



# Modélisation de la sévérité des risques opérationnels par les lois $G\&H$

Mémoire de recherche

Septembre 2012

**Anass TALOUH**

ISFA & Ecole centrale de Lyon : Master 2 Graf-Gestion des risques en assurance et en finance.

Ensimag : MIF -Mathématique Informatique pour la Finance

Contact : [anass.talouh@gmail.com](mailto:anass.talouh@gmail.com)

## Table des matières

<b>I-ENVIRONNEMENT ET CONTEXTE</b>	<b>4</b>
1. LA REGLEMENTATION BANCAIRE	4
1.1 Bâle I	4
1.2 Bâle II	5
1.3 Bâle III	6
2. LE RISQUE OPERATIONNEL	9
2.1 Définition	9
2.2 Les approches réglementaires	10
3. LA METHODE LDA	12
4. LA DISTRIBUTION DE SEVERITE DES PERTES	14
4.1 Propriétés souhaitées en risque opérationnel	14
4.2 Distribution de probabilité à queue épaisse	15
4.3 Contraintes spécifiques	15
4.4 Tests d'adéquation	16
Lois de probabilité	18
<b>II- MODELISATION DE LA SEVERITE PAR DES DISTRIBUTIONS <math>\mathcal{G}\&amp;\mathcal{H}</math></b>	<b>20</b>
1. DEFINITION DES LOIS $\mathcal{G}\&\mathcal{H}$	20
Introduction	20
1.1 Asymétrie et lois $\mathcal{G}$	20
1.2 Aplatissement et lois $\mathcal{H}$	21
1.3 Lois $\mathcal{G}\&\mathcal{H}$	22
2. ESTIMATION DES PARAMETRES	27
2.1 Prise en compte du seuil de collecte	27
2.2 Méthode Inter-quantile (IQ)	28
2.3 Méthode du maximum de vraisemblance (ML)	29
2.4 Méthode des moments généralisée (GMM)	30
2.5 Méthode de la distance-quantile (QD)	31
2.6 Propriétés théoriques des méthodes d'estimation	33
3. DEPLOIEMENT DU MODELE EN ENVIRONNEMENT REEL	39
3.1 Statistiques descriptives des pertes	40
3.2 Estimation des paramètres	40
3.3 Critères d'ajustement	41
3.4 Charges en capital	46
<b>IV- CONCLUSION</b>	<b>48</b>
<b>V- BIBLIOGRAPHIE</b>	<b>49</b>
<b>VI- ANNEXES :</b>	<b>50</b>
1. RECIPROQUE DE LA FONCTION $Y$ DE TUKEY	50
1.1 Présentations des algorithmes :	50
1.2 Méthodologie	56
2. LOI NORMAL-INVERSE-GAUSSIAN (NIG)	58
2.1 Loi Inverse-Gaussian (IG)	58
2.2 Loi NIG	59
3. OPTIMISATION PAR RECUIT SIMULE	61
3.1 Principe	61
3.2 L'algorithme de Métropolis	62
3.3 Optimisation itérative- probabiliste	63
4. TRAITEMENT DES PERTES ABERRANTES	64
4.1 Réduction de la variance	64
4.2 Choix du nombre de pertes à externaliser	66

## Résumé

*Bien que les lois paramétriques simples (deux paramètres au plus) soient très utilisées pour modéliser la sévérité des risques opérationnels, il arrive souvent que les données soient complexes pour ces distributions, nécessitant de ce fait l'introduction de nouvelles lois plus flexibles et capables de s'adapter aux particularités (asymétrie positive et queue épaisse...etc.) de ces pertes. La présente étude se veut une étude de loi  $G\&H$ , appartenant à une famille de distributions paramétriques dites généralisées. Nous traitons dans un premier temps les propriétés élémentaires de la famille  $G\&H$  dans un contexte univarié, puis nous discuterons la manière la plus adéquate d'estimer ses paramètres. Enfin nous comparons, sur données réelles, avec les résultats de la loi log-normale utilisée actuellement par la majorité des structures financières.*

## Abstract

*Although simple parametric distributions (at most two parameters) are widely used to model the severity of operational risk, the data is often 'too severe' for these distributions, thus requiring the introduction of new distributions more flexible and able to adapt to the characteristics (positive skewness and heavy tail ... etc.) of those losses. The present study introduces this family of distribution in the operational risk context. We discuss the basic properties of the  $G\&H$  family, and then we discuss the most appropriate way to estimate its parameters. Finally we make a comparison on real data with the results of the lognormal distribution currently used by the majority of the financial structures.*

# I-Environnement et contexte

---

## 1. La réglementation bancaire

La réglementation prudentielle a considérablement évolué ces vingt dernières années sous l'impulsion des travaux du Comité de Bâle (*Basel Committee on Banking Supervision*, BCBS) [1]. Même si celui-ci n'a aucun pouvoir décisionnel, ses recommandations sont reprises par les autorités de tutelle des différents pays concernés. En Europe, c'est la Commission Européenne qui est chargée de définir les règles prudentielles CRD (pour *Capital Requirement Directive*), dont la mise en œuvre revient aux différentes autorités nationales de tutelle. En France, c'est l'Autorité de Contrôle Prudentiel (ACP), qui contrôle l'application de la CRD, une fois que celle-ci a été adaptée en droit français par le Comité de la Réglementation Bancaire et Financière.

### Historique

Les grandes étapes de l'évolution de ces directives prudentielles sont répertoriées ci-dessous :

- 1988 Publication de « The Basel Capital Accord » définissant le ratio Cooke (Bâle I).
- 1993 Élaboration de la CAD par la Commission Européenne.
- 1996 Incorporation du risque de marché dans l'assiette des risques.
- 1999 Premier document consultatif sur le nouveau ratio McDonough.
- 2004 Publication du texte définitif de la réglementation Bâle II.
- 2006 Mise en place de la norme Bâle II.
- 2010 Publication du texte définitif de la nouvelle réglementation Bâle III.
- 2013 Mise en œuvre programmée de la norme Bâle III.

### 1.1 Bâle I

En 1988, le Comité de Bâle propose un ratio international de solvabilité qui doit permettre une adéquation des fonds propres par rapport aux risques, tout en renforçant la solidité et la stabilité du système bancaire et également d'atténuer les inégalités concurrentielles entre les banques. Il s'agit du ratio Cooke (du nom du président du Comité de Bâle de l'époque) qui correspond au rapport entre le montant des fonds propres et celui des encours pondérés de crédit (*EPC*). Plusieurs niveaux de fonds propres sont définis :

1. les fonds propres de base *FP1 (Tier one)*, qui correspondent au capital et aux réserves.
2. les fonds propres complémentaires *FP2 (Tier two)*, principalement constitués d'emprunts subordonnés.
3. les fonds propres sur complémentaires *FP3 (Tier three)*.

Selon l'accord de Bâle I, les établissements financiers doivent alors respecter les contraintes  $FP1 \leq FP2$  et  $FP1/EPC \leq 4\%$ , mais également le ratio de Cooke défini par :

$$\frac{FP1 + FP2}{EPC} \geq 8\%$$

La réglementation a ensuite évolué progressivement pour prendre en compte les risques de marché, avec la proposition de deux approches, l'une forfaitaire, et l'autre autorisant les banques à utiliser un modèle interne. L'idée est d'inciter les banques à construire des modèles robustes pour évaluer leurs risques de marché et donc d'obtenir des exigences de fonds propres plus réalistes. Ce n'est qu'en janvier 1996 que le Comité de Bâle propose d'incorporer le risque de marché dans le ratio Cooke. En

1999, le Comité de Bâle publie un premier document consultatif pour réformer le ratio Cooke, puis un second en 2001, avec pour objectifs la modification du traitement du risque de crédit et l'intégration du risque opérationnel.

## 1.2 Bâle II

### 1.2.1 Définition

Le Comité de Bâle publie finalement en juin 2004 le Nouvel Accord de Bâle, résultat de plusieurs périodes consultatives auprès des institutions financières. La première motivation de l'Accord est la modification de l'assiette des risques, qui prévoit désormais la prise en compte du risque opérationnel, et la définition d'un nouveau ratio de solvabilité, le ratio McDonough :

$$R_{MD} = \frac{FP1 + FP2}{ERC + ERO + ERM} \geq 8\%$$

où *ERC*, *ERO* et *ERM* sont les exigences définies respectivement pour le risque de crédit, le risque opérationnel et le risque de marché. La seconde motivation de l'Accord est de rendre l'exigence de fonds propres plus sensible au risque réel encouru par la banque. L'idée est d'autoriser les banques, sous certaines conditions, à recourir à des modèles internes pour mesurer le risque de crédit et le risque opérationnel, de la même façon que pour le risque de marché. Cette réforme s'appuie sur trois piliers, représentés sur la figure 2.1

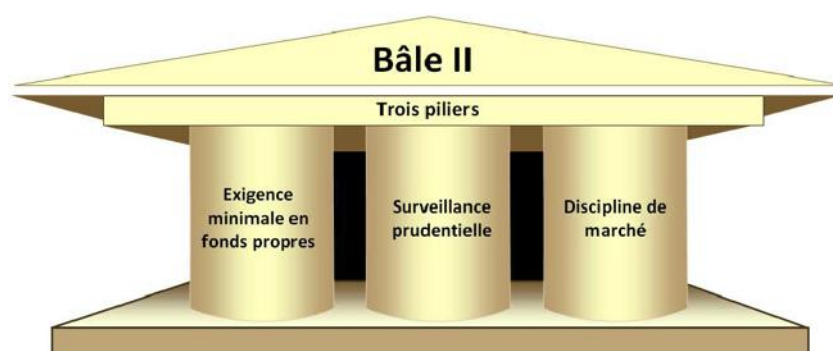


FIGURE 2.1 – Les trois piliers Bâle II

Le premier pilier correspond aux exigences minimales de fonds propres et vise à définir le ratio McDonough. Le deuxième pilier concerne le processus de surveillance et donne un pouvoir accru aux autorités de contrôle (ACP). Le troisième pilier porte pour sa part sur la communication financière et la discipline de marché, et vise à améliorer la transparence financière des banques pour permettre à des tiers d'apprécier l'adéquation de leurs fonds propres. Nous détaillons dans les sections suivantes les deux premiers piliers, auxquels se rapporte ce mémoire.

### Le pilier 1 : une exigence de fonds propres

Le pilier 1 définit les exigences minimales en fonds propres. Il vise à assurer une couverture minimale, par les fonds propres, des risques de crédit, de marché et opérationnel via l'établissement du ratio de McDonough défini précédemment. Bâle II tient non seulement à un éventail élargi de risques couverts (le risque opérationnel n'était pas implicitement appréhendé par le ratio de Cooke,

celui de Bâle I), mais également à la possibilité pour les établissements bancaires de choisir différents niveaux de sophistication pour le calcul des exigences en fonds propres.

Ainsi, il leur est possible, pour le risque de crédit et le risque opérationnel, de recourir soit à des pondérations forfaitaires graduées en fonction de la qualité de la contrepartie, soit à des notations internes.

Le travail effectué au cours du stage se place dans le cadre du pilier 1.

