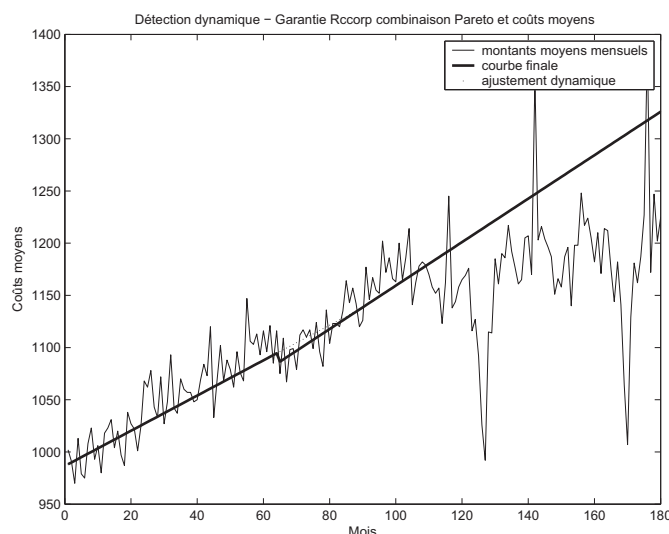


FIGURE 4.19 – Résumé de la détection dynamique de rupture pour la garantie Rcmat en combinant les deux approches. La courbe finale est obtenu à partir du modèle estimé de façon dynamique, vu de la date 180. L'ajustement dynamique matérialise le retard de détection.

4.3.7 Comparaison avec l'approche bootstrap sur les coûts moyens

D'après l'étude de ces trois séries, il semble que la méthode basée sur l'estimation des paramètres de Pareto débouche sur une détection d'un nombre plus faible de ruptures. S'il est adapté, le modèle obtenu est donc plus stable à long terme, nécessite moins de calibrations et donc d'erreurs. Néanmoins, pour les séries Rccorp et Rcmat, le modèle obtenu ne paraît pas satisfaisant. On peut expliquer ce mauvais comportement notamment par la présence de pics dans les séries des paramètres de Pareto associées à ces deux garanties.

4.4 Approche mixte

4.4.1 Intérêt potentiel d'une séparation entre "gros" sinistres et "petits" sinistres

En raison de propriétés issues de la théorie des extrêmes, on s'attend à ce que la loi de Pareto soit surtout adaptée à la description des sinistres les plus coûteux. Dans cette optique, il est possible de combiner une approche double, en séparant petits et gros sinistres. Les petits sinistres sont ensuite étudiés par la méthode de la section 4.2, tandis que les gros sinistres sont étudiés en utilisant le modèle de Pareto.

4.4.2 Applications aux différentes garanties

Pour chacune des garanties, nous établissons des "franchises" artificielles, qui sont construites pour qu'une proportion d'environ 10% de ces sinistres soit supérieure à cette

franchise. Nous précisons que cette franchise est uniquement introduite dans le but d'analyser la distribution des coûts de sinistres, en séparant les sinistres les plus petits de l'analyse de la queue de distribution. Elle n'a aucune interprétation commerciale ou opérationnelle.

4.4.2.1 Garantie Vol

Nous avons fixé le montant de la franchise à 2900 euros. Nous présentons en parallèle la série des coûts moyens et la série des paramètres de Pareto estimés sur les sinistres supérieurs à la franchise. Pour les plus petits sinistres, nous notons la même saisonnalité que celle observée précédemment. En revanche, cette saisonnalité est absente de la série correspondant aux sinistres supérieurs à la franchise.

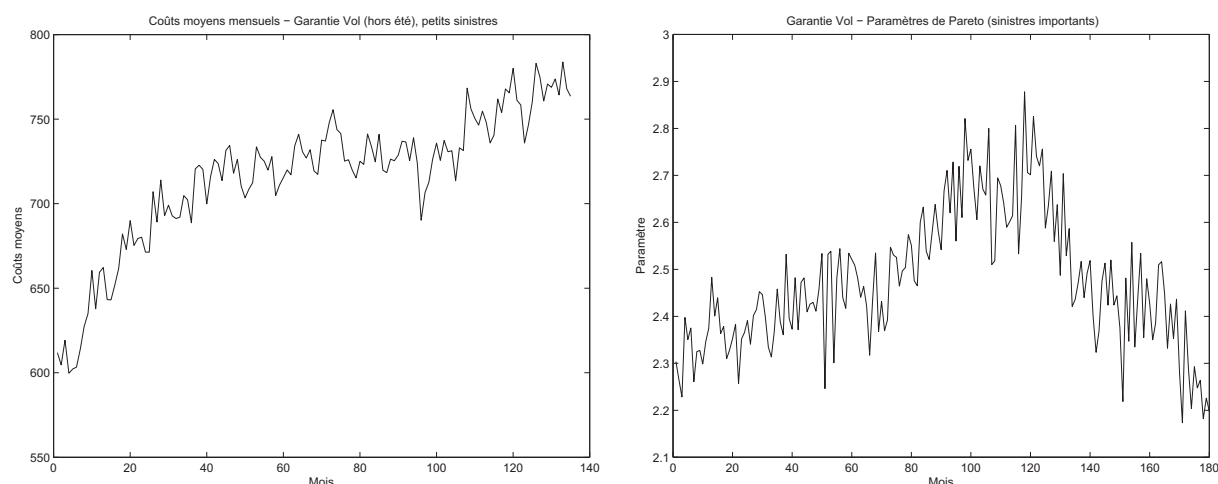


FIGURE 4.20 – Coûts moyens mensuels pour les sinistres inférieurs à la franchise pour la garantie Vol (hors été), et paramètres de Pareto estimés (été compris) pour les sinistres supérieurs à la franchise.

Sinistres faibles.

Le résumé de la détection dans le cas des petits sinistres est présenté dans le tableau suivant.

Rupture	Seuil 10%		Seuil 5%		Seuil 1%	
	Date détection	Date détectée	Date détection	Date détectée	Date détection	Date détectée
1	25	19	25	19	29	19
2	50	44	50	44	51	44
3	73	69	73	69	73	69
4	92	76	94	76	102	95
5	124	119	124	119	129	119

On remarque d'emblée des différences importantes entre les différents seuils de test, notamment pour la rupture 4. Nous gardons la date 95 plus que la date 76 en raison du graphique suivant, qui présente le comportement de la vraisemblance au voisinage de la détection.

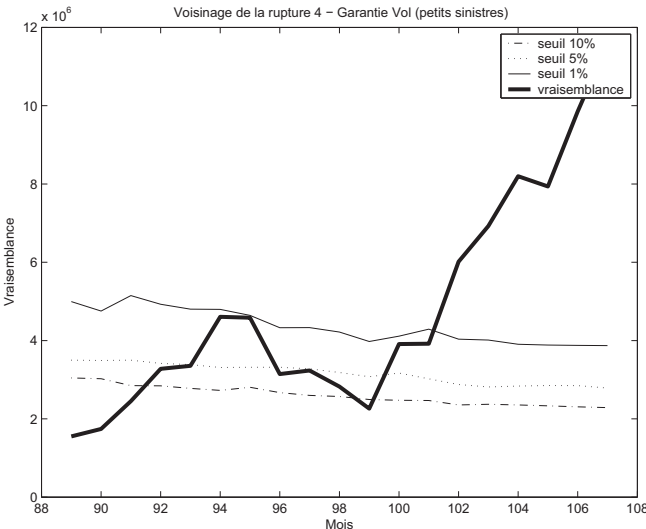


FIGURE 4.21 – Détection dynamique au voisinage de la rupture 4, petits sinistres de la garantie vol.

La détection aux seuils 10 et 5 % intervient aux dates 92 et 94. La vraisemblance diminue ensuite pour repasser assez rapidement sous les seuils. En revanche, lors de la détection à la date 102 (seuil 1%), la vraisemblance augmente très fortement (tendance qui se poursuit par la suite). A la lumière de cette courbe, nous retenons donc plutôt la date 95.

Le modèle retenu correspondant est présenté dans le tableau suivant.

Tronçon de courbe	α	β
1 à 19	593.92	4.3052
20 à 44	674.74	1.9531
45 à 69	719.43	0.16036
70 à 95	742.09	-1.0135
96 à 119	709.8	2.2815
120 à 135	758.87	0.82194

Sinistres importants.

Le résumé de la détection dans le cas des paramètres de Pareto pour les sinistres importants est présenté dans le tableau suivant.

Rupture	Seuil 10%		Seuil 5%		Seuil 1%	
	Date détection	Date détectée	Date détection	Date détectée	Date détection	Date détectée
1	93	65	99	65	>120	*
2	100	97	100	97	100	97

Dans le cas de la première rupture, si on utilise le seuil 1%, on met beaucoup de temps à détecter. Le modèle retenu pour la courbe des paramètres de Pareto est représenté dans le tableau suivant.

Tronçon de courbe	α	β
1 à 65	2.3007	0.0042556
66 à 97	2.3832	0.0096689
98 à 180	2.7287	-0.0061694

Nous représentons graphiquement les différents modèles obtenus.

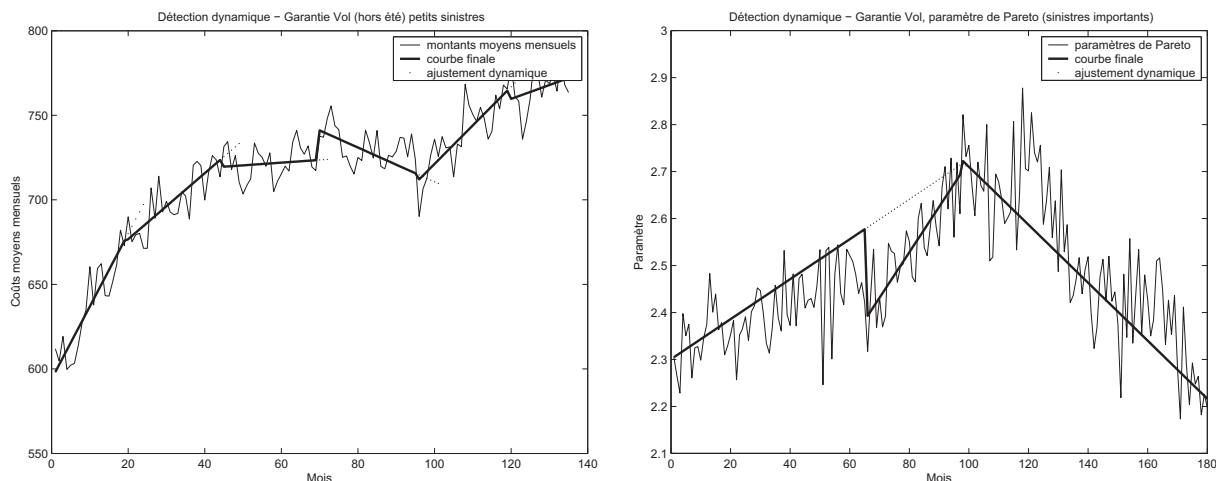


FIGURE 4.22 – Résumé de la détection dynamique pour la Garantie vol en séparant grands et petits sinistres.

Coûts moyens pour les gros sinistres.

Concernant les gros sinistres, on peut également considérer la courbe des coûts moyens correspondant en utilisant les dates de ruptures estimées pour le paramètre de Pareto. Le modèle retenu est présenté dans le tableau suivant.

Tronçon de courbe	α	β
1 à 65	4635.3	-3.6938
66 à 97	4556.1	-7.3187
98 à 180	4317.2	4.7941

Conclusions : L'absence de saisonnalité dans la série des gros sinistres, contrairement à celle des petits, est une justification supplémentaire qui plaide pour la séparation des sinistres en plusieurs catégories. On note relativement peu de ruptures pour la modélisation des sinistres les plus importants. Néanmoins, on détecte à nouveau un délai de détection important (près de 3 ans entre la première rupture et sa détection) qui conduit à une sous-évaluation du risque. En revanche, la série des petits sinistres semble modélisée de façon plus satisfaisante. Le délai de détection est relativement faible (5 à 6 unités de temps, sauf pour la dernière rupture, mais cette détection tardive aboutit à une évaluation plus prudente), on n'enregistre plus que 5 ruptures au lieu de 6 sur la série enregistrant tous les sinistres.

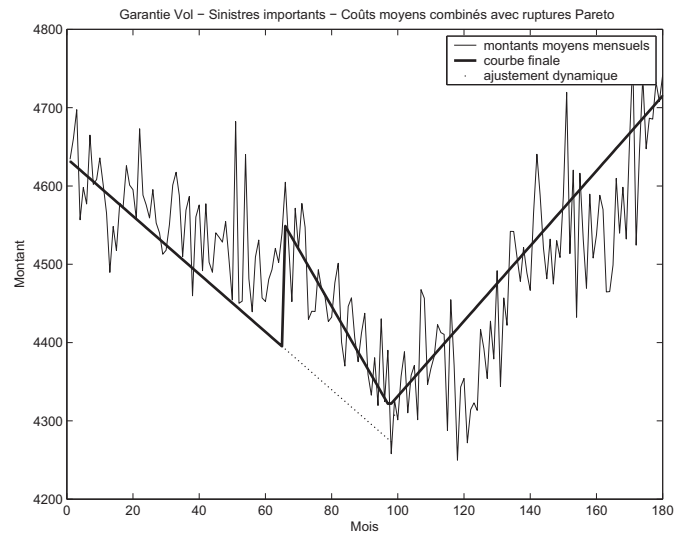


FIGURE 4.23 – Garantie Vol. Détection dynamique sur sinistres importants en combinant coûts moyens et modèle de Pareto.

4.4.2.2 Garantie Rccorp

Nous avons fixé le montant de la franchise à 13100. Nous présentons en parallèle la série des coûts moyens et la série des paramètres de Pareto estimés sur les sinistres supérieurs à la franchise.

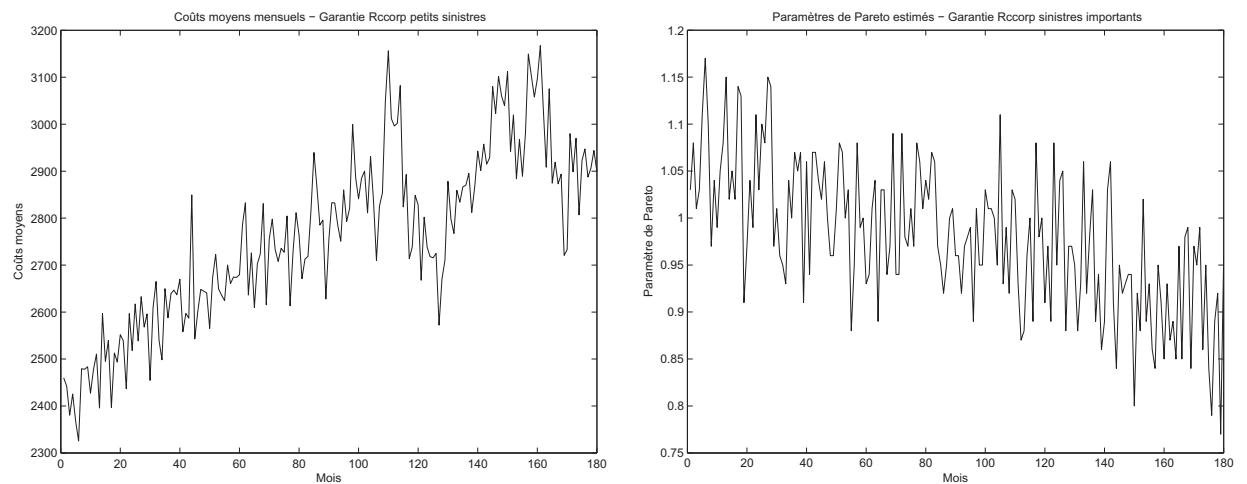


FIGURE 4.24 – Coûts moyens mensuels pour les sinistres inférieurs à la franchise pour la garantie Rccorp, et paramètres de Pareto estimés pour les sinistres importants.

Sinistres faibles.

Le résumé de la détection dans le cas des petits sinistres est présenté dans le tableau suivant.