### 1.2.2 Le pilier 2 : pour un meilleur suivi des risques

Le processus de surveillance prudentielle (*Supervisory Review Process*) relatif au Pilier 2 vise à inciter les établissements à développer leurs techniques de gestion des risques et des fonds propres. Il est aussi voué à permettre aux autorités de contrôler que les banques disposent d'un niveau de fonds propres conforme à leur profil de risques et, à défaut, d'engager des mesures correctrices. Il comprend trois composantes essentielles :

1. Le dispositif de contrôle interne (*Internal Governance*), incluant notamment le contrôle des opérations et des procédures internes, l'organisation comptable et de traitement de l'information, les systèmes de mesure des risques, la surveillance des flux d'espèces et de titres. Ce dispositif doit être adapté à la taille, à la nature et au volume des activités de l'établissement.
2. Le processus d'évaluation de l'adéquation du capital interne (ICAAP, pour *Internal Capital Adequacy Assessment Process*), mis en place par les établissements, qui a vocation à leur permettre de correctement identifier, mesurer, agréger et surveiller les risques, et d'assurer la détention d'un montant de capital interne en adéquation avec leur profil de risque
3. Le processus de surveillance et d'évaluation prudentielle (SREP, pour *Supervisory Review and Evaluation Process*), mis en œuvre par les autorités de contrôle, qui prend en compte la nature et la complexité de l'activité des établissements, peut conduire à des mesures prudentielles. Il consiste en l'évaluation du profil de risque des établissements, la vérification de l'adéquation et de la robustesse des dispositifs de contrôle interne et du processus d'évaluation de l'adéquation du capital et la vérification de l'adéquation du niveau des fonds propres au profil de risque.

### Le pilier 3 : discipline du marché

Le pilier 3 concerne la discipline de marché. Il vise à améliorer la transparence financière des banques, en leur imposant de communiquer les informations nécessaires pour permettre à des tiers d'apprécier l'adéquation de leurs fonds propres. Une meilleure discipline de marché en est espérée. Les trois volets forment un tout indissociable.

## 1.3 Bâle III

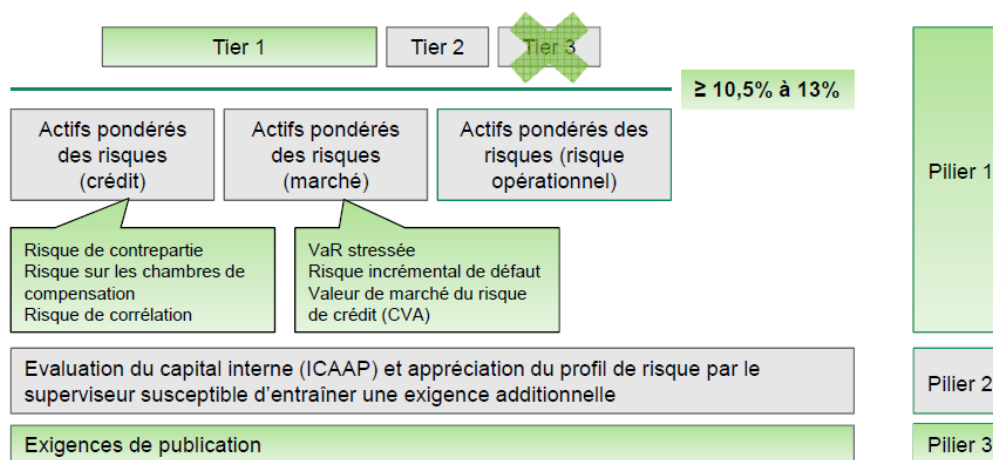
### 1.3.1 Les raisons de la réforme

La crise financière de 2007/2008 a eu des conséquences énormes pour la société en termes de destruction de richesses, d'augmentation du niveau de la dette publique et de la hausse du chômage. Elle a également démontré que les fonds propres des banques étaient insuffisants. Dans un contexte de diminution quasi continue des fonds propres des banques au fil du siècle dernier, les données empiriques, les études universitaires et plusieurs personnalités de premier plan arrivent à la conclusion que les fonds propres des banques doivent être beaucoup plus élevés. On peut résumer les raisons qui ont conduit à adopter la réforme en trois points :

- Des fonds propres (en particulier Tier 1) insuffisants et de qualité variable:
- Un effet de levier important
- Une réglementation insuffisante du risque de liquidité (exemple de Dexia)

### 1.3.2 Le contenu des évolutions réglementaires

Bâle III garde la même structure, en trois piliers, de Bâle II. Néanmoins elle apporte des modifications en termes de renforcement de la quantité et la qualité des fonds propres



### Mise en place de ratio de liquidité: LCR

Le **LCR** (*Liquidity Coverage Ratio*) est un ratio à un mois qui a pour but de permettre aux banques de résister à des crises de liquidité aiguës sur une durée d'un mois. Son principe est le suivant : les actifs liquides doivent être supérieures aux sorties de trésorerie générées par la perte des possibilités de refinancement sur le marché, par la fuite des dépôts, et par une série d'autres facteurs qui peuvent advenir lors d'une telle crise de liquidité (tirages de lignes hors-bilan, fuite de liquidité liée aux collatéraux...).

$$LCR = \frac{\text{Actis liquides}}{\text{Sorties de trésorerie}} \geq 100\%$$

### Mise en place de ratio de liquidité (NSFR)

Le **NSFR** (*Net Stable Funding Ratio*) est un ratio à un an qui a pour but de permettre aux banques de résister à un scénario de stress d'un an spécifique à l'établissement. Son principe est le suivant : le montant des ressources disponibles (*available stable funding*) doit être supérieur au montant des besoins en ressources stables (*required stable funding*).

$$NSFR = \frac{\text{Eléments fournissant un financement stable}}{\text{Eléments nécessitant un financement stable}} \geq 100\%$$

Les ratios LCR et NSFR sont calculés par les banques et reportés à l'ACP. Cette dernière complète son analyse de la liquidité par des outils de suivi complémentaires.

### Mise en place d'un ratio de levier LR

Le **LR** (Leverage ratio) est un ratio simple destiné à limiter l'effet de levier et à constituer un élément de comparaison avec le ratio de solvabilité (évite le risque de modèle). Son principe est le suivant : les fonds propres Tier1 doivent être supérieurs à 3% des actifs et engagements de hors bilan bruts (non pondérés).

$$LR = \frac{\text{Tier 1}}{\text{Actifs non pondérés}} \geq 3\%$$

### 1.3.3 Synthèse des évolutions prudentielles

Nous pouvons dresser une synthèse des principales évolutions que les plus grandes banques européennes ont connu ainsi qu'un calendrier provisoire concernant l'adoption d'une première mesure de Bâle III.

Réformes du Comité de Bâle	Réformes européennes	Thèmes	Statut CRD (adopté/non adopté)	Date d'entrée en vigueur
Bâle II	CRD 1	<ul style="list-style-type: none"><li>- Refonte du dispositif prudentiel concernant les exigences minimum de fonds propres (pilier1)</li><li>- Instauration du pilier 2 (processus de surveillance individualisé) et du pilier 3 (discipline de marché)</li></ul>	Adoptée	Fin 2006 avec option fin 2007
Bâle II.5	CRD 2	<ul style="list-style-type: none"><li>- Grands risques</li><li>- Instruments hybrides</li><li>- Transfert de risque de titrisation</li><li>- Organisation de la supervision européenne</li></ul>	Adoptée	Fin 2010
	CRD 3	<ul style="list-style-type: none"><li>- Risques de marché</li><li>- Retitrisation</li><li>- Pratiques de rémunération</li></ul>	Adoptée	Mise en vigueur pour partie fin 2010 puis fin 2011
Bâle III	CRD 4 /CRR 1	<ul style="list-style-type: none"><li>- Redéfinition des fonds propres</li><li>- Risque de contrepartie (dont CVA)</li><li>- Coussins en capital</li><li>- Ratio de levier</li><li>- Gestion de la liquidité</li></ul>	Adoptée	Début 2013



## 2. Le risque opérationnel

### 2.1 Définition

Le risque opérationnel peut être défini comme le risque de pertes résultant de procédures internes inadaptées ou défaillantes, d'erreurs humaines, de défaillance des systèmes informatiques, d'événements extérieurs. Ce sont donc des risques qui existent depuis toujours dans les banques ou les entreprises. Il peut s'agir par exemple de problèmes informatiques qui retardent l'exécution d'ordres sur les marchés (incidents courants) mais également de l'incendie de locaux ou de fraude interne pouvant entraîner de lourdes pertes. Le risque opérationnel est donc un risque qui peut s'avérer très couteux. Même s'ils sont connus depuis longtemps, la gestion des risques opérationnels est un exercice encore assez récent pour les banques, qui a largement évolué grâce à la réforme Bâle II. Le Comité de Bâle définit les risques opérationnels de manière formelle comme « *le risque de pertes dues à une inadéquation ou à une défaillance des procédures, personnels, systèmes internes ou à des événements extérieurs* ».

#### Classification

Pour répondre à la diversité des événements assimilables à des risques opérationnels et homogénéiser les données de pertes, le Comité de Bâle présente une classification des différentes catégories d'événement (EL pour *Event Line*) dont les intitulés sont regroupés dans le tableau 1.1. Ces catégories sont généralement respectées par la plupart des banques lors de la collecte de leurs pertes, dans un souci d'homogénéité et de comparaison. Il est toutefois possible que dans certaines banques la catégorie « Fraude interne » (EL2) soit différenciée entre pertes liées à la monétique (M) et les autres types de fraude externe hors monétique (HM).

Catégorie de risque	Intitulé
EL1	Fraude interne
EL2	Fraude externe
EL3	Pratiques en matière d'emploi et sécurité sur le lieu de travail
EL4	Clients, produits et pratiques commerciales
EL5	Dommages occasionnés aux actifs corporels
EL6	Interruption de l'activité et dysfonctionnement des systèmes
EL7	Exécution, livraison et gestion des processus

TABLE 1.1 – Classification des catégories de risque (ELs) pour le risque opérationnel selon Bâle II

Les différentes catégories d'événements précédentes sont observées suivant une répartition des lignes métier (*business lines*, BL) de la banque, proposée par le Comité de Bâle. Leurs intitulés figurent dans le tableau 1.2.

Ligne métier	Intitulé
BL1	Financement d'entreprise
BL2	Activités de marché
BL3	Banque de détail
BL4	Banque commerciale

BL5	Paielements et règlements
BL6	Fonctions d'agent
BL7	Gestion d'actifs
BL8	Courtage de détail

TABLE 1.2 – Classification lignes métier (BLs) pour le risque opérationnel selon Bâle II

Ces lignes métiers sont données à titre indicatif, et peuvent subir des modifications selon les différentes activités de la banque du groupe bancaire considéré.

### Exemples célèbres

Nous illustrons la définition et la classification précédente au travers d'exemple de pertes liées au risque opérationnel subies par le secteur bancaire lors de ces dernières années.

#### L'incendie du Crédit Lyonnais

Le Crédit Lyonnais a été victime de l'incendie spectaculaire de son siège parisien, le 5 mai 1996. L'incendie, parti de la salle des marchés, ravage le bâtiment et la salle des coffres. Même si certains doutes subsistent sur son caractère accidentel, il occasionne de sérieux dégâts et le Crédit Lyonnais revend ensuite son siège central 195 millions d'euros à l'assureur AIG. Cette perte rentre donc dans la catégorie 'Dommages' (EL5) de la ligne métier 'Banque de détail' (BL3).

#### L'affaire Kerviel (EL1)

Plus récemment, la Société Générale a enregistré en 2008 des pertes énormes résultant de prises de positions frauduleuses d'un de ces traders (l'affaire Kerviel). Il aurait exposé la banque à un risque de marché en accumulant des positions acheteuses sur des contrats à terme sur indices. Il aurait surtout dissimulé ces opérations de marché en introduisant dans le système informatique de la Société Générale des opérations inverses fictives les compensant. Les causes de ces importantes pertes sont la fraude interne (EL1) et une défaillance dans les contrôles sur les activités de marché (BL2).

## 2.2 Les approches réglementaires

L'accord de Bâle II propose trois méthodes pour mesurer le risque opérationnel au titre du Pilier 1.