Rupture	Seuil 10%		Seuil 5%		Seuil 1%	
	Date détection	Date détectée	Date détection	Date détectée	Date détection	Date détectée
1	55	44	57	44	60	44
2	75	62	75	62	77	62
3	89	70	89	84	102	83
4	110	105	110	105	110	105
5	137	126	138	127	139	127
6	152	144	153	150	153	150

On remarque, pour la sixième rupture, un grand écart entre les estimateurs de la date suivant le seuil 10% et les deux autres seuils. Ceci n'est pas surprenant : la date de détection 152 est trop proche de la date de rupture constatée par la suite (150) pour détecter avec précision. Le modèle retenu est présenté dans le tableau suivant.

Tronçon de courbe	α	β
1 à 44	2397.4	6.151
45 à 62	2587.6	7.4566
63 à 84	2694.2	2.4955
85 à 105	2783.2	4.9213
106 à 126	2973	-10.299
127 à 144	2714.1	16.168
145 à 180	3075.4	-10.551

Sinistres importants.

Le résumé de la détection dans le cas des paramètres de Pareto pour les sinistres importants est présenté dans le tableau suivant.

Rupture	Seuil 10%		Seuil 5%		Seuil 1%	
	Date détection	Date détectée	Date détection	Date détectée	Date détection	Date détectée
1	70	41	72	41	79	41
2	89	78	98	72	103	72
3	118	99	119	99	124	99
4	162	117	165	117	168	117

Le modèle retenu pour la courbe des paramètres de Pareto est représenté dans le tableau suivant.

Tronçon de courbe	α	β
1 à 41	1.0711	-0.0017669
42 à 72	1.0274	-0.00179446
73 à 99	1.0271	-0.0028538
100 à 117	1.0047	-0.0023985
118 à 180	0.97776	-0.0016014

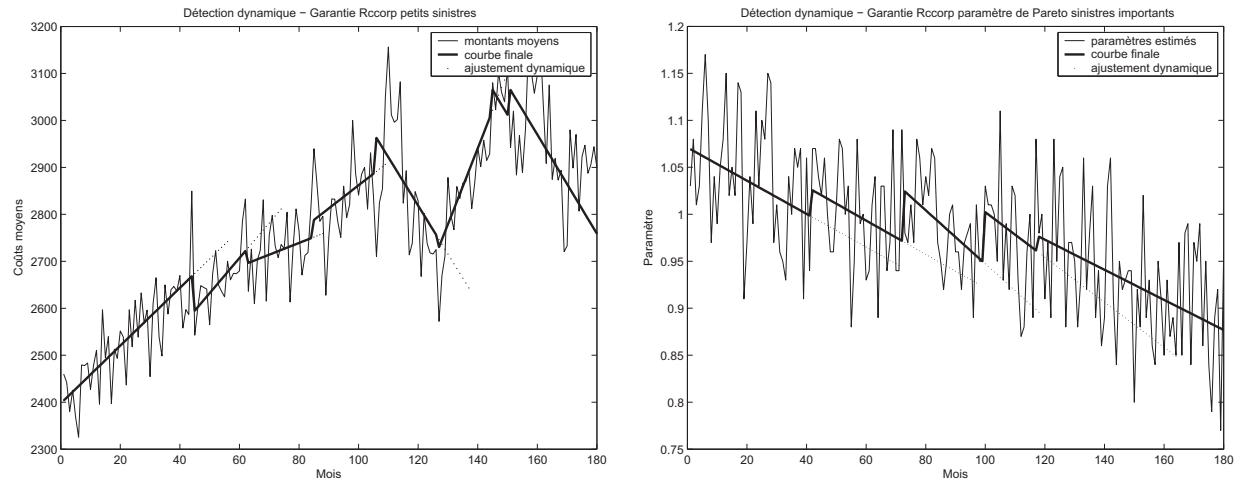


FIGURE 4.25 – Résumé de la détection dynamique pour la Garantie Rccorp en séparant grands et petits sinistres.

Nous représentons graphiquement les différents modèles obtenus.

Coûts moyens pour les gros sinistres.

Concernant les gros sinistres, on peut également considérer la courbe des coûts moyens correspondant en utilisant les dates de ruptures estimées pour le paramètre de Pareto. Le modèle retenu est présenté dans le tableau suivant.

Tronçon de courbe	α	β
1 à 41	65979	56.306
42 à 72	57260	1887.3
73 à 99	69391	1035.3
100 à 117	87455	342.3
118 à 180	88826	493.77

Conclusions : La détection dynamique sur les sinistres importants est peu concluante, avec des délais de détection particulièrement importants, qui conduisent aussi bien à une sous-estimation qu'à une surestimation. La trop forte volatilité de la courbe semble être à l'origine de ce mauvais comportement.

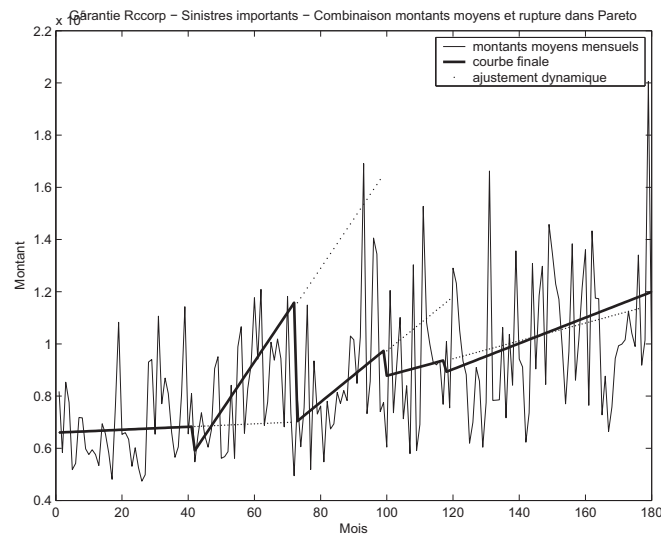


FIGURE 4.26 – Garantie Rccorp. Détection dynamique sur sinistres importants en combinant coûts moyens et modèle de Pareto.

4.4.2.3 Garantie Rcmat

Nous avons fixé le montant de la franchise à 2200. Nous présentons en parallèle la série des coûts moyens et la série des paramètres de Pareto estimés sur les sinistres supérieurs à la franchise.

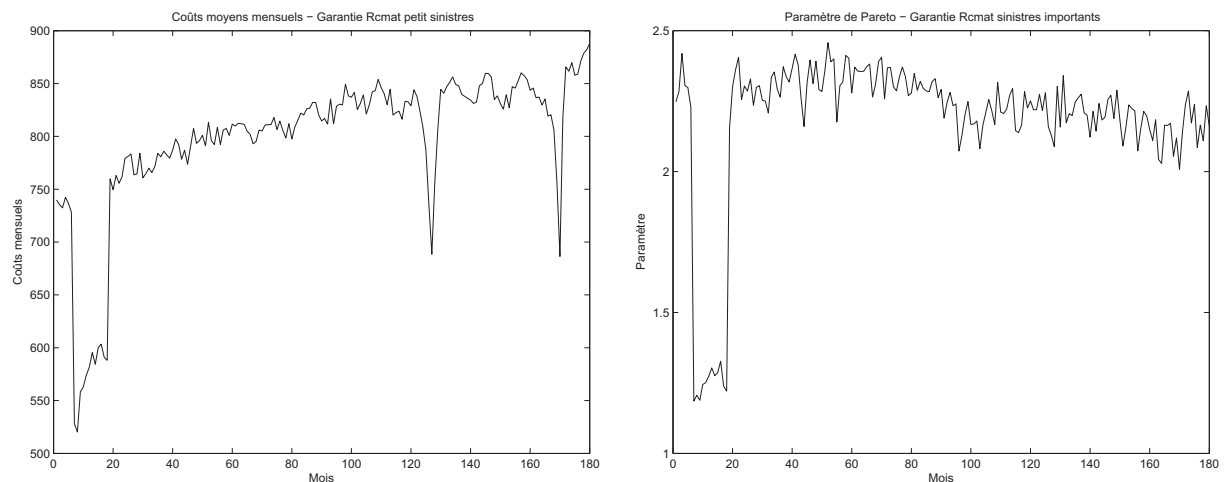


FIGURE 4.27 – Coûts moyens mensuels pour les sinistres inférieurs à la franchise pour la garantie Rcmat, et paramètres de Pareto estimés pour les sinistres importants.

Sinistres faibles.

Le résumé de la détection dans le cas des petits sinistres est présenté dans le tableau suivant.

Rupture	Seuil 10%		Seuil 5%		Seuil 1%	
	Date détection	Date détectée	Date détection	Date détectée	Date détection	Date détectée
1	7	6	7	6	7	6
2	19	18	19	18	19	18
3	73	65	74	65	76	65
4	114	108	115	108	115	108
5	128	127	128	127	128	127
6	171	170	171	170	171	170

Les deux premières ruptures correspondent en fait à une baisse immédiatement identifiable. Les deux dernières sont en fait des pics vers le bas qui sont situées après des pics instantanément identifiés.

Tronçon de courbe	α	β
1 à 6	740.27	-1.2689
7 à 18	532.07	6.4455
19 à 65	758.66	1.1334
66 à 108	798.86	0.96144
109 à 127	846.69	-1.9514
128 à 170	817.63	2.0515
171 à 180	837.14	5.0888

Sinistres importants.

Le résumé de la détection dans le cas des paramètres de Pareto pour les sinistres importants est présenté dans le tableau suivant.

Rupture	Seuil 10%		Seuil 5%		Seuil 1%	
	Date détection	Date détectée	Date détection	Date détectée	Date détection	Date détectée
1	7	6	7	6	7	6
2	19	18	19	18	19	18
3	86	78	86	78	86	78
4	123	115	123	115	123	115

Les deux premières détections sont donc les changements brutaux. Le modèle retenu pour la courbe des paramètres de Pareto est représenté dans le tableau suivant.

Tronçon de courbe	α	β
1 à 6	2.3133	0.0045974
7 à 18	1.1647	0.0068433
19 à 78	2.2774	0.0016675
79 à 115	2.3417	-0.0079563
116 à 180	2.2098	0.000345

Nous représentons graphiquement les différents modèles obtenus.

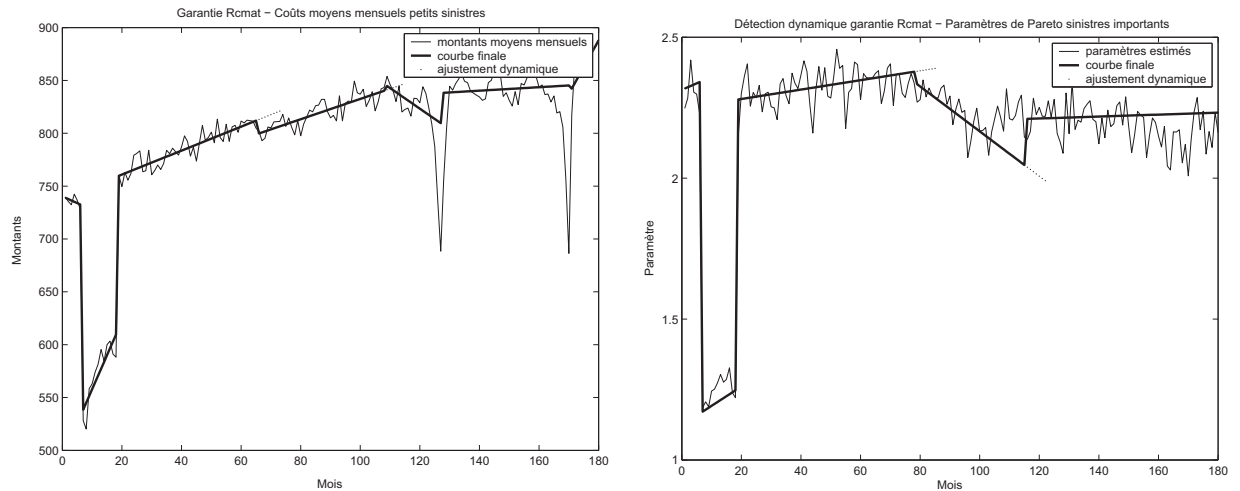


FIGURE 4.28 – Résumé de la détection dynamique pour la Garantie Rccorp en séparant grands et petits sinistres.

Coûts moyens pour les gros sinistres.

Concernant les gros sinistres, on peut également considérer la courbe des coûts moyens correspondant en utilisant les dates de ruptures estimées pour le paramètre de Pareto. Le modèle retenu est présenté dans le graphique et le tableau suivants.

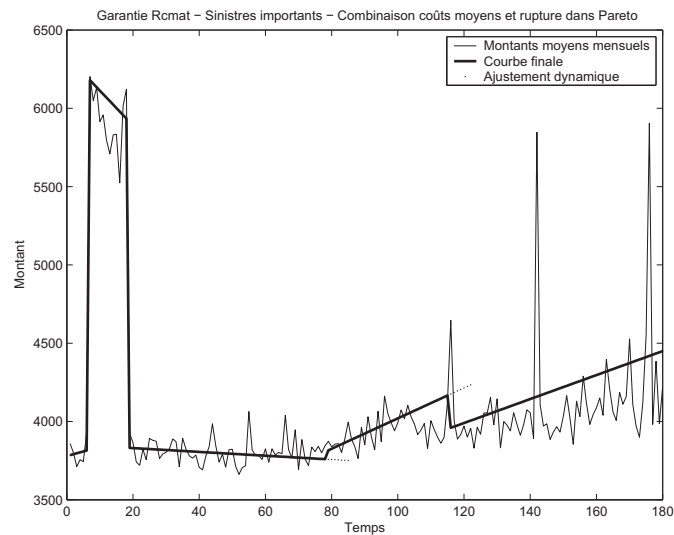


FIGURE 4.29 – Garantie Rccmat. Détection dynamique sur sinistres importants en combinant coûts moyens et modèle de Pareto.

Tronçon de courbe	α	β
1 à 6	3779.3	5.8383
7 à 18	6068	-22.197
19 à 78	3832.4	-1.2174
79 à 115	3806.1	9.7207
116 à 180	3951.8	7.6546

Conclusions : Le modèle retenu semble représenter assez fidèlement la tendance des différentes courbes, avec des délais de détection relativement faibles. La présence de pics perturbe par endroits la procédure. Par exemple, sur la figure 4.4.2.3, on semble surestimer la tendance. On observe cependant que le modèle retenu a tendance à légèrement surestimer le coût moyen, ce qui rend l’approche prudente.

4.5 Conclusion

Si l’on s’en tient à une vision graphique, les courbes ajustées en utilisant la méthode mixte de la section 4.4 semblent mieux représenter l’évolution de nos données. D’un point de vue statistique, une explication de ce phénomène vient sans doute du fait que les queues de distribution des résidus de nos séries sont lourdes, et de ce fait peuvent dégrader considérablement l’estimation en occultant ce qui se passe dans les autres parties de la distribution. De ce point de vue, segmenter la distribution en plusieurs parties semble une bonne stratégie. Cette intuition sera confortée par les résultats concernant le provisionnement qui sont présentés au chapitre suivant.

Précisons que la façon dont nous avons segmenté notre distribution des coûts de sinistres est un peu fruste, et pourrait sans doute être améliorée. Nous avons établi des franchises artificielles de sorte qu’environ 10% des sinistres soient supérieurs à ces franchises. Il serait sans doute plus avisé (et plus délicat) de déterminer, à partir des données, un seuil de séparation entre petits et gros sinistres. Une perspective intéressante nous semblerait être d’adapter une technique présentée par Grama et Spokoiny (2008). La technique proposée par ces auteurs consiste à identifier, à partir de la distribution empirique des coûts de sinistres, la partie de la distribution sur laquelle on peut considérer que les données suivent approximativement une loi de Pareto. Si l’utilisation de cette technique était couronnée de succès, les sinistres importants pourraient être réellement modélisés par une variable de Pareto (ce qui n’est pas le cas de nos observations qui rejettent nettement le test de Kolmogorov-Smirnov). La série des paramètres de Pareto estimés aurait donc plus de valeur dans ce cadre.