### 2.2.1 L'approche indicateur de base

C'est une approche forfaitaire. Le Comité de Bâle propose de retenir un indicateur d'exposition ( $IE$ ) basé sur le produit net bancaire ( $PNB$ ) moyen sur les trois dernières années. L'indicateur d'exposition de l'année  $n$  est donné par :

$$IE_n = \frac{1}{3} \sum_{i=1}^3 \max(PNB_{n-i}, 0)$$

L'exigence de fonds propres FP pour l'année  $n$  est alors égale à :

$$FP_n = 15\% * IE_n$$

Le calcul du capital alloué au risque opérationnel se fait à partir d'un indicateur uniquement relié au résultat de la banque. Ce qui fait de cette mesure une mesure très simpliste.

### 2.2.2 L'approche standardisée

C'est un prolongement plus fin de l'approche précédente en déclinant ce type de calcul par type d'activité. Dans cette approche, les fonds propres de la banque pour le risque opérationnel sont égaux à la somme des fonds propres de chaque catégorie d'activité. Nous avons pour l'année  $n$  :

$$FP_n = \sum_i FP_n(i) = \sum_i \beta(i) \times IE_n(i)$$

Le Comité de Bâle propose pour les coefficients  $\beta(i)$

Intitulé	Pondération
Financement d'entreprise	18%
Activités de marché	18%
Banque de détail	12%
Banque commerciale	15%
Paielements et règlements	18%
Fonctions d'agent	15%
Gestion d'actifs	15%
Courtage de détail	12%

### 2.2.3 L'approche AMA (Advanced Measurement Approach)

Le choix de la méthode utilisée par la banque pour calculer ses besoins en fonds propres est laissé libre par le Comité de Bâle. Cette méthode doit être fondée sur des critères qualitatifs et quantitatifs rigoureux, qui sont ensuite régulièrement contrôlés par l'ACP. Quelle que soit l'approche retenue, la mesure du risque opérationnel doit correspondre à un horizon d'un an et à un niveau de confiance de 99,9%. Cette approche tend en général à faire diminuer les fonds propres par rapport aux autres approches bien que cela ne soit pas toujours vérifié. Le Comité envisage trois méthodes différentes.

#### Internal Measurement Approach

On suppose qu'il existe une relation (linéaire ou non) entre la mesure de risque et la perte espérée. Pour une catégorie d'activité  $i$  et un type de risque  $j$ , la charge en capital est définie par exemple de la façon suivante :

$$FP(i, j) = EL(i, j) \times \gamma(i, j)$$

où  $EL$  représente la perte moyenne et  $\gamma$  est un facteur d'échelle. Notons que l'on peut calculer la perte moyenne comme le produit de trois termes :

$$EL(i, j) = IE(i, j) \times PE(i, j) \times LGE(i, j)$$

où  $IE$  est l'indicateur d'exposition,  $PE$  est la probabilité d'occurrence d'une perte unitaire et  $LGE$  est le montant de la perte unitaire. La grande difficulté est l'estimation du paramètre  $\gamma$ .

#### Scorecard Approach

Cette méthode consiste à se donner pour un type de risque ou pour une ligne de métier une valeur initiale de l'exigence en capital pour couvrir le risque opérationnel, puis de modifier cette valeur à l'aide de méthodes de notations (dites de scoring). Ces notations anticipent l'évolution de la gestion des risques au sein des lignes de métiers (par exemple), et donc de pouvoir donner une meilleure

exigence en capital pour les années futures. Le Comité de Bâle n'a aucune formulation mathématique pour cette approche. Néanmoins, des formules circulaient à l'époque à l'intérieur des banques. Elles étaient de la forme :

$$FP(i, j) = IE(i, j) \times w(i, j) \times RS(i, j)$$

avec  $IE$  l'indicateur d'exposition,  $RS$  le score de risque et  $w$  un facteur d'échelle.

### Loss Distribution Approach

La dernière méthode (Loss Distribution approach ou LDA) est la plus utilisée par la majorité des grands groupes bancaires ainsi que pour la tarification des contrats d'assurance non vie. Elle permet une meilleure appréciation du risque opérationnel et conduit généralement à réduire l'exigence en fonds propres de la banque.

## 3. La méthode LDA

La méthode LDA est dérivée d'une approche actuarielle ancienne et largement utilisée en assurance. On rappelle que le but de la méthode est de pouvoir calculer pour l'année à venir le quantile à 99,9% de la distribution de perte annuelle, soit l'exigence réglementaire en fonds propres appelée aussi charge en capital notée CaR pour Capital at Risk (ou VaR pour Value at Risk). La CaR représente en fait la couverture d'un risque millénaire. Le découpage se fait toujours selon les lignes métier et les catégories de risque et la méthode s'applique à chacune des mailles. L'idée générale est de modéliser la distribution de la perte globale liée au risque opérationnel pour une période donnée (1 an) en s'appuyant sur l'ensemble des pertes collectées dans les bases de données internes à l'établissement bancaire. La difficulté réside dans le fait que cette perte annuelle ne correspond pas à une occurrence mais qu'elle est le résultat de plusieurs pertes unitaires successives. Il convient alors de distinguer deux aspects du modèle :

- la distribution de la fréquence (nombre des pertes)
- la distribution de la sévérité (montant des pertes)

Mathématiquement, cela se traduit par l'introduction de deux variables aléatoires :

- $N$  la variable aléatoire discrète de comptage des pertes annuelles. Sa fonction de masse est  $p(k) = P(N = k)$ .
- $X$  la variable aléatoire continue représentant le montant d'une perte individuelle. Sa distribution est caractérisée par une fonction de répartition  $F$  et par une densité  $f$  de paramètre  $\theta$ .

L'hypothèse sur laquelle repose la construction du modèle suppose les pertes indépendantes entre elles, identiquement distribuées et indépendantes de  $N$ . La variable aléatoire représentant la perte annuelle peut alors s'écrire :

$$L = \begin{cases} 0 & \text{si } N = 0 \\ \sum_{n=1}^N X_n & \text{sinon} \end{cases}$$

Soit  $G$  la fonction de répartition de  $L$  alors :

$$G(x) = \begin{cases} p(0) & \text{pour } x = 0 \\ \sum_{n=1}^{\infty} p(n) F^{*n}(x) & \text{pour } x > 0 \end{cases}$$

où  $F_n$  est la fonction de répartition de la somme des  $n$  pertes  $(X_1, \dots, X_n)$ , qui s'obtient par auto-convolution d'ordre  $n$  de la distribution de sévérité. La charge en capital  $CaR$  de niveau  $\alpha$  s'obtient finalement en calculant le quantile d'ordre  $\alpha$  de la distribution de perte annuelle :

$$CaR_{\alpha} = \inf \{x \mid G(x) \geq \alpha\} = G^{-1}(\alpha)$$

Pour l'exigence en fonds propres du pilier 1, le quantile réglementaire est fixé à  $\alpha = 99.9\%$ .

La figure 2.2 illustre les différentes distributions qui composent le modèle ainsi que la charge en capital en fonction du niveau de quantile.

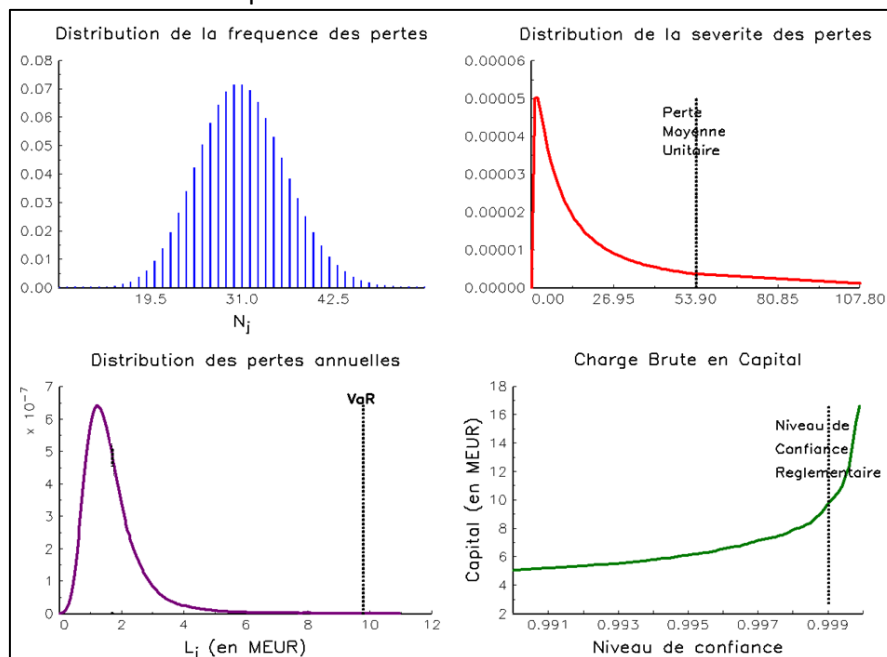


Figure 2.2 Les étapes du modèle LDA pour l'estimation de la charge en capital

Comme la fonction de répartition  $G$  n'admet pas d'écriture analytique simple (sauf pour la loi Gamma), la  $CaR$  est généralement calculée en utilisant des simulations de Monte-Carlo. Cela consiste à simuler un grand nombre ( $N_{MC}$ ) de pertes annuelles à partir des distributions de fréquence et de sévérité puis à prendre le quantile à 99.9% de l'échantillon simulé. Il peut aussi être envisagé de recourir à la F.F.T. (Fast Fourier Transform), ou la méthode de Panjer pour estimer la  $CaR$  et ce afin de diminuer le temps de calcul.

**Algorithm :** Calcul de la  $CaR$  par simulations de Monte-Carlo

- 1: Pour  $i = 1$  à  $N_{MC}$
- 2: Simuler un nombre  $n$  selon la distribution de fréquence

- 3: Simuler un  $n$ -échantillon  $x_1, \dots, x_n$  selon la distribution de sévérité
- 4: Poser  $L_i = \sum_{k=1}^N x_k$  la perte annuelle
- 5: Calculer la CaR comme le quantile à 99.9% de l'échantillon  $(L_1, \dots, L_{N_{MC}})$

## Récapitulatif des principales étapes de la méthode LDA

- Estimation de la distribution de sévérité C'est l'étape la plus importante de la méthode LDA et c'est aussi celle qui est la plus complexe. Comme nous le verrons dans le chapitre suivant, la loi de la sévérité actuellement utilisée en production est la loi log-Normale dont les paramètres  $\mu$  et  $\sigma$  sont estimés par la méthode du maximum de vraisemblance (ML) ou par la méthode des moments généralisée (GMM).
- Estimation de la distribution de fréquence La fréquence des pertes est supposée suivre une loi de Poisson de paramètre  $\lambda$ . Ce paramètre s'estime facilement par la méthode du Maximum de Vraisemblance qui n'est autre que la moyenne annuelle du nombre de pertes.
- Calcul de la charge en capital La CaR est calculée, comme expliqué précédemment, par simulations de Monte-Carlo. Il faut alors prendre en général un nombre de simulations de l'ordre de  $N_{MC} = 5.10^6$  pour avoir une précision de 1 %.

Enfin, deux étapes supplémentaires complètent le calcul global des fonds propres de la banque pour le risque opérationnel :

## Intégration de scénarios d'experts

Cette étape permet de compléter l'information des bases de données internes. En s'inspirant des pertes extrêmes subies par d'autres groupes bancaires ou des événements historiques ayant affecté le bilan de la banque (exemple : incendies, fraude d'un trader), les experts créent de façon qualitative des scénarios si l'historique des pertes semble insuffisant.

## Agrégation des charges en capital

Une fois, les CaRs estimées pour chaque couple ligne métier / catégorie de risque plus les scénarios, la CaR totale est agrégée en tenant compte des corrélations entre les différentes mailles de la matrice à l'aide de la théorie des copules par exemple.

L'objet de ce stage porte principalement sur la distribution de sévérité des pertes. La question est de déterminer comment des lois généralisées comme la loi  $\mathcal{G}\&\mathcal{H}$  peuvent être employées pour modéliser les montants de pertes et quel va en être l'impact sur la qualité de l'ajustement et le calcul de la CaR par la méthode LDA.

## 4. La distribution de sévérité des pertes

La grande difficulté de la méthode LDA est la spécification de la distribution de sévérité  $F$  ainsi que l'estimation de ses paramètres. Ce chapitre se focalise sur l'état des lieux concernant l'ajustement de cette distribution ainsi que sur les directives proposées par le Comité de Bâle sur ce sujet. Par la suite, on notera  $x_1, \dots, x_n$  l'échantillon de pertes à modéliser.

### 4.1 Propriétés souhaitées en risque opérationnel

Il est indispensable que la distribution choisie pour modéliser la sévérité soit continue et à support

positif. Ensuite la performance de notre modélisation sera mesurée par son respect des propriétés suivantes :

- Bon ajustement : le modèle doit vérifier un certain nombre de tests d'adéquation. En particulier il doit modéliser aussi bien la queue que le corps des distributions.
- Stabilité : La modélisation ne doit pas être sensible au nombre de données, en particulier celui de la queue. Par exemple lorsqu'on modélise séparément, corps et queue de distribution, le seuil de troncature dépend du nombre de données extrêmes.
- Flexibilité : Recouvrir plusieurs lois de probabilité connues et permettre ainsi d'utiliser le même modèle pour des lignes métiers différentes.
- Simplicité : Il faut que les méthodes d'estimation des paramètres et les méthodes de simulation soient implémentables.
- Réalisme : Permettre de calculer des valeurs de charges en capital ayant un sens du point de vue métier.

## 4.2 Distribution de probabilité à queue épaisse

Dans cette partie, on présente une caractérisation analytique des queues de distributions dites épaisses. Intuitivement, lorsqu'une variable aléatoire suit une distribution à queue épaisse, sa probabilité de prendre de grandes valeurs est loin d'être négligeable. Nous formalisons cela dans la définition suivante, qui se basent sur la théorie de Karamata.

Une variable aléatoire continue et réelle  $X$ , admettant une densité  $f$  et une fonction de répartition  $F$ , est à queue épaisse s'il existe  $p > 0$  tel que

$$\lim_{t \rightarrow \infty} \frac{t f(t)}{1 - F(t)} = p > 0$$

Nous allons nous baser sur ce résultat pour déterminer pour quelles valeurs de paramètres les lois étudiées sont à queue épaisse ou à queue fine.

## 4.3 Contraintes spécifiques

Les données de pertes opérationnelles possèdent des profils très atypiques. Malgré la classification des risques et l'amélioration de la collecte des pertes, il existe toujours une hétérogénéité non négligeable à l'intérieur des catégories mais aussi entre les catégories. Elles contiennent de nombreuses pertes de montants faibles ainsi que quelques événements extrêmes, ce qui les rend difficiles à ajuster. De plus, elles comportent des particularités engendrant une perte d'information importante dont il faut tenir compte dans l'inférence statistique.

### 4.3.1 Le seuil de collecte

Comme le permet le Comité de Bâle, et également à des fins de fiabilité, les pertes liées au risque opérationnel ne sont collectées que si leur montant excède un seuil de collecte noté  $H$ . Ce seuil de collecte est aussi le seuil à partir duquel on choisit de mettre en place l'inférence statistique. Il est donc nécessaire de le prendre en compte afin que le modèle soit statistiquement correct. C'est pourquoi au lieu de considérer la densité  $f$  de la distribution de sévérité lors de l'inférence, on considère la densité conditionnelle par rapport à  $H$  :

$$\tilde{f}_\theta(x|x \geq H) = \frac{f_\theta(x)}{1 - F_\theta(H)} \mathbb{I}_{x \geq H}$$

Cela permet d'intégrer la présence de données tronquées dans l'estimation de la distribution de la sévérité. Le paramètre  $\theta$  estimé est ainsi le paramètre de la distribution globale, c'est-à-dire des pertes à partir de 0. La distribution de fréquence doit aussi être modifiée puisqu'on collecte alors un nombre plus faible de pertes que le nombre réel. Le paramètre de fréquence est donc corrigé, après le calibrage de la distribution de sévérité :

$$\hat{\lambda}_{\text{corrigé}} = \frac{\hat{\lambda}_{\text{Expert}}}{1 - F_{\hat{\theta}}(H)}$$

Une fois les deux distributions estimées grâce à ces corrections, il est possible de simuler un échantillon de pertes représentatif de la réalité.

Remarque : le paramètre de fréquence  $\hat{\lambda}_{\text{Expert}}$  est calculé à partir du nombre de pertes moyen, dans le cas d'une distribution de Poisson. Ce paramètre est calculé par des experts métier qui le rendent le plus représentatif possible du nombre moyen des pertes en prenant en compte les évolutions attendues en termes de profil de risque (aspect forward-looking du modèle)

#### 4.3.2 Les agrégats

Les bases de données de pertes opérationnelles présentent également la particularité de contenir des agrégats. En effet, certains montants de pertes sont reportés comme étant la somme de plusieurs événements dont les montants ne sont pas connus. Les bases de données sont constituées de cette façon pour des raisons de confidentialité (par exemple dans le cas de litiges juridiques) ou pour des raisons métiers (par exemple dans le cas de la fraude à la carte bancaire). Il en résulte une perte d'information importante puisqu'une perte de 10000€ composée de deux agrégats peut aussi bien être la somme de 9999 € + 1 € que de 5000 € + 5000 €.

Le vecteur du nombre d'événements associés à l'échantillon est noté  $(\kappa_1, \dots, \kappa_n)$ . On choisit donc de considérer pour l'ensemble de l'étude l'échantillon "réduit"  $X^\kappa = \left(\frac{x_1}{\kappa_1}, \dots, \frac{x_n}{\kappa_n}\right)$ .

### 4.4 Tests d'adéquation

Avant de présenter les lois de sévérité les plus utilisées nous introduisons les tests d'ajustement qui nous permettront de juger de la qualité de l'ajustement entre données et modèle. Les tests utilisés dans cette étude sont des tests d'adéquation de loi très couramment employés en statistique. Ils tiennent naturellement compte des caractéristiques des distributions, en particulier de la troncature et de leur caractère leptokurtique. Le dernier test présenté ADup est, à cet égard, une correction du test d'Anderson-Darling standard qui se focalise essentiellement sur l'ajustement en queues de distribution. Notations pour ce paragraphe :

$$z_H = F_\theta(H), z_j = F_\theta(\xi_j)$$

#### Test de Kolmogorov-Smirnov (KS)



Le test de Kolmogorov mesure l'écart maximal (en valeur absolue) entre la distribution empirique et la distribution estimée sur le jeu de données. Il s'agit donc d'un test local puisque son résultat peut être grandement conditionné par l'ajustement d'une minorité d'observations du jeu de données. En effet, si l'ajustement peut être parfait sur presque l'ensemble du support, une donnée aberrante ou même extrême peut conduire au rejet de l'hypothèse nulle. En présence de données censurées à gauche, on démontre aisément que KS a pour expression :

$$KS = \frac{\sqrt{n}}{1 - z_H} \max \left\{ \sup_j \left( z_H - z_j + \frac{j}{n} (1 - z_H) \right), \sup_j \left( z_j - z_H - \frac{j-1}{n} (1 - z_H) \right) \right\}$$

### Test de Cramer-Von-Mises (CVM)

Test appartenant, comme son nom le suggère, à la famille de Cramer-von-Mises, il s'agit d'un test quadratique mesurant la moyenne des écarts résidant entre la distribution empirique et la distribution estimée au carré, dans le cas standard où la pondération vaut 1 quelle que soit l'observation. En présence de données censurées à gauche, on démontre que CVM s'exprime par :

$$CVM = \frac{n}{3} + \frac{nz_H}{1 - z_H} + \frac{1}{n(1 - z_H)} \sum_{j=1}^n (1 - 2j)z_j + \frac{1}{(1 - z_H)^2} \sum_{j=1}^n (z_j - z_H)^2$$

### Test d'Anderson-Darling (AD)

Test appartenant également à la famille de Cramer, il s'agit donc d'un test quadratique. Il correspond au cas où la pondération vaut  $\psi(\hat{F}(x)) = [\hat{F}(x)(1 - \hat{F}(x))]^{-1}$ . Par cette pondération, la statistique d'Anderson-Darling accorde un poids plus important aux queues gauches et droites des distributions. En présence de données censurées à gauche, on démontre qu'AD a pour expression :

$$AD = -n + 2n \ln(1 - z_H) - \frac{1}{n} \sum_{j=1}^n (1 + 2(n - j)) \ln(1 - z_j) + \frac{1}{n} \sum_{j=1}^n (1 - 2j) \ln(z_j - z_H)$$

### Test d'Anderson-Darling modifié (AD<sub>up</sub>)

A l'instar des deux tests précédents, ce test appartient une nouvelle fois à la famille de Cramer. Test quadratique, donc, il correspond au cas où la pondération vaut  $\psi(\hat{F}(x)) = [1 - \hat{F}(x)]^{-2}$ . Par l'intermédiaire de cette pondération, le test d'Anderson-Darling modifié attribut un surcroît d'importance très significatif aux queues droites des distributions correspondant à la région des pertes très sévères (nettement plus que dans le test d'Anderson-Darling non modifié). En présence de données censurées à gauche, on démontre qu'AD<sub>up</sub> a pour expression :

$$\mathbb{D}AD_{up} = -2n \ln(1 - z_H) + 2 \sum_{j=1}^n \ln(1 - z_j) + \frac{1 - z_H}{n} \sum_{j=1}^n \frac{(1 - 2j)(1 + 2(n - j))}{1 - z_j}$$

### Implémentation

Il n'existe pas de forme analytique pour les p-values de ces tests, nous suivront alors la méthodologie de l'article d'Anna Chernobai [14]. Cette méthode consiste à calculer les p-values et les valeurs critiques par simulation Monte-Carlo. La valeur de la p-value nous permet de décider si nous rejetons ou non l'hypothèse nulle stipulant le bon ajustement de la distribution aux données. Ainsi, les valeurs critiques vont être calculées à partir d'échantillons générés. Il s'agit d'accomplir donc les étapes de l'algorithme suivant.

Pour un test K :

#### Algorithme

1. calculer la statistique  $K_0$ , telle que définie en annexe, avec la fonction de répartition dont les paramètres déjà estimés sont  $\widehat{\theta}_0$ .
2. Générer, à partir des paramètres estimés de la distribution à tester, un échantillon de montants de pertes de même taille que l'échantillon de départ.
3. A partir de l'échantillon généré, estimer les paramètres  $\widehat{\theta}$  de la même distribution
4. Calculer la statistique  $K_i$  avec la fonction de répartition dont les paramètres nouvellement estimés sont  $\widehat{\theta}$ .
5. Comparer les statistiques  $K_0$  et  $K_i$ . Si  $K_0 < K_i$  alors un compteur  $j$  s'incrmente de 1.
6. Répéter les étapes 2, 3, 4 et 5 un grand nombre de fois  $N$  (ex.  $N = 1000$ ).
7. Calculer la p-value ( $pv$ ) comme étant  $pv = \frac{j}{N}$ .
8. Rejeter l'hypothèse nulle (l'ajustement), si la p-value est plus petite que le niveau de confiance (5%)

## Lois de probabilité

De nombreuses lois de probabilités sont utilisées pour modéliser la sévérité des pertes opérationnelles. Ces lois doivent être à support positif et asymétriques avec une queue à droite plus épaisse. Ci-dessous, une liste des lois les plus couramment utilisées en risque opérationnel.