Chapitre 5

Conséquences sur la mise en œuvre opérationnelle

Dans les chapitres 3 et 4, nous avons modélisé la fréquence de sinistres et leurs coûts. Ce dernier chapitre consiste à confronter nos estimations de charges sinistres aux données réelles et aux autres méthodes d'estimation plus courantes. Le but est à la fois de valider la technique en la confrontant à des exercices passés (où la sinistralité est mieux connue), et d'essayer d'apporter une réponse à l'une des questions initiales concernant les changements observés sur le marché au cours de l'année 2009.

Dans la section 5.1, nous présentons les prévisions de charges déterminées à partir de nos modèles de Fréquence/coût, qui s'ajustent ou non en fonction des techniques de détection de rupture proposées. Nous comparerons ensuite ces résultats aux données réelles constatées sur les sinistralités antérieures à 2009 afin de valider ou d'invalidier nos méthodes. Toutefois, les données de sinistres contenues dans la base étant une vision arrêtée au 31 décembre 2009, le niveau d'information est d'autant plus complet et stabilisé que les exercices sont anciens et que les branches sont à développement court. Ainsi, nous pouvons nous attendre à ce que nos écarts de prédictions soient plus importants sur les exercices anciens.

Dans la section 5.2, nous nous concentrons sur la prévision de charge de l'exercice 2009, telle qu'elle aurait été conduite avec les informations disponibles au 31 décembre 2008. En particulier, nous comparons cette prédiction à la vision fournie au 31 décembre 2009 (donc avec une année supplémentaire de recul) par notre base de triangles de règlements. Le but est de diagnostiquer un éventuel écart significatif entre la prévision et l'exercice 2009, qui semble marquer un renversement de tendance du marché (fin de la baisse).

La section 5.3 présente un compte-rendu de nos échanges avec les actuaires de la MACIF afin de mieux définir la façon dont les outils proposés par ce mémoire peuvent être intégrés mises à profit. On s'intéressera d'une part à des perspectives d'application immédiates (reporting, comparaison à la tendance du marché, construction de résultats intermédiaires...), d'autre part à des applications plus éloignées de la problématique initiale mais où le recul apporté par les techniques de détection de rupture pourrait s'avérer un outil précieux.

5.1 Comparaison des prédictions de charge par rapport aux sinistres réalisés

Dans cette section, nous rappelons tout d'abord comment sont construites nos différentes prévisions (section 5.1.1). En particulier, nous souhaitons distinguer les méthodes basées sur les techniques de détection de rupture, et celles qui n'en tiennent pas compte. La section 5.1.2 évalue la qualité des prévisions en la comparant aux données historiques, garantie par garantie.

5.1.1 Construction des prévisions de charge

Notre modèle fréquence/coût n'est pas basé sur la même échelle de temps pour les coûts de sinistres que pour leur nombre. Du point de vue de la fréquence, l'année est découpée en quatre parties de longueur inégale, tandis que les coûts de sinistres sont étudiés mensuellement. Pour passer des modèles que nous avons décrits aux prévisions, il faut donc procéder en trois temps.

1. Définir une prédiction du nombre de sinistres mois par mois (en utilisant le modèle à détection de rupture ou non).
2. Définir une prédiction du coût moyen d'un sinistre pour chaque mois (en utilisant les différents modèles du chapitre 4).
3. Combiner les deux pour obtenir une prévision de la charge mensuelle.

5.1.1.1 Nombre de sinistres

Rappelons (section 3.3.2) que nous avons deux propositions pour prédire le nombre de sinistres, l'une basée sur la détection de rupture qui est censée intégrer les évolutions du risque, l'autre basée sur l'utilisation de l'ensemble des données historiques sans distinction depuis 1994. Par la suite, nous appellerons Méthode 1 (M1) la méthode basée sur la détection de rupture, et Méthode 2 (M2) l'autre méthode.

Nous explicitons ici comment passer de la donnée des λ_i (section 3.3.2) et de la donnée des coefficients $\alpha_{i,g}$ (qui reflètent la saisonnalité à l'intérieur d'une année, section 3.3.1). Nous rappelons que nous supposons le montant de primes acquises l'année suivante connu, ce qui n'est pas le cas en pratique et doit être remplacé par une prévision de ce montant (ou, comme nous l'avons déjà indiqué, une prévision du nombre d'assurés dans le portefeuille, information qui semble plus pertinente).

On rappelle (voir section 3.3.4) que

$$\lambda_i \pi_i = \tilde{\lambda}_i \sum_{k=1}^4 \alpha_k n_k,$$

où n_k désigne le nombre de jours de la période k (de l'année i), π_i les primes acquises de l'année i , et $\tilde{\lambda}_i$ l'intensité du processus de Poisson sur la période de référence, d'où l'on peut aisément évaluer $\tilde{\lambda}_i$. A partir de cette donnée, on peut prédire un nombre moyen de

sinistre par jour, $\hat{N}_j = \tilde{\lambda}_i \alpha_k$ pour un jour j qui se trouve dans la période k de l'année. On agrège cette information au niveau mensuel pour se ramener à un nombre moyen de sinistres par mois.

5.1.1.2 Coûts moyen de sinistre

Nous considérons, quand c'est possible, 4 méthodes (deux d'entre elles étant très similaires) pour estimer le coût moyen d'un sinistre.

1. *Méthode 1 (M1) : Utilisation de la vraisemblance gaussienne sur la série des coûts moyens avec méthode bootstrap (section 4.2)*
2. *Méthode 2 (M2) : Utilisation de la série des paramètres de Pareto estimés pour déterminer des dates de rupture, puis retour sur la série des coûts moyens (section 4.3.6)*
3. *Méthode 3 (M3) : Séparation entre sinistres faibles et sinistres importants (supérieurs à un seuil F). Les sinistres faibles sont étudiés avec la Méthode M1, les sinistres importants sont modélisés par une loi de Pareto dont l'espérance est $aF/(a-1)$ où a est le paramètre de Pareto estimé.*
4. *Méthode 3 bis (M3bis) : Séparation entre sinistres faibles et sinistres importants, avec utilisation de la méthode M1 pour les sinistres faibles, et de la méthode M2 pour les sinistres importants.*

La méthode M3 n'est pas systématiquement applicable : lorsque le paramètre de la loi de Pareto est plus petit que 1, l'espérance de cette variable n'est pas définie. Sur les séries que nous avons modélisées, nous avons remarqués que tous les paramètres estimés étaient plus petits que 1 lorsque l'on considère l'ensemble des sinistres. Sur la garantie RC Corporels, on remarque le même phénomène, on se contentera donc de la méthode M3bis.

5.1.2 Comparaison des résultats sur l'historique

Nous combinons à présent les différentes méthodes, en distinguant les résultats entre les différentes garanties. Dans les tableaux qui suivent, la méthode M_i - M_j désigne la combinaison de la méthode M_i ($i = 1, 2$) sur les nombres de sinistres, et de la méthode M_j ($j = 1, 2, 3, 3bis$) sur les coûts.

5.1.2.1 Garantie Vol

Après consolidation des différentes méthodes de modélisation de Fréquence/coûts proposées, les prévisions de charge qui en découlent sont présentées dans le tableau de synthèse ci-dessous.

Les méthodes de prévisions offrent une vision de la charge généralement prudente. Toutefois, elles présentent sur certaines années des écarts significatifs, avec des niveaux de charges relativement élevés au regard des données réelles (notamment 2003-2006). Ces écarts sont d'autant plus significatifs qu'ils ne peuvent pas être imputés à un manque d'information (IBNR) puisque la garantie vol est une branche à développement court.

Année	Charge Réelle - Source base sinistres	Prévision de charges			
		Fréq : Méthode 1 / CM : Méthode 1	Fréq : Méthode 1 / CM : Méthode 2	Fréq : Méthode 1 / CM : Méthode 3	Fréq : Méthode 1 / CM : Méthode 3 bis
1998	125 423 995	147 802 644	148 355 699	133 035 409	129 896 850
1999	122 467 554	142 599 187	142 872 766	132 520 783	129 531 599
2000	122 046 322	150 498 572	143 512 690	134 992 528	132 112 012
2001	136 527 361	156 957 545	142 796 402	136 077 755	133 244 498
2002	133 801 062	170 005 539	156 795 871	141 961 727	139 371 831
2003	119 647 950	173 203 271	172 393 346	145 294 473	142 671 522
2004	106 001 234	154 147 580	158 172 624	130 486 888	127 909 436
2005	101 136 392	158 991 849	158 968 039	129 800 207	127 002 911
2006	94 644 909	158 780 744	155 682 907	124 867 163	121 901 444
2007	87 890 157	110 484 313	108 842 373	85 703 504	83 439 240
2008	79 440 835	110 867 882	108 838 771	83 264 867	80 729 842

Année	Charge Réelle - Source base sinistres	Prévision de charges			
		Fréq : Méthode 2 / CM : Méthode 1	Fréq : Méthode 2 / CM : Méthode 2	Fréq : Méthode 2 / CM : Méthode 3	Fréq : Méthode 2 / CM : Méthode 3 bis
1998	125 423 995	147 802 644	148 355 699	133 035 409	129 896 850
1999	122 467 554	142 599 187	142 872 766	132 520 783	129 531 599
2000	122 046 322	150 498 572	143 512 690	134 992 528	132 112 012
2001	136 527 361	156 957 545	142 796 402	136 077 755	133 244 498
2002	133 801 062	170 005 539	156 795 871	141 961 727	139 371 831
2003	119 647 950	173 203 271	172 393 346	145 294 473	142 671 522
2004	106 001 234	154 147 580	158 172 624	130 486 888	127 909 436
2005	101 136 392	158 991 849	158 968 039	129 800 207	127 002 911
2006	94 644 909	158 780 744	155 682 907	124 867 163	121 901 444
2007	87 890 157	154 221 685	151 929 750	119 630 908	116 470 292
2008	79 440 835	156 125 659	153 268 237	117 254 717	113 684 861

Par ailleurs, si on s'intéresse à une borne de confiance sur la prévision (cette question sera discutée plus avant dans la section 5.2), on se rend compte que sur ces années, la charge réalisée est largement à l'extérieur de l'intervalle de confiance fourni par les modèles (sur la garantie Vol, on constate une taille de l'intervalle de confiance à 99% centré autour de la prédiction de l'ordre de 6 millions d'euros). En revanche, les méthodes M1-M3 et M1-M3bis semblent donner de bonnes prévisions sur les derniers exercices (10 % au-dessus de réel, la charge de sinistre rentrant dans l'intervalle de confiance prévu).

Ce résultat n'est pas surprenant, puisqu'il confirme les résultats qui semblaient apparaître à l'issue des chapitres 3 et 4 : on gagne à séparer les sinistres faibles des sinistres graves pour éviter les effets levier causés par la présence d'un sinistre important. Par ailleurs, nous avons d'ores et déjà constaté une mauvaise prévision du nombre de sinistres sur les années 2003-2006. Si nous avons pu réduire le délai de détection de rupture (rupture estimée à 2002, détectée 4 ans plus tard), nous aurions pu réduire considérablement cet écart.

Des pistes existent pour combler cet écart. Nous avons remarqué que, dans la procédure de détection de rupture que nous mettons en place (3.3.2) les estimateurs de la date de rupture changent brutalement puis se stabilisent rapidement autour de 2002, ce qui peut représenter un signal d'alerte. Par ailleurs, le fait que les prévisions de charges ne rentrent

pas dans l'intervalle de confiance de prévision est en soi une alerte qui doit inviter à réviser le modèle. Nous aurons l'occasion de discuter ce point plus en détail en s'attardant sur le cas de l'année 2009.

5.1.2.2 Garantie RC Corporels

Nous présentons à nouveau les résultats sous forme de tableau, en rappelant que la Méthode M3 pour les coûts de sinistres ne peut être utilisée du fait de paramètres de Pareto estimés plus petits que 1.

Année	Charge Réelle - Source base sinistres	Prévision de charges					
		Fréq : Méthode 1 / CM : Méthode 1	Fréq : Méthode 1 / CM : Méthode 2	Fréq : Méthode 1 / CM : Méthode 3 bis	Fréq : Méthode 2 / CM : Méthode 1	Fréq : Méthode 2 / CM : Méthode 2	Fréq : Méthode 2 / CM : Méthode 3 bis
1998	348 913 433	332 486 300	332 486 300	340 800 565	332 486 300	332 486 300	340 800 565
1999	397 502 682	363 360 323	363 360 323	371 634 840	363 360 323	363 360 323	371 634 840
2000	335 684 069	400 696 439	400 696 439	603 224 810	400 696 439	400 696 439	603 224 810
2001	410 642 246	400 527 483	423 629 047	698 021 941	400 527 483	423 629 047	698 021 941
2002	361 562 638	472 200 898	449 536 062	530 757 465	472 200 898	449 536 062	530 757 465
2003	340 349 090	500 964 067	442 826 046	539 321 198	500 964 067	442 826 046	539 321 198
2004	297 434 997	500 600 666	422 343 353	468 015 808	500 600 666	422 343 353	468 015 808
2005	329 131 083	494 482 366	411 619 896	464 385 433	494 482 366	411 619 896	464 385 433
2006	399 827 045	479 340 856	394 015 915	456 295 650	479 340 856	394 015 915	456 295 650
2007	375 535 406	468 075 917	380 189 049	444 672 733	468 075 917	380 189 049	444 672 733
2008	379 966 999	376 763 533	302 528 160	362 026 355	475 199 609	381 568 944	456 612 085

Sur la garantie RC Corporels et sur l'ensemble des méthodes, la largeur de l'intervalle de confiance à 99 % est de l'ordre de 107 millions d'euros. Cette amplitude, bien plus importante que sur la garantie Vol, était attendue, du fait de la forte variance des coûts des sinistres corporels.

Sur l'exercice 2008, ce sont les méthodes basées sur la détection de rupture dans le nombre de sinistres qui donnent les meilleurs résultats (les autres méthodes sur-évaluent très fortement la charge). Le tableau présente des prédictions plus faibles que la charge réalisée, mais il faut tempérer ce bémol par le fait que l'amplitude des intervalles de confiance de prévision est importante, et que la prévision doit toujours être accompagnée d'une marge prudentielle. Ce sont les méthodes M1-M1 et M1-M3bis qui semblent donner les meilleurs résultats sur l'exercice récent.

Sur les exercices plus anciens, on enregistre des écarts parfois très importants, notamment sur la méthode M1-M3bis et les années 2000-2001. Cet écart est vraisemblablement dû à un phénomène déjà identifié dans le chapitre 4 : pour les sinistres importants, une rupture apparaît au mois 72 (année 2000), qui se caractérise par un ralentissement de la croissance du coût moyen des sinistres associés (voir section 4.4.2.2). Cette rupture est détectée assez tard (près de deux ans après). Il n'est donc pas étonnant de retrouver ce mauvais comportement sur la méthode M3bis.

En résumé, pour ce qui concerne la garantie RC Corporels, une difficulté supplémentaire vient du fait que la modélisation des coûts de sinistres paraît plus délicate que pour la garantie Vol.

5.1.2.3 Garantie RC Matériels

Pour la garantie RC Matériels, la détection de rupture sur les coûts de sinistre (Chapitre 3) ne conduit pas à une détection (même si les résultats de la section 3.3.2 semblent indiquer

qu'une rupture correspondant à une baisse s'est peut être produite en 2001).