Loi de probabilité	Densité $f(x)$
Exponentielle	$\lambda \exp - \lambda x$
Weibull	$\alpha \theta x^{\theta-1} e^{-\alpha x \theta}$
Gamma	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\lambda x)$
Log-normale	$\frac{1}{\sqrt{2\pi}x\sigma} \exp \left[ -\frac{1}{2} \left( \frac{\log x - \mu}{\sigma} \right)^2 \right]$
Log-logistique	$\frac{\eta(x - \alpha)^{\eta-1}}{[1 + (x - \alpha)^\eta]^2}$
Pareto (translatée)	$\frac{\alpha}{\theta} \left( 1 + \frac{x}{\theta} \right)^{-(\alpha+1)}$
Loi de Burr	$ck \frac{x^{c-1}}{(1 + x^c)^{k+1}}$

Les lois paramétriques simples (deux paramètres au plus) utilisées pour modéliser les sévérités des risques en assurance et en finance, offrent aux analystes de risques, des outils efficaces afin de mieux appréhender les pertes futures. Il existe néanmoins de nombreux phénomènes qui n'obéissent pas à ce type de lois, et qui nécessitent l'introduction de distributions qui s'adaptent à plus de complexité. On s'intéresse alors à des familles de distributions, dites généralisées, afin de pouvoir apporter plus d'ajustement entre données et modèles. Une autre motivation à vouloir généraliser des lois de probabilité est que souvent, les lois paramétriques simples adéquates aux données varient d'une catégorie de risque à une autre et d'une année à l'autre, ceci pose un réel problème de gestion aux établissements financiers et rend ainsi la tâche d'automatisation du processus de gestion des risques difficile.

La famille de distributions généralisées contient des lois ayant au moins quatre paramètres, et de laquelle plusieurs autres distributions paramétriques simples peuvent être dérivées ou approchées, comme cas particulier.

Les quatre premiers paramètres représentent, typiquement :

- La localisation : (médiane ou moyenne)

- L'échelle (volatilité ou écart type)

- L'asymétrie (Skewness)

- L'aplatissement (Kurtosis)

## II- Modélisation de la sévérité par des distributions $\mathcal{G}\&\mathcal{H}$

### 1. Définition des lois $\mathcal{G}\&\mathcal{H}$

#### Introduction

Proposées par J.Tukey, Les lois  $\mathcal{G}\&\mathcal{H}$  ont été ensuite étudiées par Hoaglin et Peters (1979), Martinez et Iglewicz [7]. Hoaglin [5], a comparé les techniques d'estimation de ces paramètres, Dutta et Perry [6] étudient quant à eux l'application du modèle sur les risques opérationnels. Le fait que ces lois soient une simple transformation de la loi normale centrée réduite, facilite de nombreux calculs tels que l'évaluation des quantiles et la simulation Monte-Carlo.

Dans ce chapitre, on commencera par introduire deux sous-familles de ces lois qui sont : les lois  $\mathcal{G}$ , qui permettent de décrire des variables continues asymétriques et les lois  $\mathcal{H}$ , qui peuvent modéliser les queues épaisses. Les lois  $\mathcal{G}$  et  $\mathcal{H}$  sont ensuite réunies pour former les lois  $\mathcal{G}\&\mathcal{H}$ , des lois permettant de décrire des structures d'asymétrie et d'aplatissement plus complexes.

#### 1.1 Asymétrie et lois $\mathcal{G}$

Nous modélisons une variable aléatoire asymétrique  $X$  de loi  $\mathcal{G}$  et de paramètres  $A, B$  et  $g$ , à l'aide d'une fonction monotone d'une variable normale centrée réduite  $Z$ . Les paramètres  $A$  et  $B$  tiennent compte, respectivement, de la localisation et de l'échelle de  $X$ . Nous notons  $X \sim \mathcal{G}(A, B, g)$ .

Nous écrivons :

$$X = A + B \cdot Y$$

Où  $A$  et  $B$  sont des scalaires et  $Y$  est une variable aléatoire de loi  $\mathcal{G}(0, 1, g)$ . La médiane de  $Y$  est choisie égale à zéro de sorte que  $A$  soit la médiane de  $X$ .

Nous écrivons maintenant  $Y$  comme fonction de  $Z$ ,  $Y = Y(Z)$ . Une façon de procéder, d'après J.Tukey [11] est d'introduire une fonction de modulation  $G$  qui affecte les valeurs positives de  $Z$  différemment des valeurs négatives.

$$Y = G(Z) \cdot Z$$

Le cas  $G(z) = 1$  correspond à la loi normale centrée réduite. Nous prendrons, afin de représenter l'asymétrie,  $G$  tel que  $G(-z) \neq G(z)$ , pour tout  $z \neq 0$ . Ensuite pour prendre en compte le fait que l'effet d'asymétrie est léger près de la médiane, il est nécessaire d'exiger  $G(z) \approx 1$  au voisinage de 0.

Une famille pratique de fonctions, proposée par Tukey et vérifiant ces propriétés est définie par :

$$Y_g(z) = \frac{e^{gz} - 1}{gz} \cdot z = \frac{e^{gz} - 1}{g}$$

Remarque : MacGillivray et Cannon's [12] et Rayner et MacGillivray [9] ont proposé d'autres familles de fonctions vérifiant les propriétés énoncées précédemment. L'étude se fera sur les lois  $\mathcal{G}$  de Tukey.

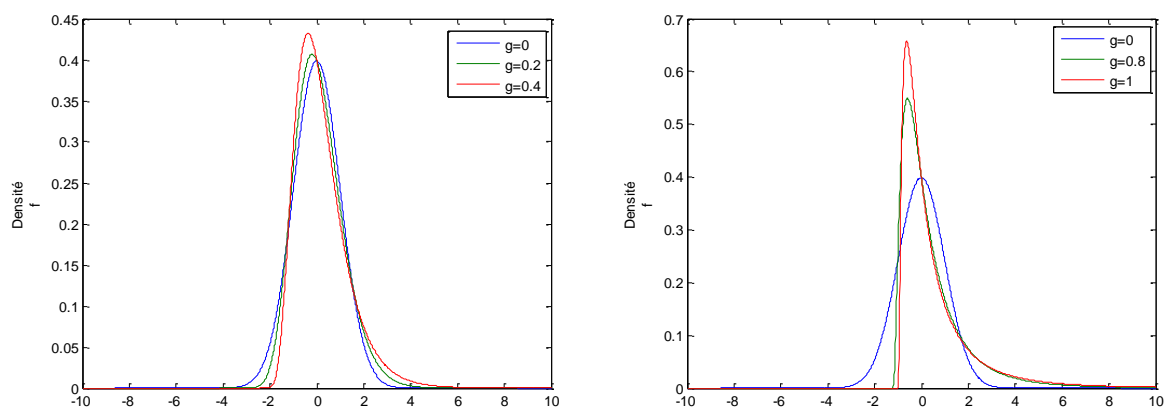
**Définition :** Une variable aléatoire continue  $X$  de loi  $\mathcal{G}$  (de Tukey), et de paramètres  $A, B$  et  $g$  notée :  $\mathcal{G}(A, B, g)$  s'écrit :

$$X = A + B \cdot \frac{e^{gZ} - 1}{g}$$

Où  $Z$  est une variable aléatoire normale centrée réduite,  $A$  correspond à la médiane,  $g$  au paramètre d'asymétrie, et  $B$  est un paramètre d'échelle. Le cas  $g \rightarrow 0$  correspond à loi normale.

### Asymétrie pour diverses valeurs de $g$

Afin de visualiser l'effet du paramètre  $g$  sur l'asymétrie nous traçons les densités de loi  $\mathcal{G}(0,1,g)$  pour  $g = 0, 0.2, 0.4, 0.8$  et  $1$



Notons que des valeurs négatives de  $g$  produisent une asymétrie (négative) dans le sens contraire à celui des schémas ci-dessus.

## 1.2 Aplatissement et lois $\mathcal{H}$

Afin de modéliser des lois ayant des queues plus épaisses que celles de la loi normale, nous introduisons une transformation de cette dernière qui octroie plus de poids aux valeurs extrêmes. Nous modélisons une variable aléatoire asymétrique  $X$  de loi  $\mathcal{H}$  et de paramètres  $A, B$  et  $h$ , à l'aide d'une fonction d'une variable normale centrée réduite  $Z$ . Les paramètres  $A$  et  $B$  tiennent compte, respectivement, de la localisation et de l'échelle de  $X$ . On note  $X \sim \mathcal{H}(A, B, h)$ .

Nous écrivons :

$$X = A + B \cdot Y$$

De la même façon que pour les lois  $\mathcal{G}$  on introduit une fonction  $H$ , qui modélise l'aplatissement. Le choix de  $H$  se fait de manière à permettre d'étirer les queues, tout en préservant la symétrie. Cela exige que  $H$  soit une fonction paire et strictement positive.

$$H(z) = e^{hz^2/2}$$

De plus pour que l'aplatissement opère, il faut que  $H$  soit croissante pour  $z^2 \geq -\frac{1}{h}$ .

Une famille simple de fonctions, proposée par Tukey, ayant le comportement souhaité est définie par :  
D'où :

$$Y_h(z) = z \cdot H(z) = z \cdot e^{hz^2/2}$$

**Définition :** Une variable aléatoire continue  $X$  de loi (de Tukey), et de paramètres  $A, B$  et  $h$ , notée :  $\mathcal{H}(A, B, h)$  s'écrit :

$$X = A + B \cdot Z \cdot e^{hZ^2/2}$$

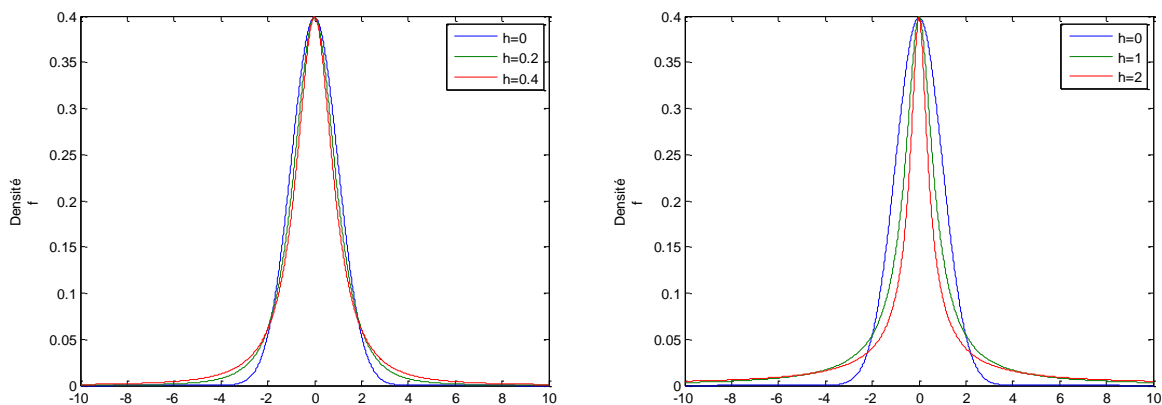
Où  $Z$  une variable aléatoire normale centrée réduite,  $A$  correspond à la médiane et  $B$  à un paramètre d'échelle et  $h$  contrôle l'importance et la direction de l'aplatissement.

Remarques :

- Le cas  $h = 0$  correspond à la loi normale, et donc à une absence d'aplatissement.
- Une valeur négative de  $h$  constitue une difficulté numérique car  $Y_h(z)$  n'est plus monotone pour  $z^2 > -\frac{1}{h}$ . On l'observe sur les résultats obtenus lors de la simulation d'échantillons de loi  $\mathcal{H}$  et lors du calcul de sa fonction de densité. Toutefois, les données en risque opérationnel se caractérisent « souvent » par des queues épaisses, ce qui correspond à un  $h$  positif. Nous décrivons en annexe, de façon détaillée, le traitement que l'on utilise pour remédier à ce problème.

### Aplatissement pour diverses valeurs du paramètre $h$

Afin de visualiser l'effet du paramètre  $h$  sur l'asymétrie nous allons tracer les densités de loi  $\mathcal{H}(0,1,h)$  pour  $h = 0, 0.2, 0.4, 1$  et  $2$ .



### 1.3 Lois $\mathcal{G}$ & $\mathcal{H}$

Pour pouvoir rassembler les lois  $\mathcal{G}$  et les lois  $\mathcal{H}$ , il faut permettre à la distribution résultante d'être à la fois asymétrique et étalée. En effet, traiter simultanément ces deux aspects en plus de la localisation et de l'échelle permet plus de flexibilité.

Pour combiner ces deux aspects, nous utilisons de nouveau la multiplication, en écrivant comme précédemment :

$$X = A + B \cdot Y$$

$$Y(z) = z \cdot G(z) H(z)$$

Nous prenons les mêmes choix de fonctions particulières, introduits par Tukey, que les lois  $\mathcal{G}$  et les lois  $\mathcal{H}$ , définis précédemment.

Nous écrivons alors :

$$Y(z) = \left( \frac{e^{gz} - 1}{g} \right) e^{hz^2/2}$$

Pour les raisons expliquées dans la partie précédente, nous choisissons de ne traiter que les  $h$  positifs ou nuls.

**Définition :** Une variable aléatoire continue  $X$  de loi  $\mathcal{G}\&\mathcal{H}$  et de paramètres  $A, B, g$  et  $h$  notée :  $\mathcal{G}\&\mathcal{H}(A, B, g, h)$  s'écrit :

$$X = A + B \cdot \left( \frac{e^{gZ} - 1}{g} \right) e^{hZ^2/2}$$

Où  $Z$  une variable aléatoire normale centrée réduite,  $A$  correspond à la médiane,  $g$  au paramètre d'asymétrie,  $h$  au paramètre d'aplatissement et  $B$  est un paramètre d'échelle.

### 1.3.1 Quelques lois obtenues par différentes combinaisons de A, B, g et h

Les lois  $\mathcal{G}\&\mathcal{H}$  offrent par leur flexibilité un grand potentiel de modélisation. D'après la partie précédente, ces lois permettent d'approcher de nombreuses distributions théoriques. En effet, Martinez et Iglewicz [8] montrent comment plus de douze lois univariées, parmi lesquelles, Log-Student, Weibull, Log-Normale peuvent être approchées en choisissant de façon appropriée les paramètres  $g$  et  $h$ .

#### Loi Log-Normale :

La loi log-normale est un cas particulier d'une distribution  $\mathcal{G}\&\mathcal{H}$ , lorsque  $g$  est constant et positif. Ainsi, il est possible de trouver des relations entre les paramètres des deux distributions.

Si  $X$  est une variable aléatoire de distribution  $\mathcal{LN}(\mu, \sigma)$  et  $Z \sim \mathcal{N}(0,1)$  alors on a :

$$X = e^{\mu + \sigma Z}$$

Donc :

$$X = e^{\mu} + \sigma e^{\mu} \left( \frac{e^{\sigma Z} - 1}{\sigma} \right)$$

Par identification par rapport à une variable aléatoire  $T$  de loi  $\mathcal{G}\&\mathcal{H}(A, B, g, 0)$

$$T = A + B \left( \frac{e^{gZ} - 1}{g} \right)$$

On en déduit que  $X$  est de loi  $\mathcal{G}\&\mathcal{H}(e^{\mu}, \sigma e^{\mu}, \sigma, 0)$ .

De la même façon, on obtient pour quelques distributions de probabilité les valeurs des paramètres  $A, B, g$  et  $h$  des lois  $\mathcal{G}\&\mathcal{H}$  qui leurs correspondent.

Distribution	Paramètres	$\mathcal{G}\&\mathcal{H}$			
		$A$	$B$	$g$	$h$
Cauchy	$\mu, \sigma > 0$	$\mu$	$\sigma$	0	1
Normale	$\mu, \sigma$	$\mu$	$\sigma$	0	0
Log-normale	$\mu, \sigma$	$e^\mu$	$\sigma e^\mu$	$\sigma$	1

### 1.3.2 Analyse de flexibilité : Skewness-Kurtosis (asymétrie-aplatissement)

Outre la caractérisation par leur nombre de paramètres, les lois de probabilité peuvent aussi être caractérisées par l'évaluation de l'épaisseur de leurs queues et leur asymétrie. Dans le contexte des distributions de sévérité, la queue correspond à la partie qui se trouve au-dessus d'un certain seuil « élevé ». Une distribution est dite à queue épaisse si la vraisemblance de tomber sur une grosse perte est « levée ». Ceci étant dit, il existe plusieurs manières formelles de définir une loi à queue épaisse, l'une des plus connues est celle que nous avons décrite dans le 1<sup>er</sup> chapitre.

Une façon pratique de caractériser une distribution, utilisée par Dutta et Perry [6] est basée sur une étude de ses moments d'ordre  $k$ . Les deux premiers représentent la localisation et l'échelle, le 3<sup>ème</sup> mesure l'asymétrie (Skewness) de la loi, et enfin le 4<sup>ème</sup> caractérise l'épaisseur de la queue ou l'aplatissement (Kurtosis) de la distribution.

Les deux derniers sont définis respectivement par :

$$S = \frac{\mu_3}{\mu_2^{3/2}}$$

Et :

$$K = \frac{\mu_4}{\mu_2^2}$$

Où :

$$\mu = E(X) \quad \text{et} \quad \mu_k = E\{(X - \mu)^k\}$$

Si :

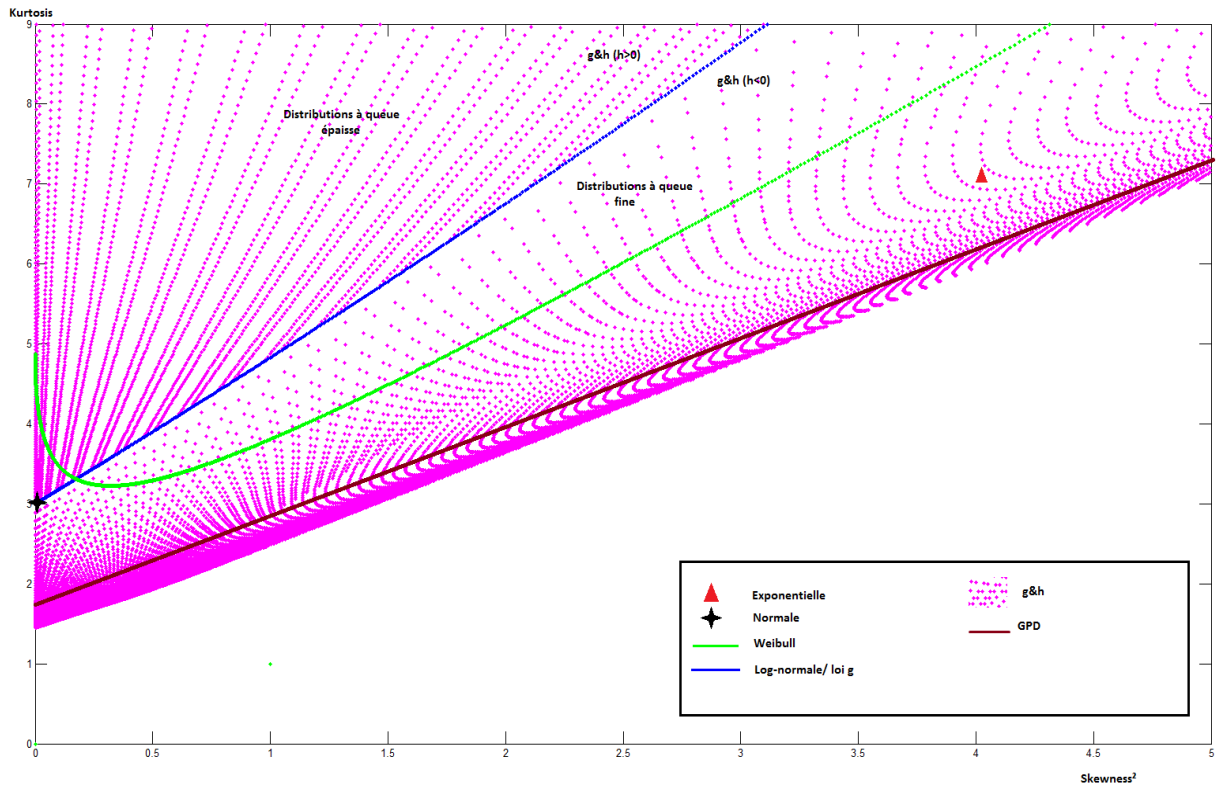
$K < 3$  : distribution platikurtique.

$K = 3$  : distribution mésokurtique.

$K > 3$  : distribution leptokurtique.

Une façon intéressante de visualiser la flexibilité d'une loi de probabilité est de tracer un nuage de points sur un axe  $Skewness^2 - Kurtosis$ . Ce dernier montre l'emplacement des couples de ces mesures que la distribution peut représenter pour différentes valeurs de ces paramètres. Plus la surface balayée par la loi est grande plus la loi est flexible. On obtient alors un point, une courbe ou une surface selon si le Skewness ou le Kurtosis sont des fonctions de zéro, un ou plusieurs paramètres.





Sur la figure ci-dessus, on voit clairement que les lois  $\mathcal{G}\&\mathcal{H}$  peuvent balayer une surface importante de couple Skewness-Kurtosis. Elles permettent aussi par leur paramètre d'aplatissement  $h$  de séparer distributions à queue épaisse et celles à queue fine (i.e. :  $h = 0$  constitue une frontière entre distributions à queue fine et celles à queue épaisse).

### 1.3.3 Fonction de répartition

Soit  $X$  une variable aléatoire de distribution  $\mathcal{G}\&\mathcal{H}(A, B, g, h)$ . La fonction de répartition de  $X$  s'écrit :

$$F(t) = \Phi \left[ Y^{-1} \left( \frac{t - A}{B} \right) \right], \quad \forall t \in \mathbb{R}$$

Où  $\Phi$  est la fonction de répartition d'une  $N(0,1)$  et  $Y^{-1}$  la réciproque de la fonction de Tukey :

$$Y(z) = \left( \frac{e^{gz} - 1}{g} \right) e^{hz^2/2}$$

En effet :

$$\begin{aligned} F(t) &= P(X \leq t) \\ &= P(A + B.Y \leq t) \\ &= P\left(Y \leq \frac{t - A}{B}\right) \\ &= P\left(Z \leq Y^{-1}\left(\frac{t - A}{B}\right)\right) \\ &= \Phi\left[Y^{-1}\left(\frac{t - A}{B}\right)\right] \end{aligned}$$

### 1.3.4 Densité

Par dérivation de ce que nous avons obtenu précédemment, nous obtenons la densité de  $X$  :

$$f(t) = \frac{\varphi \left[ Y^{-1} \left( \frac{t-A}{B} \right) \right]}{B \cdot Y' \left[ Y^{-1} \left( \frac{t-A}{B} \right) \right]}, \quad \forall t \in \mathbb{R}$$

Où  $\varphi$  est la fonction densité d'une  $\mathcal{N}(0,1)$  et  $Y'(z) = e^{gz+hz^2/2} + hz \left( \frac{e^{gz}-1}{g} \right) e^{hz^2/2}$ .