Année	Prévision charge	Prévision de charges			
		Fréq : Méthode 1 / CM : Méthode 1	Fréq : Méthode 1 / CM : Méthode 2	Fréq : Méthode 1 / CM : Méthode 3	Fréq : Méthode 1 / CM : Méthode 3 bis
1998	242 868 019	228 587 381	227 617 572	236 935 716	234 000 355
1999	249 227 778	207 752 107	207 176 929	213 489 669	211 073 602
2000	244 498 882	223 762 389	223 464 751	228 020 707	225 666 959
2001	248 618 294	237 907 176	241 545 900	241 378 211	239 081 408
2002	236 741 633	236 353 575	250 620 905	251 349 756	248 799 322
2003	219 015 185	244 509 129	262 564 777	263 519 782	260 130 996
2004	211 070 212	271 202 119	270 269 100	269 728 631	264 922 224
2005	225 985 511	247 162 505	270 088 996	255 743 632	254 127 351
2006	217 579 672	255 685 858	266 369 570	249 939 546	251 080 481
2007	221 934 639	242 654 183	257 866 298	241 571 557	244 653 204
2008	230 531 356	232 326 813	251 742 169	235 470 400	240 331 359

La largeur de l'intervalle de prédiction est de l'ordre de 10 millions d'euros pour toutes les méthodes. On remarque que les différentes méthodes ne fournissent pas de prévision véritablement convenable, les méthodes M1 et M3 semblant être celles qui se comportent le moins mal. Néanmoins, les prévisions s'avèrent toujours plus prudentes. Pour les années les plus anciennes, la mauvaise performance peut s'interpréter comme une méconnaissance du risque du fait de la faible ancienneté de nos bases. En revanche, à partir de 2001, on peut vraisemblablement imputer la mauvaise performance au retard de détection. Les mêmes remarques que pour les deux autres garanties s'appliquent, concernant une éventuelle modification de la procédure pour anticiper cette détection.

L'année 2008 est convenablement prédite par la méthode 1 et la méthode 3, du fait d'une hausse de la sinistralité pour l'année 2008 qui vient compenser le fait que le modèle n'a toujours pas intégré la baisse entamée en 2001.

5.2 Prédiction de l'exercice 2009

Comme nous l'avons indiqué en introduction, l'année 2009 a été reconnue par le marché comme une année d'inversement d'une tendance globalement à la baisse dans les charges de sinistres. En comparant la sinistralité enregistrée sur l'année 2009 (telle que vue au 31 décembre 2009) avec la prévision qu'on aurait pu en faire au 31 décembre 2008 à partir de nos modèles, nous nous demandons si ce changement observé est bien significatif, ou rentre dans les limites de ce qu'on pouvait attendre.

Pour effectuer cette comparaison, nous confrontons notre prédiction à la prévision de charge ultime pour 2009 fournie par notre base de triangles. Rappelons que le but de la méthode des triangles n'est pas le même que celui que se fixe notre approche prédictive : la méthode des triangles se base sur la connaissance d'un nombre de sinistres effectivement réalisés dans l'année, et donc sur une information supplémentaire par rapport à notre vision (fin de l'année courante pour prédire l'année suivante).

5.2.1 Estimation de la charge ultime par liquidation de triangles

5.2.1.1 Pr sentation des triangles et liquidation par la m thode Chain Ladder

En fonction de la nature des garanties, les projections de triangles seront r alis es   partir de triangles de r glement et/ou de charges. En effet, pour les branches   d veloppement court (RC mat et Vol), l'utilisation des triangles de r glements avec 13 ann es d'historique suffit   obtenir une estimation stable de la charge (les triangles de r glements sur ces 2 garanties montrent que la charge est stabilis e d s la 5 me ann e).

En revanche, pour la garantie RC Corp, m me au-del  de la 12 me ann e de d veloppement, les triangles de r glement ne sont pas stabilis s, c'est pourquoi il est pr f rable d'avoir recours aux triangles de charges pour d terminer les ultimes.

Le m moire n'ayant vocation   d tailler les diff rentes m thodes possible de liquidation (d terministes ou stochastiques) nous nous limiterons   la pr sentation des r sultats pour la liquidation simple de type Chain Ladder.

	Charge 2009 D/D triangles	Charge 2009 D/D Base sinistres	Ultimes Chain Ladder
Vol	80 981 816	80 931 511	82 729 154
RC Mat	237 239 570	237 230 901	228 937 115
RC Corp	320 648 344	320 635 121	389 778 125

En fonction des informations qui seront compl t es au fur et   mesure que le temps passe, les estimations de charge courante seront revues au gr  des nouvelles diagonales aliment es dans les triangles. Ainsi nous allons d terminer dans la section suivante un intervalle de confiance sur cette  valuation de la charge ultime.

5.2.1.2 Intervalle de confiance sur la charge ultime

L'intervalle de confiance   99% a  t  r alis  en utilisant la technique d'estimation de la variance d crite dans Mack (1993) sous l'hypoth se que les charges suivent des lois normales.

	Borne inf	Ultimes Chain Ladder	Borne sup
Vol	61 603 173	82 729 154	103 855 135
RC Mat	182 928 603	228 937 115	274 945 628
RC Corp	300 789 541	389 778 125	478 766 708

5.2.2 Prédiction de la charge 2009 à partir des modèles de détection de rupture

Pour construire des intervalles de confiance sur la prévision, nous proposons deux approches, l'une basée sur une approximation gaussienne, l'autre sur une méthode type bootstrap. Dans les résultats que nous proposons, nous nous sommes concentrés sur la méthode plus rapide basée sur l'approximation gaussienne, mais nous précisons comment le recours à des simulations peut éventuellement améliorer ces bornes.

Pour utiliser l'approximation gaussienne, on a besoin de déterminer la variance prévue pour la charge de sinistres de l'année 2009. On rappelle que la charge mensuelle de sinistres est modélisée par un processus de Poisson composé, c'est à dire que la charge de sinistres enregistrés dans un mois j est

$$C_j = \sum_{i=1}^{N_j} X_{ij},$$

où N_j est le nombre de sinistres enregistrés le mois j (suit une loi de Poisson), et X_{ij} le montant du i -ème sinistre, les X_i étant supposés i.i.d. indépendants de N_j . Du fait de l'indépendance des X_{ij} et du caractère markovien des processus de Poisson, les C_j sont indépendants, et la variance de la charge annuelle de sinistres est donc

$$Var\left(\sum_{j=1}^{12} C_j\right) = \sum_{j=1}^{12} Var(C_j).$$

Par ailleurs, on rappelle que $Var(C_j) = m_{2X}\lambda_j$, où λ_j désigne le paramètre de la loi de Poisson associée à N_j , où $m_{2X} = E[X_{1j}^2]$. Le coefficient m_{2X} peut être facilement estimé à partir des données de façon empirique.

L'autre méthode consiste à utiliser une technique bootstrap pour mieux évaluer le comportement de la charge annuelle. Elle peut paraître plus adaptée, notamment dans le cas où le volume de sinistres considéré est plus faible. La méthode consiste à simuler de nombreuses réalisations des variables C_j , puis de déterminer les bornes de l'intervalle de confiance à partir de ces réalisations. Nous expliquons ci-dessous comment une telle variable $\hat{C}_j$ ayant une loi proche de celle de C_j peut être simulée.

- Simuler N_j en simulant une loi de Poisson de paramètre λ_j .
- Soit $X_{ij}^{(-1)}$ le coût du i -ème sinistre du mois j au cours de la dernière année disponible, et soit $m_j^{(-1)}$ son espérance (telle qu'évaluée par le modèle). Construire $\varepsilon_{ij} = X_{ij}^{(-1)} - m_j^{(-1)}$.
- Tirer uniformément, pour $k = 1 \dots N_j$, des éléments ε_k de façon indépendante dans l'ensemble des ε_{ij} .
- Calculer $\hat{C}_j = \sum_{k=1}^{N_j} \varepsilon_k$.

5.2.2.1 Garantie Vol

Les intervalles de confiance à 99 % sur la prédiction de l'année 2009 sont présentés dans le tableau suivant.

M�thode Freq / Co�t	Borne inf	Charge 2009	Borne sup
M1 - M1	101 247 740	103 762 323	106 276 906
M1 - M2	97 483 347	99 997 930	102 512 513
M1 - M3	72 497 020	75 011 603	77 526 186
M1 - M3 bis	69 878 736	72 393 319	74 907 903
M2 - M1	145 153 512	148 158 267	151 163 022
M2 - M2	139 778 479	142 783 234	145 787 989
M2 - M3	104 101 454	107 106 210	110 110 965
M2 - M3 bis	100 362 907	103 367 662	106 372 418

De mani re g n rale, nos m thodes fournissent des intervalles de confiance d'amplitude plus faible que ceux propos s par la m thode de Mack.

Sur la garantie Vol, les m thodes M1-M3 et M1-M3bis fournissent des pr visions qui sont coh rentes avec les intervalles obtenus par la m thode de Mack (les nouveaux intervalles de confiance sont strictement inclus dans ceux obtenus par la m thode de Mack). Elles semblaient donner les meilleurs r sultats sur les exercices pass s, ce sont   nouveau celles qui sont les plus proches de la charge 2009 telle qu'elle est  valu e par les triangles (aux alentours de 82 millions d'euros). Les autres techniques surestiment fortement cette charge. On note que la pr vision la plus haute (borne sup rieure de l'intervalle de confiance) obtenue par les deux m thodes retenues (M1-M3 et M1-M3bis) est inf rieure   l'estimation de charge du chain-ladder. Rappelons que cette pr vision a lieu au 31 d cembre 2008 (contrairement   l'estimation de chain-ladder qui est r alis e au 31 d cembre 2009), et n'int gre pas les premiers  l ments sur la sinistralit  de l'ann e 2009. Ceci tend   conforter l'hypoth se avanc e par les experts d'une ann e 2009 au caract re exceptionnel (dont il resterait   savoir s'il persiste dans le temps), avec un red marrage de la sinistralit    la hausse.

Il faudrait n anmoins avoir plus de recul pour pouvoir d terminer si cette hausse est un  cart momentan  au mod le, ou si elle est amen e   perdurer.

5.2.2.2 Garantie RC Corporels

Les r sultats concernant la garantie RC Corporels sont pr sent s dans le tableau suivant.

M�thode Freq / Co�t	Borne inf	Charge 2009	Borne sup
M1 M1	320 249 119	368 531 824	416 814 530
M1 M2	244 464 563	292 747 269	341 029 974
M1 M3b	306 769 341	355 052 047	403 334 752
M2 M1	404 819 445	458 685 030	512 550 615
M2 M2	310 495 865	364 361 450	418 227 035
M2 M3b	388 042 136	441 907 721	495 773 307

Les intervalles de confiance englobent pour la plupart la valeur estim e par les triangles (390 millions d'euros environ), de plus, pour les m thodes qui semblaient le mieux fonctionner sur la garantie RC Corporels (M1-M1 et M1-M3bis), les intervalles sont inclus dans l'intervalle de confiance calcul  suivant la m thode de Mack (1993). Si l'on s'en tient

à notre modèle, la hausse de la sinistralité (390 millions contre environ 360 prévus par nos modèles) reste dans les limites de tolérance fixées par le modèle. Bien que l'amplitude de nos intervalles de confiance (96 millions) soit bien plus faible que ceux calculés par méthode de Mack (178 millions), nous avons déjà constaté qu'ils restent encore relativement conséquents. Il serait donc nécessaire de se demander si une meilleure modélisation des coûts de sinistres ne permettrait pas d'augmenter notre précision.

5.2.2.3 Garantie RC Matériels

Les résultats concernant la garantie RC Matériels sont présentés dans le tableau suivant.

Méthode Freq / Coût	Borne inf	Charge 2009	Borne sup
M1 - M1	215 657 895	220 387 947	225 117 999
M1 - M2	243 272 301	248 002 352	252 732 404
M1 - M3	223 522 054	228 252 105	232 982 157
M1 - M3 bis	231 753 122	236 483 174	241 213 225

Nos intervalles de confiance sont tous inclus dans l'intervalle de confiance de Mack, et présentent des amplitudes bien plus faibles (10 millions contre 90 pour Mack). Les méthodes qui semblaient les mieux adaptées étaient M1 et M3. Pour la méthode M1, la charge 2009 estimée par triangle est en-dehors des bornes de l'intervalle de confiance, et on conclut donc bien à une année de hausse de la sinistralité. En revanche, pour la méthode M3, la charge rentre dans les bornes de l'intervalle de confiance. Ce phénomène est sans doute dû au fait que le test du rapport de vraisemblance (Chapitre 3, section 3.3.2) ne concluait pas à une rupture alors que d'autres indicateurs semblaient la suggérer. De ce fait, n'ayant pas intégré une baisse, la méthode M3 est moins sensible à la hausse qui lui fait suite (c'est aussi pour cette raison qu'elle se comportait relativement bien sur l'année 2008, où l'on ressentait déjà un début de hausse).

5.3 Mises en œuvre opérationnelles

Dans cette dernière partie nous nous sommes attachés, en collaboration avec les équipes techniques de la MACIF, à présenter différentes modalités d'application possibles de notre outil de détection de rupture.

Comme nous l'avons présenté dans le chapitre précédent, notre outil n'a pas vocation à se substituer aux méthodes actuelles de provisionnement mais permet de construire des indicateurs d'alerte.

Dans ce cadre, la mise en place de modèles de détection de rupture peut contribuer à améliorer les processus opérationnels et les prises de décisions dans 3 grands domaines que sont :

- Les **reportings opérationnels** de surveillance de portefeuille
- Les **provisionnements** des sinistres (gestion et inventaire)
- Les constructions de scénarii pour la **projection de résultats**

5.3.1 Reportings et surveillance de portefeuille

Dans le cadre de l'élaboration de reportings à dates régulières, l'outil de détection de rupture vient éclairer des tendances observées sur l'évolution de la sinistralité. Si notre étude se concentre ici sur la question de la fréquence des sinistres et de leurs coûts, cette démarche peut être étendue à d'autres indicateurs (flux des primes, évolutions structurelles du portefeuille...).

Affiner les comparaisons entre l'évolution du portefeuille et celle du marché.

En l'état actuel, le test de détection de rupture étudié fournit un délai de détection important du point de vue de la fréquence des sinistres. Néanmoins, du strict point de vue de l'estimation d'une date de rupture, les résultats semblent relativement satisfaisants. L'outil peut donc être utilisé en complément d'une analyse du marché, afin de saisir avec plus de précision la date à laquelle une tendance observée sur l'ensemble du marché a commencé à impacter le portefeuille. Ce type d'information peut permettre ensuite aux opérationnels d'affiner les comparaisons entre la situation du portefeuille et celle du marché. Il permet également d'affiner les prévisions de sinistres pour les années ultérieures, en évitant de s'appuyer sur des années avant rupture qui témoignent d'une dynamique à présent obsolète.

Segmentation du portefeuille

Le fait d'avoir utilisé une modélisation des sinistres par un processus de Poisson permet à la technique d'être appliquée au cadre de portefeuilles de petite taille (par rapport à des techniques usuelles basées sur une approximation gaussienne). Il est donc possible de l'appliquer sur un portefeuille segmenté au plus près des besoins d'analyse. Ainsi, dans le cas spécifique du portefeuille MACIF pour lequel un suivi régionalisé de la sinistralité est nécessaire, notre outil permettrait de construire des indicateurs de sinistralité selon les croisements "Zone géographique * Type de véhicules * Produits * Garanties".

Surveillance de la sinistralité

Enfin, dans le cadre de travaux de scoring et de développements de nouveaux produits, la surveillance de la sinistralité selon des critères plus fins de type tranche d'âge du conducteur / sexe / ancienneté du permis pourrait fournir des alertes sur les besoins de refonte de produits pour certains profils au regard des dérives de comportements et d'apporter des indications sur les éventuelles révisions tarifaires nécessaires.

5.3.2 Provisionnement en gestion et en inventaire

Bien que notre outil ne constitue pas en tant que tel un outil de provisionnement, sa mise en place offrirait des indicateurs permettant d'affiner les méthodes de provisionnements.

Forfaits d'ouverture

Les résultats sur les délais de détection de rupture des coûts moyens (inférieurs à 12 mois sur les sinistres Vol - cf section 4.4.2.1) semblent suffisamment satisfaisant pour justifier la construction d'indicateurs d'alerte sur la mise à jour des forfaits d'ouverture des sinistres au sein des systèmes de gestion. Actuellement les forfaits d'ouverture de la MACIF sont revus annuellement en fonction des boni/ mali dégagés sur ces petits sinistres de masse. La mise en place d'une procédure de détection dynamique de rupture sur les coûts moyens permettrait de donner des signaux d'alerte quantifiables (probabilité de rupture avérée/ intensité de la rupture) pour calculer un ajustement des forfaits d'ouverture au plus près de l'évolution des coûts.

La possibilité de calibrer dans la procédure le niveau des seuils de détection (1%, 5%...) permet aux opérationnels de déterminer, en fonction de leur besoin, l'équilibre entre nombre de détections / fausses alertes et délais de détection.

Branches à délais de déclaration longs

Dans le cadre, des branches à délais de déclaration longs (RC médicale / Construction) une analyse de l'évolution des délais de déclaration des sinistres permettrait de fournir des éléments sur l'évolution des taux de d'IBNR et sur les besoins d'ajustement des provisions d'inventaires. Ce type d'étude pourrait également être étendu au suivi des taux de recours (analyse de la stabilité des taux de recours dans le temps) particulièrement significatif sur les branches Responsabilité Civile et Dommage Ouvrage.

5.3.3 Construction de scénarii et projections de résultats techniques

Les outils présentés dans ce mémoire trouvent une application directe dans l'élaboration de résultats intermédiaires grâce notamment à la meilleure compréhension des caractéristiques de saisonnalité du risque sous-jacent. Ils répondent également à un besoin d'aide à l'élaboration de scénarii plus ou moins crédibles. A plus longue échéance, la technique pourrait devenir un élément d'analyse et de comparaison par rapport à une référence (référence qui reste encore à définir).

Résultats intermédiaires

L'analyse des effets de saisonnalités sur la sinistralité du portefeuille, telle que traitée dans le mémoire, semble apporter une solution concrète aux problématiques de prévisions de résultats intermédiaires (trimestriel / mensuel). En effet, compte tenu de l'impact de la saisonnalité, le résultat annuel n'évolue pas de manière linéaire dans le temps ; par exemple, pour la garantie Vol, la fréquence des sinistres sur la période de mars à juin est significativement plus importante que sur les 2 derniers mois de l'année (cf section 3.3.1), générant une croissance de la charge plus rapide sur le deuxième trimestre et un ralentissement sur novembre-décembre.

Ainsi, les travaux réalisés sur la décomposition de l'année en périodes de fréquence de sinistres homogènes (détermination du nombre de sous-périodes et estimation du paramètre d'ajustement pour chaque période) devrait permettre, sur la base des résultats des premiers

mois d'une année, de donner une prévision plus fine des résultats attendus sur les différents trimestres à venir et donc d'en assurer un meilleur pilotage (notre étude ayant montré une certaine stabilité des découpages par période au cours des années).

Appui à l'élaboration de scénarii

L'utilisation de la procédure de détection de rupture à la fois sur la fréquence et les coûts permet de construire, sur les données historiques, des courbes de tendances ajustées plus finement et offre en mode de détection dynamique, un ajustement « automatique » des trajectoires dès le franchissement du seuil d'alerte. La définition de différents niveaux de seuils d'alerte, et par conséquent de courbes de seuil, peut permettre de construire plusieurs jeux de courbes correspondant à des probabilités de fausses alertes (sur-détections) plus ou moins importants.

Ainsi, dans le cadre de l'élaboration des Business Plans et du développement d'un modèle interne, la procédure pourrait permettre de proposer différents jeux de scénarii sur les évolutions de la sinistralité et présenter les déformations des marges attendues en cas de rupture avérée de tendances et/ou d'objectiver les scénarii retenus.

Comparaison à une référence

Dans le cadre d'une mise en œuvre à plus long terme, la procédure de détection de rupture pourrait être utilisée pour comparer l'évolution du portefeuille par rapport à un "benchmark" et générer des alertes en cas de dérives. A titre d'exemple, l'évolution des coûts des sinistres pourrait être modélisée par rapport à un indicateur inflationniste, la procédure de détection de rupture dynamique donnerait le cas échéant, les éléments d'alertes pour recalibrer l'indicateur (taux d'inflation + x%) en cas de dérives entre la sinistralité observée et le benchmark. Toutefois, une telle utilisation nécessite en amont de définir et/ou de créer le benchmark de référence présentant sur l'historique un comportement assez proche de la tendance à modéliser.

5.3.4 Conclusions

Les différents axes d'utilisation présentés devraient apporter une meilleure visibilité sur les évolutions du portefeuille dans la mesure où la procédure permet d'« objectiver » des inflexions de tendances observées (rupture avérée / phénomène ponctuel). Outre l'amélioration constatée des outils de prévision, la technique pourrait déboucher sur la construction d'indicateurs, et d'outils de pilotage, que ce soit pour tenir compte de l'évolution des coûts comme de la fréquence de sinistres. Concernant cette dernière, la segmentation de l'année en sous-périodes comparables permet notamment une meilleure compréhension de la dynamique des sinistres, et une meilleure prise en compte des effets saisonniers sous-jacents.

5.4 Conclusion

Sur les données historiques, nous confirmons les conjectures effectuées dans les chapitres précédents. Les méthodes qui consistent à segmenter sinistres importants et sinistres faibles semblent donner de meilleurs résultats pour modéliser la charge de sinistres. En revanche,

sur certains exercices, on peut imputer la mauvaise performance des méthodes par le mauvais comportement du test du rapport de vraisemblance pour la détection de rupture dans la fréquence de sinistres. Ce mauvais comportement est dû notamment au caractère saisonnier des observations (qui diminue le nombre d'observations, et donc la performance de la méthode). Les approches par test du rapport de vraisemblance sont trop peu puissantes du point de vue statistique lorsque les observations sont en nombre si faible. Par ailleurs, nous avons noté sur la garantie RC Matériels (voir Chapitre 3) que cette absence de détection était peut-être due à notre niveau d'information insuffisant (dans notre modèle, les fluctuations des primes peuvent masquer le phénomène, pour cette raison il serait plus intéressant de connaître le nombre d'assurés dans le portefeuille).

Pour cette raison, il paraît important de ne pas se limiter à cette approche pour détecter des éventuelles ruptures de sinistralité. Une première piste, qui consisterait à ne regarder que la fréquence de sinistres, consisterait à observer plus attentivement les résultats de la section 3.3.2 (dans une détection dynamique de rupture, la fluctuation de la date de rupture estimée au cours du temps peut être informative, et nécessiterait d'être utilisée). Le présent chapitre suggère également une deuxième possibilité. On peut également comparer la prédiction de charge de sinistre en début d'année, avec celle effectuée à partir de la méthode Chain-Ladder en fin d'année. Un écart significatif entre les deux peut être un signal d'alerte indiquant qu'il devient nécessaire de recalibrer le modèle. Sur les garanties RC Corporels et Vol, il apparaît que des écarts significatifs apparaissent entre ces prévisions précisément à partir de la date estimée de rupture dans la fréquence de sinistres.

Conclusion et perspectives

Dans ce mémoire nous avons essayé d'appliquer des techniques récentes permettant de tester la stabilité d'un risque en assurance automobile. Les techniques de détection de rupture que nous avons mises en œuvre ont pour but de réagir de façon rapide mais néanmoins prudente à des évolutions de la sinistralité, et nous les avons appliquées à la fois à une modélisation des fréquences de sinistres mais aussi des coûts. Plusieurs constats s'imposent à la suite de cette expérimentation.

1. **Limites liées aux données :** Afin de modéliser les fluctuations du portefeuille liées à une baisse ou une hausse du nombre d'assurés, nous avons tenu compte des primes acquises dans notre modélisation. Cette approche ne nous paraît pas pleinement satisfaisante. Même si l'évolution des primes acquises reste globalement stable, le manque d'information sur l'évolution des politiques tarifaires entre 1994 et 2008 peut masquer un certain nombre de phénomènes. Nous pensons donc que la technique que nous proposons serait sans doute plus performante si on l'associait à la donnée du nombre d'assurés présents dans le portefeuille. Dans ce cas, le nombre de sinistre ne serait plus rapporté à l'euro de prime acquise, mais au nombre d'assurés. Quoi qu'il en soit, la mise en œuvre du modèle de prédiction nécessitera une prédiction du nombre d'assurés présents dans le portefeuille l'année suivante (ou des primes acquises, si l'on en reste au modèle considéré dans ce mémoire).
2. **Rupture dans la fréquence des sinistres :** les méthodes de détection de rupture multiples nous ont permis de segmenter l'année en 4 périodes distinctes de sinistralité comparable. Cette segmentation semble fonctionner de manière raisonnable (on a vu que le modèle identifie une structure globalement cohérente d'une année sur l'autre). Par ailleurs, cette segmentation nous a permis d'anticiper la détection de rupture pour l'effectuer en cours d'année. Du point de vue de l'anticipation de la détection, nous avons constaté un retard qui pourrait sans doute être diminué en basant la détection non seulement sur le test du rapport de vraisemblance, mais en utilisant d'autres informations disponibles (évolution des estimateurs de la date de rupture au cours du temps, écarts entre prévision de la charge et charge ultime conjecturée par les triangles à la fin de l'année...).
3. **Coûts de sinistres :** Il apparaît intéressant de traiter séparément les petits et gros sinistres. Des progrès pourraient vraisemblablement être obtenus en modélisant de façon plus précise les lois des sinistres (particulièrement les plus importants). En effet, nous ne sommes pas parvenus à ajuster de lois simples à ces coûts (les données ne sont pas en adéquation avec la loi de Pareto). Une piste intéressante serait de s'orienter vers des mélanges de lois (de type exponentiel ou Pareto), qui

permettraient de rendre compte de l'hétérogénéité de la distribution suivant que l'on se place dans la queue de distribution ou vers les petites valeurs.

4. **Commentaires sur les différentes garanties :** Les résultats que nous obtenons semblent relativement bien adaptés aux branches courtes telles que Vol et RC Matériels. Pour la garantie RC Corporels, la qualité des résultats est moins évidente, pour deux raisons : la qualité des données est moins bonne (nous effectuons l'approximation que la charge ultime de sinistre est connue instantanément, ce qui n'est pas le cas pour cette branche où la vision de la charge ultime est susceptible d'évoluer fortement au cours du temps et donc de polluer la procédure), par ailleurs la variance sur les coûts de sinistres est particulièrement importante et fournit de grandes incertitudes.
5. **Année 2009 :** Parmi les techniques que nous proposons, celles qui semblent fonctionner le mieux (basées sur une détection de rupture dans les fréquences, et sur une segmentation des gros et petits sinistres) confirment l'avis général d'une fin de la baisse de la sinistralité pour les sinistres matériels (Vol et RC Matériels), puisque les estimations de la charge vue du 31 décembre 2009 sont éloignées des prévisions obtenues fin 2008. Pour la garantie RC Corporels, nos estimations ne permettent pas de statuer sur une modification de la sinistralité en 2009. Ceci peut notamment s'expliquer par les difficultés rencontrées dans la modélisation des coûts des sinistres (cette branche se caractérise par une volatilité importante des coûts de sinistres), générant une grande marge d'incertitude dans nos prévisions.
6. **Différents cadres d'application de ces techniques :** Les outils proposés sont voués à s'appliquer dans un large nombre de domaines, dont nous avons effectué l'examen plus détaillé dans la section 5.3. Ils apparaissent comme prometteurs dans trois champs principaux : reporting, provisionnement, et projection. La segmentation de l'année en périodes de sinistralité comparable s'avère importante pour une meilleure compréhension des phénomènes saisonniers. Par ailleurs, l'utilisation d'estimateurs des dates de rupture peut permettre de confronter l'évolution de la sinistralité d'un portefeuille avec une tendance observée sur le marché. Elle peut permettre également de mieux évaluer la pertinence de scénarii utilisés dans le cadre de la projection de résultats.

Mentionnons également quelques prolongements des techniques présentées dans ce mémoire. Nous nous sommes concentrés ici sur des portefeuilles de taille considérable. Il nous paraîtrait intéressant d'étudier le comportement sur des volumes plus faibles. En effet, l'écart avec l'hypothèse gaussienne (hypothèse classique de la théorie de détection de rupture) devrait être plus marqué encore sur des petits volumes, et devrait rendre les techniques présentées d'autant plus précieuses. Une autre question intéressante serait de confronter ces techniques de détection de rupture avec des chocs, comme on peut en constater notamment avec d'autres types de garanties telle que celle des risques climatiques. Les garanties climatiques peuvent être caractérisées par des chocs importants qui ne sont pas nécessairement dus à un changement durable de la sinistralité, mais simplement à un phénomène ponctuel. Il serait intéressant de voir comment les méthodes statistiques de détection de rupture pourraient permettre de distinguer un choc d'une tendance durable, notamment en regardant au bout de combien de temps le choc est "absorbé" par la procédure.

Enfin, nous nous sommes essentiellement concentrés sur un changement dans la moyenne des séries considérées (même si le changement de paramètre dans un processus de Poisson signifie également un changement de variance). Pour certaines séries que nous avons considérées, la variance semble évoluer au cours du temps, et il serait sans doute utile de mettre en place une technique de détection similaire pour saisir des fluctuations brutales de variance. Notons que ce problème de détection de rupture dans la variance pourrait également être d'un intérêt considérable pour des applications financières, où la bonne modélisation de la volatilité joue un rôle crucial.

Annexe

Annexe A

Statistiques descriptives

Nous présentons dans cette annexe quelques informations supplémentaires sur la volumétrie du portefeuille que nous considérons. Nous décrivons également les incohérences que nous avons relevées entre la base sinistre et la base triangles de règlements.

A.1 Répartition des sinistres par mois

Nous présentons dans cette section, garantie par garantie, les fréquences mensuelles sur le portefeuille considéré. Nous excluons les sinistres classés sans suite.

A.1.1 Garantie Vol

Nb sinistres	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Moyenne	Ecart-type
Janvier	10745	9953	9360	8275	8649	8650	7828	8635	8816	7154	6781	5754	5052	4910	3984	3659	7393	2455
Février	9166	9108	8455	8318	8155	7787	7853	7948	8276	6638	6429	5082	4680	4488	3829	3589	6864	1924
Mars	10284	9786	9468	8876	8689	8336	8231	8784	8846	7456	6842	5853	5337	4739	3885	4058	7461	2083
Avril	9855	8901	8679	8238	8188	8451	7935	8655	8713	7034	6242	5773	5238	4614	3788	3906	7139	1947
Mai	10265	9536	9324	8253	8540	8749	8372	8539	8727	8028	6537	6110	5590	4885	4222	4189	7555	1989
Juin	10349	9720	8952	8958	8556	8781	8411	8810	7611	7439	6825	6034	5258	4889	3983	4222	7429	1966
Juillet	9360	9042	8262	8713	8900	8506	8659	8724	7915	7408	6883	6382	5150	4621	4039	4102	7338	1923
Août	9216	8384	8076	7036	7508	7519	7359	7687	7143	6285	5802	5460	4913	4107	3549	3540	6472	1747
Septembre	9381	8733	8356	7646	8070	7853	7684	8212	7131	6883	5809	5273	4524	3981	3388	3508	6680	1957
Octobre	10181	9403	9172	8517	9107	8471	8843	9108	8049	7632	6233	5695	5126	4388	3888	3681	7335	2478
Novembre	10240	8771	8823	8681	8522	8331	8529	8922	7689	7081	6326	5254	4908	4154	3912	3602	7405	2410
Décembre	8455	8844	8370	8381	8138	7417	7785	8027	7282	6780	5887	5119	4586	3838	3587	2885	6640	2072
Moyenne	9875	9181	8658	8409	8449	8245	8132	8516	8017	7146	6361	5648	5091	4455	3883	3745	7115	1993
Ecart-type	535	487	468	588	429	484	445	429	657	488	404	410	324	382	235	379	370	1984