### 1.3.5 Fonction de répartition inverse (fonction quantile)

Nous démontrons que, pour un niveau de quantile  $\alpha$ , que la fonction quantile s'obtient en appliquant la même transformation qui définit une variable aléatoire de distribution  $\mathcal{G}\&\mathcal{H}$ , c'est-à-dire :

Pour  $0 < \alpha < 1$ , on a

$$F^{-1}(\alpha) = A + B \cdot [Y(\Phi^{-1}(\alpha))]$$

### 1.3.6 Epaisseur de la queue

Afin de déterminer pour quelles valeurs des paramètres une loi  $\mathcal{G}\&\mathcal{H}$  est à queue épaisse ou fine, on va utiliser la caractérisation énoncée dans la première section.

**Propriété :** Si  $X$  une variable aléatoire de distribution  $\mathcal{G}\&\mathcal{H}(A, B, g, h)$  alors la loi de  $X$  est à variation régulière d'indice  $\frac{1}{h}$ .

Pour simplifier, nous développons la démonstration pour  $A = 0$  et  $B = 1$ , i.e.  $X \sim \mathcal{G}\&\mathcal{H}(0, 1, g, h)$ .

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{xf(x)}{1-F(x)} &= \lim_{x \rightarrow \infty} \frac{x \varphi[Y^{-1}(x)]}{[1 - \Phi[Y^{-1}(x)]] \cdot Y'[Y^{-1}(x)]} \\ &= \lim_{u \rightarrow \infty} \frac{Y(u) \cdot \varphi(u)}{[1 - \Phi(u)] \cdot Y'(u)}, \quad u := Y^{-1}(x) \\ &= \lim_{u \rightarrow \infty} \frac{(e^{gu} - 1)\varphi(u)}{[1 - \Phi(u)] \cdot [ge^g + hu(e^{gu} - 1)]} \\ &= \frac{1}{h} \end{aligned}$$

La loi  $\mathcal{G}\&\mathcal{H}$  est donc à queue épaisse pour  $h$  strictement positif.

Remarque : Les lois  $\mathcal{G}$  (dont la log-normale) ne sont quant à elles pas à variation régulière et donc ne sont pas à queue épaisse selon la caractérisation de Karamata. Elles sont donc limitées pour ajuster des distributions de risques opérationnels. L'introduction d'un paramètre supplémentaire  $h$  permet de modéliser cette particularité.

### 1.3.7 Moments d'ordre $n$

Le moment d'ordre  $n$  d'une distribution  $\mathcal{GH}(A, B, g, h)$  pour  $g \neq 0$  et  $0 \leq h < \frac{1}{n}$  est donné, d'après Dutta et Perry (2007), par :

$$E(X^n) = \sum_{i=0}^n \binom{n}{i} A^{n-i} B^i \frac{\sum_{r=0}^i (-1)^{-1} \binom{i}{r} e^{\frac{\{(i-r)g\}^2}{2(1-ih)}}}{g^i \sqrt{1-ih}}$$

On en déduit l'espérance et la variance :

$$E(X) = A + B \cdot \left[ \frac{1}{g \sqrt{1-h}} (e^{g^2/\{2(1-h)\}} - 1) \right], \text{ pour } h < 1$$

$$var(X) = B^2 \cdot \left[ \frac{1}{g^2 \sqrt{1-2h}} (e^{2g^2/\{1-2h\}} - 2e^{g^2/\{2(1-2h)\}} + 1) - \frac{1}{g^2 (1-h)} (e^{g^2/\{2(1-h)\}} - 1)^2 \right], \text{ pour } h < \frac{1}{2}$$

Compte tenu de la complexité des moments, les skewness et kurtosis ne peuvent avoir de formes intéressantes. Il est cependant possible de les obtenir numériquement grâce aux précédents résultats.

## 2. Estimation des paramètres

### 2.1 Prise en compte du seuil de collecte

Les pertes ne sont collectées qu'à partir d'un certain seuil  $H$ , ce qui affecte l'estimation des paramètres puisque la distribution empirique est différente de la vraie distribution. Il faut donc relier la vraie distribution à sa version empirique, en considérant la densité conditionnelle :

$$\tilde{f}_{\theta|H}(t) = \frac{f_{\theta}(t)}{\int_H^{+\infty} f_{\theta}(u) du} \mathbb{I}_{\{t \geq H\}} = \frac{f_{\theta}(t)}{1 - F_{\theta}(H)} \mathbb{I}_{\{t \geq H\}}$$

En remplaçant la densité et la fonction de répartition par leurs valeurs, on obtient :

$$\tilde{f}_{\theta|H}(t) = \frac{\varphi \left[ Y^{-1} \left( \frac{t-A}{B} \right) \right]}{B \cdot Y' \left[ Y^{-1} \left( \frac{t-A}{B} \right) \right] * \left[ 1 - \Phi \left[ Y^{-1} \left( \frac{H-A}{B} \right) \right] \right]} \mathbb{I}_{\{t \geq H\}}$$

$$\text{Avec : } Y'(\mathbb{R}) = B \left\{ e^{gz+hz^2/2} + hz \left( \frac{e^{gz}-1}{g} \right) e^{hz^2/2} \right\}$$

On en déduit, la fonction de répartition et la fonction quantile :

$$\tilde{F}_{\theta|H}(t) = \frac{F_{\theta}(t) - F_{\theta}(H)}{1 - F_{\theta}(H)} \mathbb{I}_{\{t \geq H\}}$$

Donc :

$$\tilde{F}_{\theta|H}(t) = \frac{\Phi\left[Y^{-1}\left(\frac{t-A}{B}\right)\right] - \Phi\left[Y^{-1}\left(\frac{H-A}{B}\right)\right]}{1 - \Phi\left[Y^{-1}\left(\frac{H-A}{B}\right)\right]} \mathbb{1}_{\{t \geq H\}}$$

Et :

$$\tilde{F}_{\theta|H}^{-1}(\alpha) = F^{-1}\left[\left[1 - \Phi\left[Y^{-1}\left(\frac{H-A}{B}\right)\right]\right] * \alpha + \Phi\left[Y^{-1}\left(\frac{H-A}{B}\right)\right]\right]$$

Donc

$$\tilde{F}_{\theta|H}^{-1}(\alpha) = F^{-1}\left[\alpha + (1 - \alpha) \cdot \Phi\left[Y^{-1}\left(\frac{H-A}{B}\right)\right]\right]$$

Avec :

$$F^{-1}(\alpha) = A + B \cdot [Y(\Phi^{-1}(\alpha))]$$

## 2.2 Méthode Inter-quantile (IQ)

Nous commençons par décrire une approche simple et pratique, présentée par J.Drouin dans [3], permettant d'estimer les paramètres  $A, B, g, h$ . On considère  $X_{g,h}$  et  $Z$  des variables aléatoires de distributions  $\mathcal{G}\&\mathcal{H}(A, B, g, h)$  et  $\mathcal{N}(0,1)$ . On note  $x_p$  et  $z_p$  leurs quantiles de niveau  $p$ .

Nous avons :

$$x_p = A + B \left( \frac{e^{gz_p} - 1}{g} \right) e^{hz_p^2/2} \quad (1)$$

On en déduit par symétrie de la loi normale :

$$x_{1-p} = A + B \left( \frac{e^{-gz_p} - 1}{g} \right) e^{hz_p^2/2} \quad (2)$$

(2)/(1) nous donne, avec  $x_{0.5} = A$ , la médiane de  $X_{g,h}$  :

$$\frac{x_{1-p} - x_{0.5}}{x_p - x_{0.5}} = \frac{x_{1-p} - 1}{e^{gz_p} - 1} = -e^{-gz_p}$$

Donc, pour  $0 < p < 0.5$  on a :

$$g_p = -\frac{1}{z_p} \ln \left( \frac{x_{1-p} - x_{0.5}}{x_{0.5} - x_p} \right)$$

Dans ce résultat, on voit bien que  $g$  dépend de  $p$ . Hoaglin [5] a suggéré de choisir le paramètre  $g$  égale à la médiane des  $g_p$ .

Maintenant que nous avons défini  $A$  et  $g$ , occupons-nous de  $B$  et  $h$ .

(1) – (2) nous donne :

$$\begin{aligned} x_p - x_{1-p} &= B \left( \frac{e^{gz_p} - e^{-gz_p}}{g} \right) e^{hz_p^2/2} \\ \ln \left( \frac{g(x_p - x_{1-p})}{e^{gz_p} - e^{-gz_p}} \right) &= \ln(B) + \frac{hz_p^2}{2} \end{aligned}$$

Etant donné que les distributions pertes en risque opérationnel sont positivement asymétriques ( $g > 0$ ) et leptokurtiques, il est plus approprié d'exprimer le terme à gauche à l'aide du demi-spread

supérieur ( $UHS$ ), obtenu à partir de (2) seulement (ceci est plus intéressant dans le sens où on se focalise sur la queue droite de la distribution) :

$$UHS = \frac{g(x_{1-p} - x_{0.5})}{e^{gz_p} - 1}$$

Les valeurs de  $B$  et  $h$  sont alors estimées en effectuant une régression linéaire de  $\ln(UHS)$  en fonction de  $\frac{z_p^2}{2}$ . L'estimateur de  $B$  est égal à l'exponentielle de l'ordonnée à l'origine et celui de  $h$  est égal à la pente de la régression.

Il ne s'agit toutefois pas d'une méthode d'estimation de paramètres utilisable en pratique car elle se base sur un résultat graphique et dépend de plusieurs hypothèses difficiles à vérifier. En effet cette procédure donne des résultats aberrants dès que les échantillons traités ne proviennent pas d'une simulation d'une distribution  $\mathcal{G}\&\mathcal{H}$ , ceci la rend encore moins efficace vis-à-vis de données réelles. Néanmoins, nous faisons le choix d'utiliser le résultat de cette procédure comme point de départ des algorithmes d'optimisation des méthodes d'estimations étudiées, son temps de calcul immédiat (formules fermées) ne pouvant avoir qu'une influence positive sur les programmes d'optimisation.

## 2.3 Méthode du maximum de vraisemblance (ML)

La méthode du maximum de vraisemblance est la plus connue en estimation paramétrique, elle est en effet celle qui possède les meilleures propriétés théoriques.

Rappelons que si la fonction de vraisemblance  $L(\theta; x)$  admet un maximum unique au point  $\hat{\theta}(x)$ , alors l'application  $x \mapsto \hat{\theta}(x)$  est appelée maximum de vraisemblance et  $\hat{\theta}(X)$  est l'estimateur de maximum de vraisemblance de  $\theta$ .

$$\hat{\theta} = \arg \max_{\theta} L(\theta; X).$$

On préfère maximiser le logarithme de la vraisemblance :

$$\hat{\theta} = \arg \max_{\theta} \ln(L(\theta; X)).$$

La densité de probabilité d'une variable aléatoire  $X \sim \mathcal{GH}(A, B, g, h)$  en prenant en compte le seuil  $H$ :

$$X = A + B \left( \frac{e^{gz} - 1}{g} \right) e^{hz^2/2} \quad (3)$$

$$\begin{aligned} f_{X|H}(x|A, B, g, h) &= f_{Z|H}(z|A, B, g, h) \left| \frac{dz}{dx} \right| \\ &= \frac{f_{Z|H}(z|A, B, g, h)}{\left| \frac{dx}{dz} \right|} \\ (4) \quad &= \frac{\varphi \left[ Y^{-1} \left( \frac{x-A}{B} \right) \right]}{B \cdot Y' \left[ Y^{-1} \left( \frac{x-A}{B} \right) \right] * \left[ 1 - \Phi \left[ Y^{-1} \left( \frac{H-A}{B} \right) \right] \right]} \mathbb{1}_{\{x \geq H\}} \end{aligned}$$

Cette densité n'aboutit pas à une forme analytique explicite en fonction de  $x$  et doit par conséquent être évaluée numériquement.