A.1.2 Garantie RC Corporels

Nb sinistres	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Moyenne	Ecart-type
Janvier	2970	3002	2848	2472	3020	3041	3104	2993	2620	2313	2260	2222	2041	2082	2051	1926	2583	431
Février	2280	2634	2512	2333	2451	2614	2792	2460	2273	1736	1762	1762	1668	1797	1848	1823	2460	404
Mars	2602	2664	2591	2783	2753	2987	3086	3085	2857	2117	2070	2068	2048	2083	1901	1974	2475	424
Avril	2766	2678	2464	2793	3078	2966	2975	2744	2382	2201	2046	2084	1921	1998	1888	1889	2430	427
Mai	2773	2963	2721	3081	3026	3168	3057	2772	2675	2357	2323	2357	2098	2116	2086	2054	2684	402
Juin	2931	2951	3047	3271	3142	3516	2909	2960	2638	2375	2480	2436	2387	2250	2041	2303	2732	416
Juillet	2680	2953	2846	3087	2908	3129	2984	2887	2625	2383	2232	2248	2059	2178	2045	2083	2584	400
Août	2684	2680	2809	3013	2904	3118	2635	2560	2541	2174	2152	2133	1932	1930	1748	1820	2427	443
Septembre	3138	3086	3086	3048	3660	3521	3192	2978	2737	2492	2484	2287	2288	2250	2117	2233	2782	478
Octobre	3428	3311	3460	3570	3883	3588	3382	3251	2975	2658	2645	2476	2434	2348	2381	2383	3011	531
Novembre	2985	3180	3471	3569	3486	3413	3337	3152	2837	2316	2373	2457	2341	2274	2071	1981	2829	560
Décembre	3284	3086	3085	3402	3330	3611	3339	2904	2518	2257	2401	2302	2304	2219	2038	1255	2745	683
Moyenne	2879	3000	2908	3035	3129	3224	3065	2887	2623	2282	2272	2243	2128	2127	2019	1960	2612	449
Ecart-type	314	302	331	383	388	310	226	226	187	225	238	209	230	160	161	304	226	508

A.1.3 Garantie RC Matériels

Nb sinistres	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Moyenne	Ecart-type
Janvier	19158	9953	18111	18893	18794	18921	19155	18838	17599	17473	17271	16211	16071	15288	16186	18383	17295	2310
Février	16880	9108	18230	15045	15508	17227	17853	16028	15480	13352	13271	14443	13031	13429	15084	14781	14908	2228
Mars	16881	9766	15885	16458	17506	18501	18893	19204	17288	15225	15675	16088	16041	16011	16455	16720	16370	2485
Avril	17953	8901	14848	17383	19140	18888	17826	17818	16334	15275	15030	14815	14203	14801	16058	16212	15883	2380
Mai	17235	9536	16838	18539	17871	18852	18920	17195	17886	16088	15888	16588	15311	15719	16404	16988	16582	2481
Juin	18048	9720	16971	19215	18620	19351	17312	18208	16757	16585	16785	16725	16087	16675	16286	18224	16973	2491
Juillet	9350	16241	16816	16988	16738	17377	17738	16957	16338	15627	15231	15185	14837	15628	15930	16785	15814	1986
Août	9216	14241	14531	15082	15201	15885	14832	14771	14258	13314	13413	13228	13188	13138	13681	14305	13887	1473
Septembre	9381	18488	16578	16814	18750	19408	17885	17885	16284	15934	15588	15354	15315	15583	16530	17084	16485	2342
Octobre	10181	17445	18816	19888	21253	20734	18638	19023	18438	17584	16884	15732	16020	16228	17778	17928	17717	2585
Novembre	10240	18259	20278	19626	20459	19929	18888	18826	17955	15474	15752	16207	16088	16219	16282	16041	17330	2614
Décembre	9455	22628	20338	21282	20875	23071	20337	20675	16878	16783	17527	17277	17158	18101	18378	16285	18572	3287
Moyenne	13645	13881	17289	17919	18488	18914	18281	17986	16778	15719	15671	15683	15284	15588	16254	16634	16479	1594
Ecart-type	4242	4777	1922	1947	2085	1872	1498	1574	1149	1385	1361	1135	1285	1344	1474	1250	1264	2570

A.2 Répartition des coûts moyens mensuels

Nous présentons dans cette section, garantie par garantie, le coût moyen mensuel des sinistres enregistrés dans le portefeuille.

A.2.1 Garantie Vol

Coût moyen	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Moyenne	Ecart-type
Janvier	1 282	1 305	1 312	1 375	1 225	1 268	1 281	1 329	1 393	1 456	1 401	1 434	1 643	1 664	1 688	2 052	1 446	15
Février	1 350	1 223	1 394	1 332	1 302	1 264	1 287	1 372	1 348	1 396	1 390	1 558	1 642	1 732	1 688	1 883	1 446	14
Mars	1 338	1 288	1 286	1 292	1 300	1 260	1 293	1 269	1 453	1 415	1 463	1 515	1 570	1 790	1 790	1 972	1 458	15
Avril	1 232	1 229	1 276	1 311	1 218	1 212	1 241	1 348	1 458	1 398	1 382	1 442	1 558	1 637	1 745	1 919	1 413	14
Mai	1 219	1 270	1 288	1 255	1 222	1 205	1 225	1 376	1 398	1 349	1 358	1 442	1 574	1 577	1 518	1 796	1 379	13
Juin	1 214	1 263	1 265	1 250	1 257	1 197	1 233	1 317	1 394	1 406	1 311	1 499	1 441	1 620	1 784	1 790	1 392	14
Juillet	1 207	1 166	1 240	1 226	1 208	1 165	1 177	1 296	1 349	1 405	1 333	1 396	1 422	1 550	1 584	1 803	1 347	13
Août	1 103	1 185	1 132	1 167	1 136	1 111	1 105	1 155	1 206	1 216	1 238	1 314	1 398	1 381	1 518	1 657	1 252	13
Septembre	1 286	1 309	1 268	1 321	1 263	1 301	1 262	1 322	1 419	1 421	1 381	1 498	1 596	1 603	1 763	1 923	1 433	14
Octobre	1 233	1 242	1 318	1 322	1 258	1 269	1 296	1 356	1 452	1 402	1 481	1 571	1 693	1 726	1 887	1 812	1 457	15
Novembre	1 239	1 306	1 306	1 276	1 249	1 276	1 281	1 450	1 399	1 408	1 480	1 615	1 630	1 673	1 825	1 674	1 440	14
Décembre	1 238	1 286	1 284	1 251	1 248	1 305	1 310	1 393	1 436	1 454	1 486	1 684	1 675	1 767	1 972	1 173	1 432	15
Moyenne	1 245	1 256	1 277	1 285	1 241	1 243	1 250	1 305	1 389	1 384	1 389	1 496	1 569	1 644	1 729	1 788	1 408	14
Ecart-type	8	7	7	7	7	8	8	9	8	8	9	10	10	10	12	15	8	14

A.2.2 Garantie RC Corporels

Coût moyen	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Moyenne	Ecart-type
Janvier	8 382	6 134	8 751	7 349	10 555	10 476	9 799	9 729	15 629	9 108	13 689	10 548	14 032	13 380	12 829	9 367	10 494	53
Février	8 129	7 162	6 396	9 193	7 068	13 264	7 273	9 670	9 377	10 525	11 222	9 796	17 622	16 020	13 394	15 166	10 517	61
Mars	7 450	6 725	6 273	10 818	6 793	8 068	9 280	9 865	10 582	21 185	9 572	13 170	17 137	16 817	13 428	15 298	11 389	66
Avril	7 425	6 697	6 993	7 668	6 772	9 104	12 877	9 477	8 428	13 397	10 482	8 749	12 257	18 671	12 699	13 687	10 486	57
Mai	6 935	5 960	9 959	8 484	9 597	10 642	6 738	11 824	14 094	13 633	7 919	12 983	19 887	12 684	15 093	13 707	11 166	61
Juin	6 009	6 755	7 597	6 301	7 288	10 264	11 525	12 472	10 599	12 445	8 909	11 386	18 386	19 804	15 783	15 781	11 326	65
Juillet	9 024	11 664	11 224	8 123	11 599	12 285	9 315	11 169	11 316	12 026	11 400	18 283	17 454	16 852	15 044	18 651	12 839	58
Août	8 200	8 428	9 500	9 849	12 287	13 341	10 124	13 165	15 163	13 188	11 743	14 151	16 838	18 227	21 249	17 624	13 325	61
Septembre	6 953	6 967	8 882	8 352	8 223	8 430	7 688	16 745	9 418	10 082	8 093	12 700	12 339	10 682	13 812	11 190	9 996	53
Octobre	6 240	6 932	8 888	7 153	9 498	12 880	8 988	9 044	10 220	11 091	10 304	9 360	10 685	12 842	14 823	16 885	10 319	53
Novembre	6 823	5 953	7 032	8 010	9 204	8 751	8 048	11 785	8 109	9 080	17 982	9 414	13 714	10 002	26 229	7 442	10 456	72
Décembre	6 488	6 689	6 755	10 089	12 418	9 707	8 815	16 882	15 152	14 345	9 787	16 423	18 808	11 573	14 408	5 424	11 280	66
Moyenne	7 020	7 170	8 085	8 435	9 253	10 346	9 162	11 888	11 503	12 485	10 935	12 222	15 731	14 821	15 712	13 350	11 129	53
Ecart-type	32	30	41	36	46	47	43	52	53	57	53	54	54	57	63	65	32	61

A.2.3 Garantie RC Matériels

Coût Moyen	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Moyenne	Ecart-type
Janvier	1 002	2 082	1 062	1 057	1 088	1 096	1 110	1 164	1 172	1 179	1 189	1 186	1 204	1 217	1 064	1 288	1 496	16
Février	981	1 074	1 078	1 057	1 079	1 121	1 117	1 143	1 186	1 170	1 176	1 217	1 196	1 224	1 007	1 254	1 481	14
Mars	970	1 978	1 043	1 048	1 062	1 085	1 089	1 157	1 186	1 158	1 116	1 182	1 187	1 205	1 129	1 207	1 475	15
Avril	1 013	1 076	1 033	1 050	1 096	1 116	1 124	1 142	1 163	1 152	1 127	1 177	1 151	1 182	1 181	1 204	1 474	14
Mai	979	1 901	1 072	1 088	1 076	1 075	1 096	1 120	1 200	1 157	1 082	1 181	1 186	1 210	1 162	1 199	1 471	14
Juin	975	1 805	1 027	1 084	1 088	1 109	1 082	1 126	1 185	1 123	1 026	1 185	1 158	1 171	1 187	1 201	1 454	14
Juillet	1 708	1 030	1 048	1 073	1 147	1 067	1 136	1 177	1 186	1 162	982	1 205	1 187	1 214	1 227	1 205	1 473	13
Août	1 642	1 027	1 083	1 120	1 106	1 088	1 104	1 146	1 214	1 245	1 115	1 207	1 196	1 212	1 380	1 170	1 493	12
Septembre	1 971	1 022	1 042	1 033	1 103	1 089	1 123	1 167	1 141	1 130	1 114	1 170	1 140	1 175	1 172	1 236	1 478	15
Octobre	1 964	1 001	1 037	1 088	1 113	1 079	1 123	1 155	1 162	1 144	1 185	1 358	1 198	1 144	1 247	1 139	1 495	15
Novembre	1 677	1 025	1 070	1 102	1 083	1 112	1 120	1 152	1 178	1 158	1 161	1 203	1 198	1 182	1 202	1 079	1 469	12
Décembre	2 248	1 088	1 080	1 088	1 116	1 117	1 135	1 202	1 182	1 185	1 190	1 216	1 248	1 141	1 225	1 070	1 216	17
Moyenne	1 428	1 475	1 055	1 069	1 096	1 098	1 114	1 154	1 176	1 162	1 122	1 205	1 186	1 190	1 183	1 188	1 481	1 481
Ecart-type	22	22	4	5	5	4	4	5	4	5	8	7	5	5	10	8	4	14

Annexe B

Résumé détaillé de la preuve du Théorème 2

Nous présentons dans cette annexe les principales étapes qui conduisent à la preuve du Théorème 2. Les arguments complets peuvent être consultés dans l'article correspondant. Le but est de montrer que, dans le cas du modèle poissonnien à k ruptures, on a

$$E \left[\sup_{\theta} \exp(L(t, \mu) - L(t_0, \mu_0) + d(t, \mu, t_0, \mu_0)) \right] \leq C[\log T]^k, \quad (\text{B.1})$$

où l'on a repris les notations du Théorème 2. La preuve de ce résultat peut se ramener au cas $k = 1$ comme nous le précisons ci-dessous. L'obtention de ce résultat s'obtient à partir d'un théorème général permettant de contrôler le supremum d'un processus (section B.1). L'utilisation de ce théorème repose sur l'étude des grandes déviations du processus $L(t, \mu)$, et des arguments géométriques (section B.2).

B.1 Contrôle des variations d'un processus

B.1.1 Norme d'Orlicz

On considère une fonction ψ convexe croissante satisfaisant $\psi(0) = 0$. La norme d'Orlicz (associée à ψ) d'une variable aléatoire X est définie comme

$$\|X\|_{\psi} = \inf \left\{ C > 0 : E \left[\psi \left(\frac{|X|}{C} \right) \right] \leq 1 \right\}.$$

Les normes L^p usuelles sont des cas particuliers de normes d'Orlicz. En théorie des grandes déviations, les normes définies par $\psi_p = \exp(x^p) - 1$ sont utilisées pour cerner le comportement des queues de distribution. Nous nous contenterons de considérer les normes définies par ψ_p .

B.1.2 Nombre de recouvrements

Pour une semi-distance δ sur l'ensemble $\mathcal{A} \times [0, T]$. On définit $N(\varepsilon, \delta)$, comme le nombre maximum d'ensembles du type $B_{\varepsilon}(t) = \{t' \in \mathcal{A} \times [0, T] : \delta(t, t') \leq \varepsilon\}$ nécessaire pour recouvrir $\mathcal{A} \times [0, T]$.

Par la suite, nous mettrons en évidence une semi-distance δ qui apparaît comme une caractéristique du problème statistique que nous nous posons. Le nombre $N(\varepsilon, \delta)$ s'interprète comme une complexité de la distance (donc du problème statistique) que nous considérons.

B.1.3 Théorème général

Nous reproduisons ici le Théorème général qui est à la base de la preuve du Théorème 2. On pourra consulter Van der Vaart et Wellner (1996) pour plus de précisions (voir leur Théorème 2.2.4 et Corollaire 2.2.5).

Théorème 3. *Soit $(X_u)_{u \in \mathcal{A} \times [0, T]}$ un processus stochastique avec*

$$\|X_s - X_u\|_{\psi_p} \leq C\delta(s, t), \quad (\text{B.2})$$

pour tout s et t , une semi-métrique δ , et C une constante positive. Alors, pour tous η et $\delta > 0$,

$$\left\| \sup_{s, u} |X_s - X_u| \right\|_{\psi} \leq K \int_0^T \psi_p^{-1}(N(\varepsilon, \delta)) d\varepsilon,$$

où K est une constante qui ne dépend que de C et p .

Ce résultat permet de borner les fluctuations du supremum d'un processus, ce qui est exactement ce que nous cherchons à faire. Dans la section suivante, nous définissons quel processus X_t nous considérons, quelle est la semi-métrique d , et comment on peut en déduire le résultat final du Théorème 2.

B.2 Application

B.2.1 Processus X_u

Soit $u = (\mu, t)$, où $\mu \in \mathcal{A}$ est une amplitude de rupture, et t un instant de rupture. Le processus X_u auquel nous appliquons le Théorème 3 est défini comme suit,

$$X_u = |L(t, \mu) - L(t_0, \mu_0) + d(t, \mu, t_0, \mu_0)|^{1/p},$$

où d est la distance définie au Chapitre 3 et qui intervient dans le Théorème 2.

B.2.2 Semi-distance δ

On peut facilement trouver une distance δ qui vérifie B.2, grâce au Lemme suivant, qui relie les normes d'Orlicz aux grandes déviations.

Lemme 4. *Soit X une variable aléatoire satisfaisant*

$$\mathbb{P}(|X| > x) \leq K \exp(-Cx^p).$$

Alors

$$\|X\|_{\psi_p} \leq \left(\frac{(1+K)}{C} \right)^{1/p}.$$

Il suffit donc de borner $\mathbb{P}(|X_u - X_s| > x)$ pour obtenir une majoration de $\|X_u - X_s\|_{\psi_p}$. Cette majoration s'obtient à partir d'inégalités exponentielles pour les processus de Poisson (l'inégalité de Chernoff qui relie la probabilité de grande déviation à la transformée de Laplace). On en déduit que la distance adaptée au problème est

$$\delta(t, \mu, t', \mu') = \|f_{t, \mu} - f_{t', \mu'}\|_{\psi_1, \lambda}^{1/\eta},$$

où $f_{t, \mu}(x) = \lambda_0 + \mu \mathbf{1}_{x > t}$.

B.2.3 Calcul du nombre de recouvrement

Le calcul du nombre de recouvrement fait apparaître que

$$N(\varepsilon, \delta) = C(\varepsilon)T,$$

où $C(\varepsilon)$ ne dépend pas de T . Du fait que $\psi_p^{-1}(m) = (\log(1 + m))^{1/p}$, on en déduit (B.1).

B.2.4 Passage au cas de ruptures multiples

Ce passage peut se faire aisément en raisonnant par additivité. Un processus de Poisson à k ruptures peut être vu comme la somme de plusieurs processus de Poisson indépendants possédant au plus une rupture. Ceci permet de décomposer la vraisemblance dans le cas du modèle à rupture multiples en une somme de vraisemblances correspondant chacune à un modèle à une rupture, et on en déduit le résultat du Théorème 2.

Annexe C

Détection de ruptures multiples dans le nombre de sinistres

Dans cette annexe, nous présentons les résultats détaillés qui conduisent au découpage de l'année en quatre périodes de sinistralité comparable (voir Chapitre 3). Dans la section C.2, nous présentons quelques graphiques qui montrent le phénomène de croissance de la vraisemblance lorsque le nombre de ruptures augmente. La section C.1 fournit les résultats détaillés de la sélection multiple, garantie par garantie, et année par année.

C.1 Résultats détaillés de la sélection multiple

Nous présentons ici les estimations effectuées lors de la segmentation de l'année en 4 périodes de risque similaire. Nous ne présentons que les résultats d'estimation liés aux modèles avec 3, 4 et 5 ruptures puisque ce sont les seuls à être sélectionnés.

C.1.1 Garantie Vol

Année	Nb ruptures	Dates de ruptures	Log-vraisemblance	Pen. forte (c_1)	Pen. faible (c_2)
1994	3	13 27 44	-7.1541e+5	-1.6221e+6	-1.3528e+6
	4	15 17 27 44	-6.2818e+5	-1.8371e+6	-1.4780e+6
	5	3 13 27 34 44	-2.7084e+5	-1.7820e+6	-1.3331e+6
1995	3	4 27 44	-1.0692e+6	-2060460	-1828140
	4	3 6 27 44	-8.7906e+5	-2200740	-1890980
	5	2 13 15 27 44	-4.7955e+5	-2131650	-1.7444e+6
1996	3	9 27 44	-7.3864e+5	-1289386	-1285042
	4	9 27 30 38	-5.7372e+5	-1308048	-1302256
	5	9 27 30 32 44	-4.6838e+5	-1386290	-1379050

Année	Nb ruptures	Dates de ruptures	Log-vraisemblance	Pen. forte (c_1)	Pen. faible (c_2)
1997	3	8 26 47	-8.8469e+5	-1609490	-1556210
	4	8 26 30 37	-5.6864e+5	-1535040	-1464000
	5	8 26 31 34 41	-4.0911e+5	-1617110	-1528310
1998	3	14 27 43	-5.9618e+5	-1.2000e+6	-1.1723e+6
	4	4 14 27 43	-4.3928e+5	-1.2444e+6	-1.2075e+6
	5	9 20 22 27 43	-3.3011e+5	-1.3365e+6	-1.2903e+6
1999	3	17 27 44	-6.6169e+5	-1.3507e+6	-1.2715e+6
	4	17 22 27 44	-5.6178e+5	-1.4804e+6	-1.3748e+6
	5	7 9 15 27 44	-3.6091e+5	-1.5092e+6	-1.3772e+6
2000	3	13 27 44	-6.0896e+5	-1.3681e+6	-1.2311e+6
	4	4 11 27 44	-5.3136e+5	-1.5436e+6	-1.3609e+6
	5	4 15 27 31 38	-2.9065e+5	-1.5559e+6	-1.3275e+6
2001	3	10 27 44	-6.8402e+5	-1.4302e+6	-1.2562e+6
	4	2 10 27 44	-6.2120e+5	-1.6162e+6	-1.3841e+6
	5	12 28 22 27 44	-3.4614e+5	-1.5898e+6	-1.2997e+6
2002	3	13 27 41	-5.3727e+5	-1.0996e+6	-1.0828e+6
	4	3 10 27 44	-5.6267e+5	-1.3124e+6	-1.2901e+6
	5	2 19 27 30 37	-4.8230e+5	-1.4195e+6	-1.3916e+6
2003	3	9 27 40	-5.5434e+5	-1.0104e+6	-9.2356e+5
	4	4 9 27 39	-4.8935e+5	-1.0974e+6	-9.8164e+5
	5	2 4 16 27 39	-3.2975e+5	-1.0898e+6	-9.4511e+5
2004	3	9 27 44	-5.5608e+5	-1.0425e+6	-1.1178e+6
	4	9 22 27 44	-3.5505e+5	-1.0036e+6	-1.1040e+6
	5	2 13 24 27 44	-3.2594e+5	-1.1366e+6	-1.2621e+6
2005	3	10 27 44	-4.9102e+5	-867028	-756634
	4	3 19 27 44	-4.0209e+5	-903434	-756242
	5	10 23 27 30 46	-2.7335e+5	-900030	-716040
2006	3	10 27 44	-4.2098e+5	-8.8000e+5	-7.3712e+5
	4	4 22 27 43	-3.9120e+5	-1.0032e+6	-8.1271e+5
	5	10 20 22 27 43	-2.2262e+5	-9.8765e+5	-7.4951e+5
2007	3	9 27 44	-4.1136e+5	-7.9633e+5	-6.7412e+5
	4	3 22 27 44	-3.9983e+5	-9.1312e+5	-7.5019e+5
	5	9 17 27 30 49	-2.5915e+5	-9.0076e+5	-6.9709e+5
2008	3	9 27 44	-3.5672e+5	-700214	-675350
	4	9 27 30 37	-2.9904e+5	-757032	-723880
	5	9 27 30 38 48	-2.1861e+5	-791100	-749660

C.1.2 Garantie RC Corporelle

Année	Nb ruptures	Dates de ruptures	Log-vraisemblance	Pen. forte (c_1)	Pen. faible (c_2)
1994	3	14 27 43	-2.3906e+5	-4.5076e+5	-4.3392e+5
	4	7 24 27 43	-2.2582e+5	-5.0808e+5	-4.8563e+5
	5	3 13 14 27 43	-1.8216e+5	-5.3499e+5	-5.0692e+5
1995	3	10 27 45	-2.8407e+5	-4.0711e+5	-3.8423e+5
	4	10 27 35 45	-1.7565e+5	-3.3970e+5	-3.0920e+5
	5	13 14 27 35 45	-1.4401e+5	-3.4907e+5	-3.1094e+5
1996	3	14 27 47	-3.0345e+5	-5.3661e+5	-5.1383e+5
	4	10 24 27 47	-2.6357e+5	-5.7445e+5	-5.4408e+5
	5	2 20 27 37 51	-2.1925e+5	-6.0784e+5	-5.6988e+5
1997	3	8 26 47	-3.6538e+5	-5.9854e+5	-5.7576e+5
	4	8 25 26 40	-2.7313e+5	-5.8401e+5	-5.5364e+5
	5	8 25 26 30 37	-2.4302e+5	-6.3161e+5	-5.9365e+5
1998	3	14 27 41	-2.4208e+5	-3.9061e+5	-3.4949e+5
	4	8 24 27 41	-2.1865e+5	-4.1669e+5	-3.6186e+5
	5	14 27 35 46 51	-1.6524e+5	-4.1279e+5	-3.4425e+5
1999	3	13 27 43	-2.6420e+5	-4.4019e+5	-4.1559e+5
	4	4 13 27 43	-2.1541e+5	-4.5006e+5	-4.1726e+5
	5	4 13 27 34 43	-1.2114e+5	-4.1445e+5	-3.7345e+5
2000	3	14 27 41	-2.3823e+5	-3.2235e+5	-3.0030e+5
	4	10 24 27 41	-1.9573e+5	-3.0788e+5	-2.7848e+5
	5	1 7 14 27 41	-1.6261e+5	-3.0280e+5	-2.6605e+5
2001	3	13 27 45	-2.6595e+5	-4.9378e+5	-4.6474e+5
	4	14 27 35 51	-2.3877e+5	-5.4255e+5	-5.0382e+5
	5	12 14 22 27 45	-1.7941e+5	-5.5913e+5	-5.1072e+5
2002	3	11 26 41	-2.1738e+5	-3.6588e+5	-3.8637e+5
	4	7 24 27 42	-2.0330e+5	-4.0130e+5	-4.2862e+5
	5	1 3 19 27 42	-1.8943e+5	-4.3693e+5	-4.7108e+5
2003	3	14 26 40	-1.8569e+5	-3.2278e+5	-3.3664e+5
	4	7 24 27 40	-1.7331e+5	-3.5611e+5	-3.7459e+5
	5	1 6 10 27 40	-1.7273e+5	-4.0122e+5	-4.2432e+5
2004	3	14 27 40	-1.8280e+5	-3.6414e+5	-4.7773e+5
	4	14 27 40 51	-1.7146e+5	-4.1324e+5	-4.6639e+5
	5	6 17 27 35 51	-1.3770e+5	-4.3993e+5	-4.3263e+5
2005	3	9 27 42	-2.1051e+5	-3.2882e+5	-2.9731e+5
	4	9 24 27 42	-1.6127e+5	-3.1902e+5	-2.7702e+5
	5	3 10 20 27 42	-1.0700e+5	-3.0419e+5	-2.5168e+5

Année	Nb ruptures	Dates de ruptures	Log-vraisemblance	Pen. forte (c_1)	Pen. faible (c_2)
2006	3	14 27 41	-1.8473e+5	-3.1151e+5	-2.8618e+5
	4	14 27 41 49	-1.5794e+5	-3.2698e+5	-2.9321e+5
	5	7 14 27 41 49	-1.1145e+5	-3.2275e+5	-2.8054e+5
2007	3	14 27 40	-1.7349e+5	-3.0717e+5	-2.9814e+5
	4	2 12 27 40	-1.5751e+5	-3.3574e+5	-3.2371e+5
	5	2 12 27 40 49	-1.2996e+5	-3.5275e+5	-3.3771e+5
2008	3	8 24 49	-3.0020e+5	-5.0494e+5	-4.7239e+5
	4	8 24 27 49	-2.3431e+5	-5.0730e+5	-4.6390e+5
	5	18 25 27 35 51	-1.7748e+5	-5.1872e+5	-4.6447e+5

C.1.3 Garantie RC Matérielle

Année	Nb ruptures	Dates de ruptures	Log-vraisemblance	Pen. forte (c_1)	Pen. faible (c_2)
1994	3	13 27 40	-9.9958e+5	-2.2140e+6	-2.1257e+6
	4	14 27 35 48	-7.4801e+5	-2.3673e+6	-2.2495e+6
	5	8 12 27 35 48	-5.0437e+5	-2.5285e+6	-2.3812e+6
1995	3	8 27 45	-1.4298e+6	-2.5863e+6	-1.9321e+6
	4	6 13 21 45	-1.2952e+6	-2.8372e+6	-1.9650e+6
	5	2 13 27 35 47	-4.7206e+5	-2.3996e+6	-1.3093e+6
1996	3	8 27 48	-1.7150e+6	-3.0399e+6	-2.9658e+6
	4	8 27 37 44	-7.5593e+5	-2.5224e+6	-2.4236e+6
	5	8 27 37 46 48	-6.8609e+5	-2.8942e+6	-2.7707e+6
1997	3	9 23 39	-1.1017e+6	-2.1001e+6	-1.9273e+6
	4	8 13 27 39	-7.0619e+5	-2.0374e+6	-1.8069e+6
	5	3 16 27 34 40	-4.0156e+5	-2.0656e+6	-1.7775e+6
1998	3	14 27 49	-1.8016e+6	-3.0278e+6	-2.6219e+6
	4	14 27 35 51	-1.1237e+6	-2.7586e+6	-2.2174e+6
	5	14 27 35 45 51	-5.4376e+005	-2.5874e+6	-1.9109e+6
1999	3	18 27 40	-1.2093e+6	-1227540	-802800
	4	8 18 27 45	-7.4895e+5	-1636720	-1070400
	5	8 18 27 35 45	-9.1741e+4	-2045900	-1338000
2000	3	8 27 40	-1.2681e+6	-1184820	-1399260
	4	4 8 27 37	-1.2133e+6	-1579760	-1865680
	5	4 7 8 27 37	-1.1886e+6	-1974700	-2332100

Année	Nb ruptures	Dates de ruptures	Log-vraisemblance	Pen. forte (c_1)	Pen. faible (c_2)
2001	3	14 27 40	-1.0211e+6	-1105500	-931320
	4	14 27 34 47	-7.4503e+5	-1474000	-1241760
	5	8 14 27 34 47	-3.3714e+5	-1842500	-1552200
2002	3	11 27 42	-1.0422e+6	-971280	-956700
	4	11 24 27 42	-7.3260e+5	-1295040	-1275600
	5	33 11 27 38 42	-5.1594e+5	-1618800	-1594500
2003	3	5 27 40	-2.3726e+6	-1567140	-1037100
	4	5 27 30 34	-1.5816e+6	-2089520	-1382800
	5	5 19 27 30 34	-7.6879e+5	-2611900	-1728500
2004	3	8 27 40	-1.1244e+6	-359616	-234312
	4	8 23 27 40	-6.8009e+5	-479488	-312416
	5	5 15 27 35 50	-4.8388e+5	-599360	-390520
2005	3	9 27 48	-1.5355e+6	-3.1389e+6	-2.8491e+6
	4	9 27 31 34	-1.2223e+6	-3.3601e+6	-2.9737e+6
	5	9 27 31 34 46	-6.8144e+5	-3.3536e+6	-2.8706e+6
2006	3	8 27 40	-1.1059e+6	-1096860	-959880
	4	4 8 27 40	-9.9497e+5	-1462480	-1279840
	5	15 21 27 35 51	-7.1823e+5	-1828100	-1599800
2007	3	8 27 40	-1.1286e+6	-1561140	-1752540
	4	8 27 40 51	-1.0501e+6	-2081520	-2336720
	5	2 8 27 40 51	-9.6921e+5	-2601900	-2920900
2008	3	8 26 49	-2.7384e+6	-3.4617e+6	-3.2496e+6
	4	8 24 27 49	-1.4441e+6	-3.5956e+6	-3.3129e+6
	5	19 24 27 35 51	-1.0923e+6	-3.7817e+6	-3.4283e+6

C.2 Heuristique de pente

Nous utilisons l'heuristique de pente (Birgé, Massart 2004) pour calibrer la pénalité que nous utilisons pour sélectionner le nombre de ruptures adaptés à notre problème (voir Chapitre 3 pour plus de détails). A la base de cette heuristique, on suppose que, à partir d'un certain nombre de paramètres, la vraisemblance croît de façon quasi-linéaire. En complément de la figure 3.2.2, nous reproduisons quelques exemples supplémentaires, qui se caractérisent toutes par une croissance plus lente et linéaire à partir de $k = 4$ ou $k = 5$ ruptures.

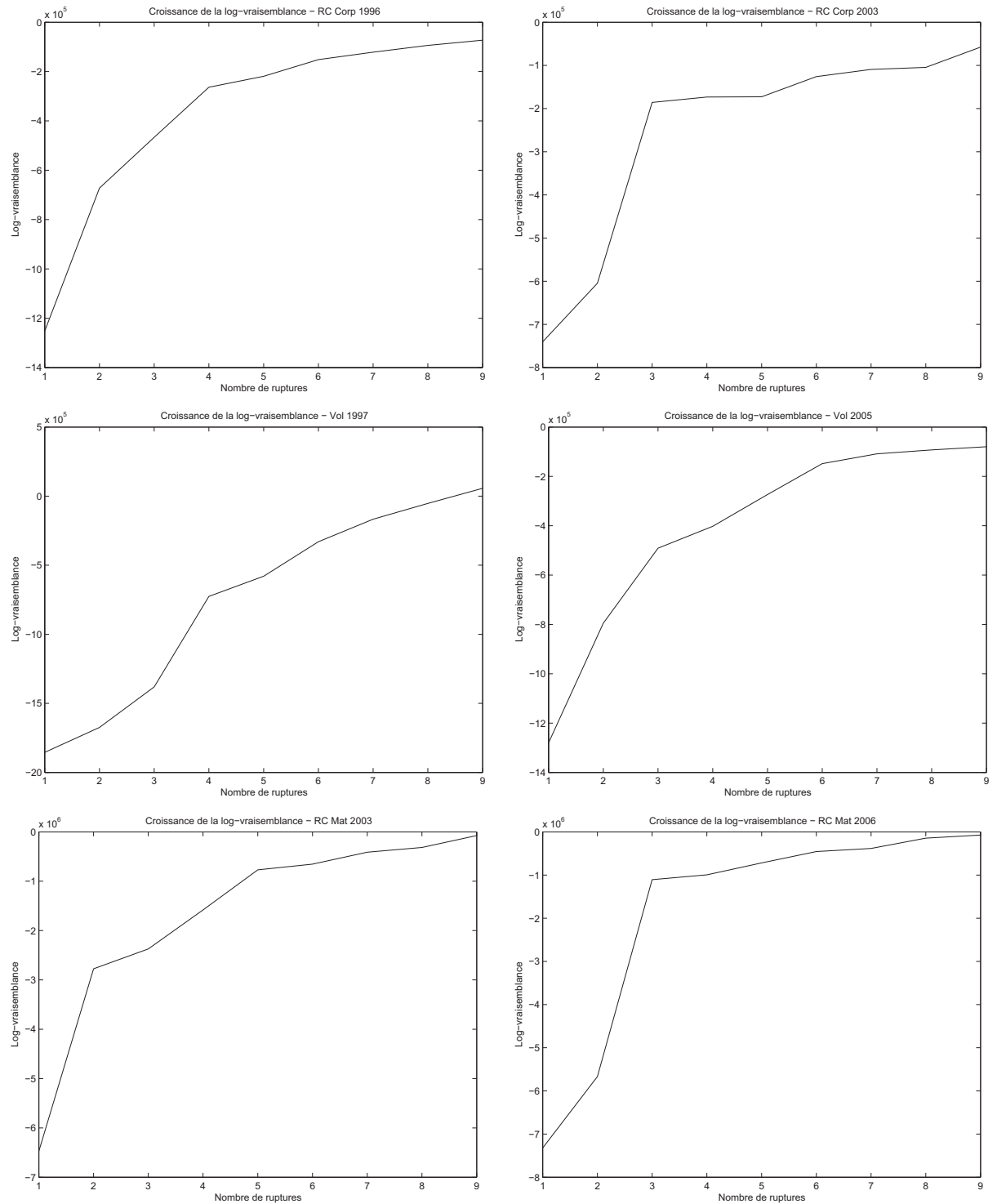


FIGURE C.1 – Illustration du phénomène de croissance de la vraisemblance.

Annexe D

Exemples de détection dynamique de rupture et données complémentaires

Dans cette annexe, nous présentons quelques détails de détection dynamique de rupture pour les séries des coûts moyens de sinistre (voir Chapitre 4), en insistant sur la représentation graphique de la détection au voisinage des ruptures. Ces résultats sont présentés dans la section D.1. La section D.2 fournit les modèles que nous retenons pour la modélisation des coûts de sinistre pour ce qui concerne les mois d'été et la garantie Vol (on rappelle que les mois d'été ont dû être traités séparément du fait de la présence de saisonnalité).

D.1 Détails de la détection dynamique de rupture

Nous nous focalisons sur la garantie Vol. Les graphiques de détection dynamique ont été présentés dans le cœur du mémoire (Chapitre 4) pour l'approche bootstrap basée sur la vraisemblance gaussienne. Nous présentons donc ceux qui concernent les deux autres méthodes.

D.1.1 Modélisation par Pareto

On rappelle que trois ruptures sont détectées.

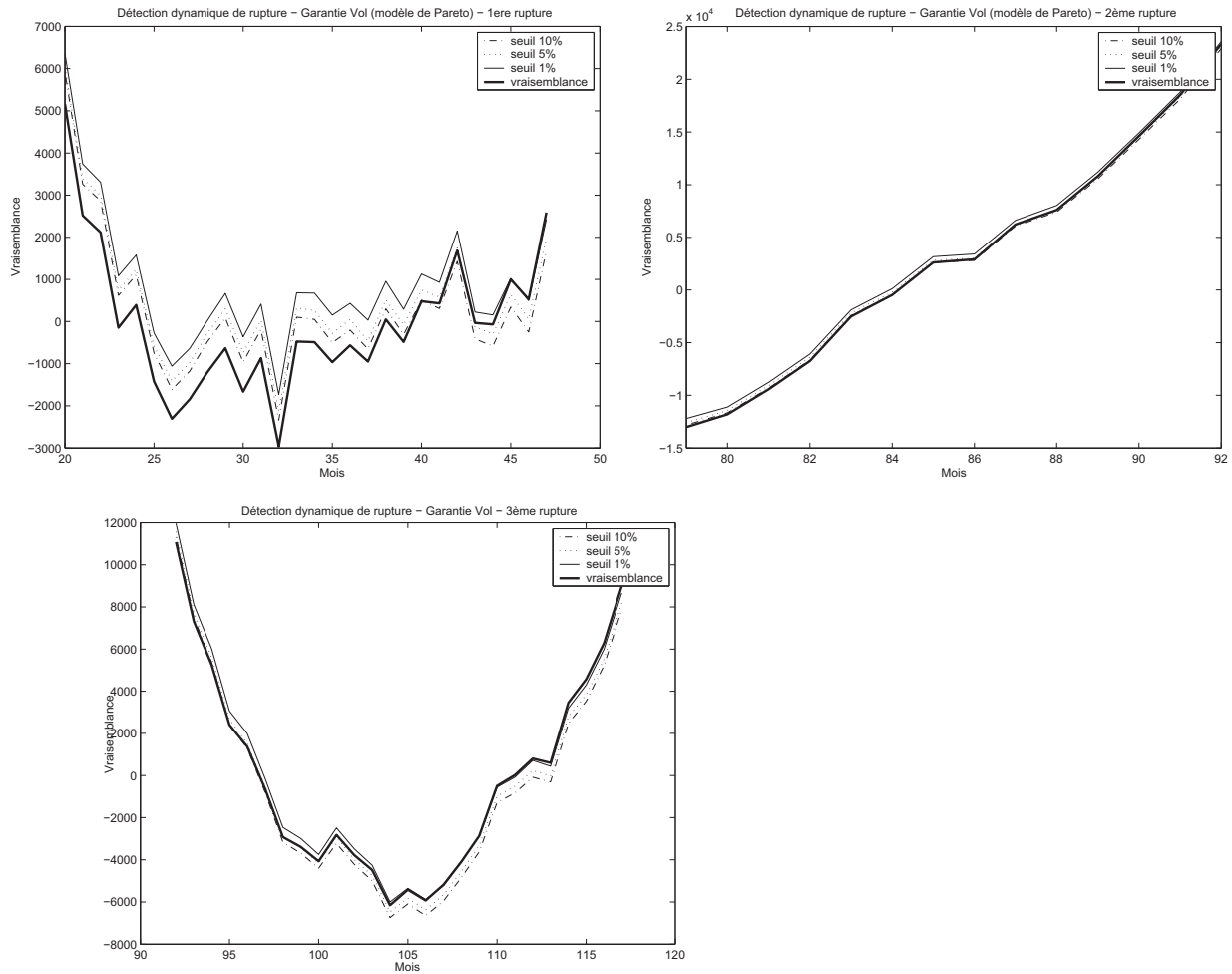


FIGURE D.1 – Garantie Vol, modélisation par Pareto. Détection dynamique au voisinage des différentes ruptures.

D.1.2 Modélisation des petits sinistres

On rappelle que cinq ruptures sont détectées. Nous ne représentons ici que les graphiques correspondant aux ruptures 1, 2, 3 et 5, la 4ème rupture étant présentée dans le corps du Chapitre 4.

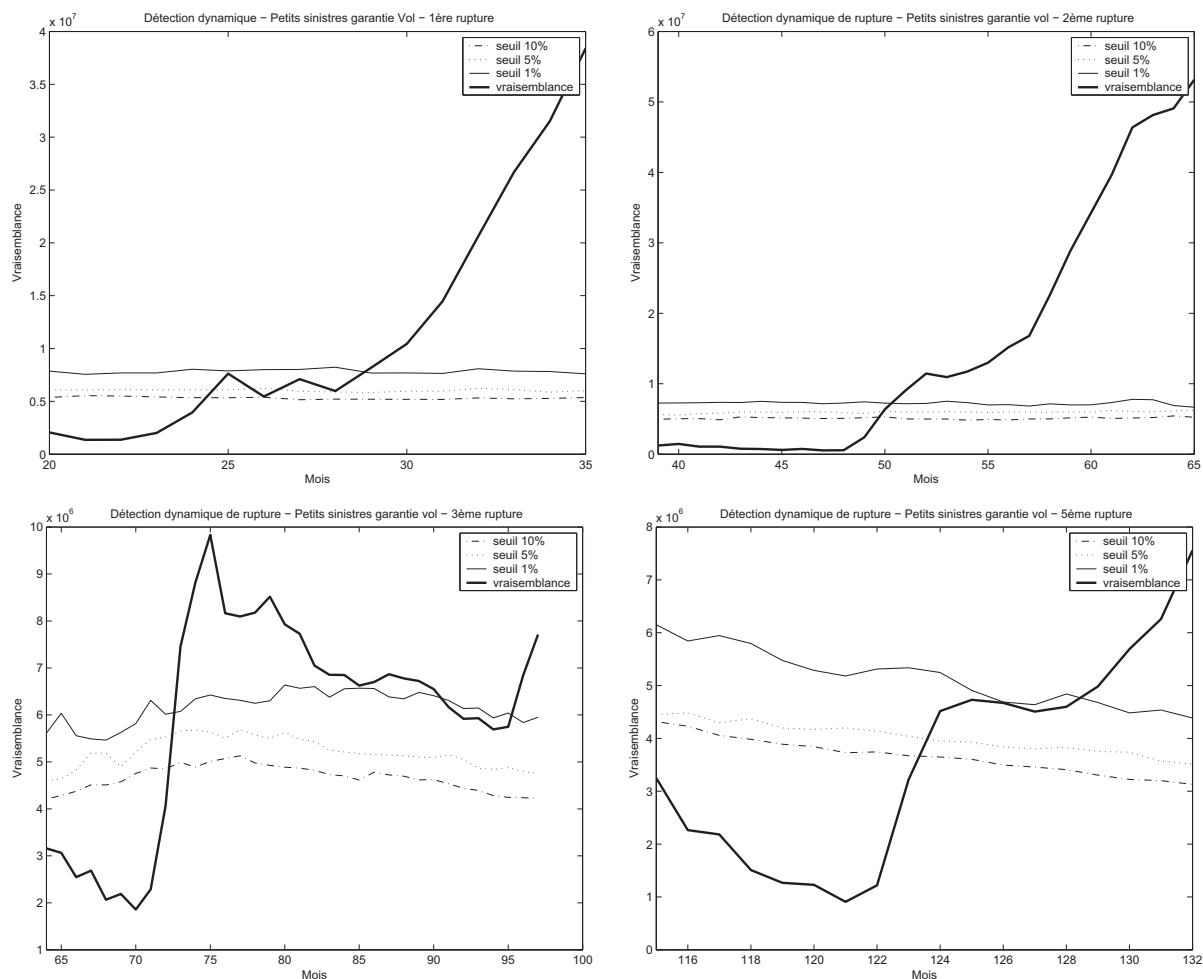


FIGURE D.2 – Garantie Vol. Petits sinistres. Détection dynamique au voisinage des différentes ruptures.

D.1.3 Modélisation des sinistres importants

On rappelle que deux ruptures sont détectées. La détection de la deuxième rupture intervient extrêmement tôt après la détection de la première, ce qui rend la représentation graphique de la détection de faible intérêt, aussi nous ne représentons que le comportement de la procédure au voisinage de la première rupture. On remarque, ainsi qu'il l'a été indiqué au Chapitre 4 une absence de détection au seuil 1%.

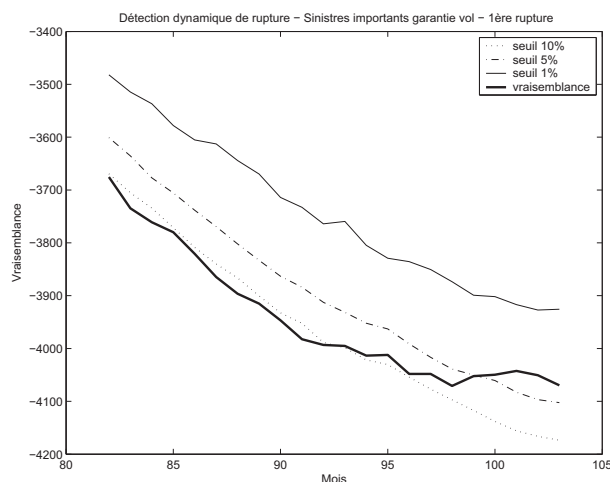


FIGURE D.3 – Garantie Vol. Sinistres importants. Détection dynamique au voisinage des différentes ruptures.

D.2 Modélisation des mois d'été pour la garantie Vol

Nous donnons ici, à titre indicatif, les modèles que nous avons retenus pour modéliser l'évolution des coûts de sinistres correspondant à la période d'été. Ces résultats ne font intervenir aucune détection de rupture (on peut mettre en œuvre les techniques présentées, mais elles n'aboutissent à aucune détection), et se résument à des modèles de régression linéaire simple. Pour chacun des mois considérés, X_i désigne le coût moyen l'année i .

Mois de juin, modèle retenu :

$$X_i = 30.854i + 1118.6.$$

Mois de juillet, modèle retenu :

$$X_i = 27.107i + 1100.1.$$

Mois d'août, modèle retenu :

$$X_i = 23.871i + 1034.2.$$

Bibliographie

- [1] Basseville, M., Nikiforov, I.V. (1993), *Detection of Abrupt Changes - Theory and Application*, Prentice-Hall.
- [2] Birgé, L., Massart, P. (2007), *Minimal penalties for Gaussian model selection*, Probab. Theory Related Fields, **138** p.33–73.
- [3] Davison, A.C., Hinkley, D.V. (1997) *Bootstrap Methods and their Application*, Cambridge University Press.
- [4] Faugeras, O., Lopez, O. (2010) *Multiple change-point detection for Poisson processes with applications to insurance*, disponible auprès des auteurs.
- [5] Golubev, Y., Spokoiny, V.(2009) *Exponential bounds for minimum contrast estimators*, Electron. J. Statist., **3** p.712–746.
- [6] Grama, I., Spokoiny, V. (2008) *Statistics of extremes by oracle estimation*, Ann. Statist., **36** p.1619–1648.
- [7] Haccou, P., Meelis, E., Van de Geer, S. (1988) *The likelihood ratio test for a change point problem for exponentially distributed random variables*, Stochastic Processes and Applications, **27** p.121–139
- [8] Lebarbier, E. (2005) *Detecting Multiple Change-Points in the Mean of Gaussian Process by Model Selection*, Signal Processing, **85** p.717–736.
- [9] Mack, T. (1999) *The standard error of Chain Ladder Reserve estimates : recursive calculation and inclusion of a tail factor*, ASTIN Bulletin **29** p.361–366.
- [10] Van der Vaart, A. W., Wellner, J. (1996) *Weak convergence and empirical processes*, Springer-Verlag, New York.