Le calcul consiste d'abord à évaluer la réciproque de la fonction  $Y$  en " $\frac{x-A}{B}$ ", puis à substituer les solutions obtenues dans l'équation (4). Etant donné un échantillon indépendant et identiquement distribué  $x_1, x_2, \dots, x_n$ , la vraisemblance des données sous la loi  $g\&h$  est :

$$L_X(A, B, g, h|x) = \prod_{i=1}^n \frac{\varphi \left[ Y^{-1} \left( \frac{x_i - A}{B} \right) \right]}{B \cdot Y' \left[ Y^{-1} \left( \frac{x_i - A}{B} \right) \right] * \left[ 1 - \Phi \left[ Y^{-1} \left( \frac{H - A}{B} \right) \right] \right]} \mathbb{I}_{\{x \geq H\}}$$

Lorsque  $h > 0$ , la vraisemblance peut être maximisée en utilisant des procédures numériques. Sur Matlab, la fonction *fminsearch* permet d'effectuer ce calcul à l'aide de l'algorithme du simplexe de Nelder-Mead (utilisé par Rayner et MacGillivray [9]).

Le paramètre  $A$  ne peut plus être immédiatement estimé par la médiane de l'échantillon. La maximisation s'effectue donc avec quatre paramètres :  $A, B, g$  et  $h$ . Néanmoins le fait de combiner cette optimisation avec l'inversion invoquée précédemment rend la procédure complexe trop lente. Lorsque  $h$  est négatif,  $x$  n'étant plus une fonction monotone de  $z$ , l'obtention de cet estimateur devient difficile. Même le cas de la loi log-normale ( $h = 0$ ) peut s'avérer délicat à traiter puisque les intervalles de confiance de  $h$  sont centrés autour de zéro.

Nous avons implémenté numériquement la méthode du maximum de vraisemblance. Or la fonction  $Y$  de Tukey utilisée n'est pas inversible analytiquement, cela provoque beaucoup de difficultés car même avec une version optimisée de cette inversion la méthode reste très lente pour des échantillons de taille raisonnable. Les complications viennent du fait que l'algorithme (Nelder-Mead) de maximisation de la vraisemblance utilisé fait appel à cette inversion ( $n$  fois pour un échantillon de taille  $n$ ) à chaque itération. De plus, prendre en compte le seuil de collecte, donc l'ajout d'un facteur  $1/(1 - F_\theta(H))$  à toutes les contributions des observations à la vraisemblance, un terme qui peut devenir très grand pour certaines valeurs des paramètres, complexifie considérablement l'optimisation. La méthode du maximum de vraisemblance n'est donc pas utilisable en pratique pour ajuster une distribution  $g\&h$  sur des échantillons de taille conséquente ou pour faire des études type Monte-Carlo qui nécessitent un grand nombre d'estimations. Néanmoins il est tout à fait possible d'améliorer ses performances en parallélisant la recherche du maximum de la fonction de vraisemblance.

## 2.4 Méthode des moments généralisée (GMM)

La méthode des moments généralisée consiste à déterminer le vecteur de paramètres qui minimise l'écart entre les moments théoriques et les moments empiriques. À l'ordre  $p$ , les conditions de moments sont définies par :

$$G_p(\theta) = \frac{1}{n} \sum_{k=1}^n ((x_k)^p - E(X^p))$$

Où  $x_k$  est la  $k^{\text{ième}}$  perte et  $E(X^p)$  est le moment d'ordre  $p$ , ce dernier dépend de  $\theta$ .

Martinez et Iglewicz [8] ont obtenu le moment d'ordre  $p$  par rapport à l'origine, pour  $h \leq \frac{1}{p}$  et  $g \geq 0$  :

$$(5) \quad E(X^p) = \sum_{i=0}^p \binom{p}{i} A^{p-i} B^i \frac{\sum_{r=0}^i (-1)^{-r} \binom{i}{r} e^{\{(i-r)g\}^2 / \{2(1-ih)\}}}{g^i \sqrt{1-ih}}$$

La méthode GMM présente plusieurs propriétés intéressantes en particulier celle de convergence, puisque elle ne fournit pas de résultats aberrants comme la méthode du maximum vraisemblance.

La contrainte " $h \leq \frac{1}{p}$ " se traduit par un ordre de moment maximal  $p \leq \frac{1}{h}$ . Cette contrainte est difficile à respecter en risque opérationnel car, dans l'étude de Dutta et Perry 2007 [6], ils trouvent que  $h \in [0.1; 0.4]$ , ce qui correspond à  $p \in [2.5; 10]$ , étant donné que nous avons besoin d'au moins quatre moments pour estimer nos paramètres, la formule (5) ne va plus être valable pour au moins 15% des cas.

On observe expérimentalement que pour  $g \geq 1$ , les moments d'ordre supérieur ou égal à 1 sont très variables par rapport aux moments empiriques, ceci rend la méthode imprécise, vu qu'en risque opérationnel  $g \in [1.7; 2.3]$  (selon Dutta et Perry 2007 [6]).

Cette méthode est en plus sensible au seuil de collecte  $H$ , cela aura tendance à rendre les contraintes citées précédemment encore plus difficiles à respecter.

Ces limites font de la GMM une méthode inutilisable en pratique pour estimer les paramètres des distributions  $\mathcal{G}\&\mathcal{H}$  appliquées aux risques opérationnels.

## 2.5 Méthode de la distance-quantile (QD)

La forme, transformée de la loi normale, de la  $\mathcal{G}\&\mathcal{H}$  la rend adaptée aux méthodes d'estimation qui consistent à minimiser une distance particulière entre quantiles empiriques et quantiles théoriques. L'idée de base est de choisir les paramètres du modèle qui rendent minimale cette distance.

Pour un échantillon de  $n$  pertes  $\xi_1, \dots, \xi_n$ , nous nous intéressons à la distance quadratique entre  $k$  quantiles empiriques  $\hat{q}(p_1), \dots, \hat{q}(p_k)$  et  $k$  quantiles théoriques  $F_\theta^{-1}(p_i)$ , où  $F_\theta^{-1}$  est la fonction de répartition inverse et  $\theta$  le vecteur de paramètres de la loi à ajuster. La distance s'écrit :

$$Q^2(\theta, p, \omega) = \sum_{i=1}^k \omega_i \left( \hat{q}(p_i) - F_\theta^{-1}(p_i) \right)^2$$

Où  $\left( \omega_i = \frac{1}{\hat{q}(p_i)^2} \right)_{i=1..k}$  est un vecteur de pondération et  $(p_i)_{i=1..k}$  est le vecteur des niveaux des quantiles à ajuster, avec  $0 < p_1 < \dots < p_k < 1$ .

Les quantiles empiriques  $\hat{q}$  sont construits à partir du vecteur de pertes de l'échantillon, en fonction du vecteur  $p = (p_1, \dots, p_k)'$ . Le  $i^{\text{ème}}$  quantile empirique correspond à la  $i^{\text{ème}}$  perte de l'échantillon trié  $\xi_1^*, \dots, \xi_n^*$  si le nombre  $n * p_i$  est entier, et à une interpolation linéaire entre les deux pertes les plus proches sinon.

Le terme de pondération  $\omega_i$  sert à limiter l'instabilité numérique liée aux termes d'ajustement des quantiles extrêmes dans l'expression de la distance.

L'objectif de l'estimation est donc de trouver le paramètre  $\theta = \hat{\theta}_{QD}$  minimisant la distance  $Q^2(\theta, p, w)$ , afin d'obtenir une bonne adéquation entre les pertes observées (quantiles empiriques) et les pertes estimées par le modèle paramétrique (quantiles théoriques).

Dans le cadre de la loi  $\mathcal{G}\&\mathcal{H}$ , et le vecteur de paramètres à ajuster est donc  $\theta = (A, B, g, h)$ , et la fonction quantile vaut

$$F^{-1}(\alpha) = A + B \cdot [Y(\Phi^{-1}(\alpha))]$$

avec :

$$Y(z) = \left( \frac{e^{gz} - 1}{g} \right) e^{hz^2/2}$$

### Intégration du seuil

La prise en compte du seuil de collecte  $H$ , se traduit par une transformation des niveaux de quantiles à ajuster. En effet, le quantile de niveau  $\alpha$  de l'échantillon tronqué correspond en fait à un niveau de quantile  $\alpha^H$  de la distribution théorique complète.

Or on a :

$$\tilde{F}_{\theta|H}^{-1}(\alpha) = F^{-1} \left[ \alpha + (1 - \alpha) \cdot \Phi \left[ Y^{-1} \left( \frac{H - A}{B} \right) \right] \right]$$

Nous en déduisons le lien entre  $\alpha^H$  et  $\alpha$ :

$$\alpha^H = \alpha + (1 - \alpha) \cdot F_{\theta}(H)$$

En conséquence, nous modifions les quantiles théoriques à ajuster, ce qui implique que l'expression de la distance à minimiser est :

$$Q^2(\theta, p, \omega) = \sum_{i=1}^k \frac{1}{\hat{q}(p_i)^2} (\hat{q}(p_i) - F_{\theta}^{-1}[p_i + (1 - p_i)F_{\theta}(H)])^2$$

Avec :  $F_{\theta}(H) = \Phi \left[ Y^{-1} \left( \frac{H - A}{B} \right) \right]$

Néanmoins, cette formule nécessite la connaissance du paramètre  $\theta = (A, B, g, h)$ , afin d'évaluer le niveau de quantile décalé  $\alpha^H$ . Cela rend l'optimisation impossible. Afin de résoudre ce problème V. Lehérisse et A. Renaudin [13] utilise un estimateur  $\tilde{\theta}$  de  $\theta$  dans le terme correcteur  $F_{\theta}(H)$ .

L'estimateur  $\tilde{\theta}$  est obtenu par minimisation d'une distance-quantile tenant compte du seuil de collecte en décalant linéairement les quantiles empiriques vers l'origine :

$$\hat{q}^H(p_i) = \hat{q}(p_i) + (1 - p_i)H$$

Cette modification permet d'utiliser des quantiles théoriques de la distribution  $\mathcal{G}\&\mathcal{H}$  non tronquée et donc d'éviter tout problème numérique dans la minimisation de la distance :

$$\tilde{\theta} = \arg \min_{\theta} \sum_{i=1}^k \frac{1}{\hat{q}(p_i)^2} ((\hat{q}(p_i) - (1 - p_i) \cdot H) - F_{\theta}^{-1}(p_i))^2$$

Une fois  $\tilde{\theta}$  ainsi déterminé, l'estimateur final  $\hat{\theta}_{QD}$  du paramètre de la distribution de sévérité est obtenu en minimisant la distance suivante :



$$\hat{\theta}_{QD} = \arg \min_{\theta} \sum_{i=1}^k \frac{1}{\hat{q}(p_i)^2} \left( \hat{q}(p_i) - F_{\theta}^{-1} \left( p_i + (1 - p_i) F_{\hat{\theta}}(H) \right) \right)^2$$

La méthode distance-quantile, appliquée à loi log-normale dans [13] donne des résultats intéressants par rapport aux autres méthodes d'estimation. Dans la section suivante, nous prenons soin de comparer les différentes méthodes d'estimation décrites précédemment, même celles que nous avons qualifiées d'inadaptées, afin de juger de leurs précisions, stabilités et temps de calcul de façon quantitative.

*Pour aller plus loin :* D'après Dutta et Perry (2007) [6], dans de rares cas, le fait de prendre  $g$  et  $h$  constants, ne permet pas d'avoir un bon ajustement. Dans ces situations-là  $g$  et  $h$  peuvent être pris comme fonctions polynomiales de  $z_p^2$ . Une forme plus générale des lois  $\mathcal{G}\&\mathcal{H}$ :

$$X_{g,h} = A + B \left( \frac{e^{g_Z \cdot Z} - 1}{g_Z} \right) e^{h_Z \cdot (Z^2/2)}$$

Avec  $g_Z$  et  $h_Z$  des fonctions polynomiales en  $Z^2$ .

Par exemple :

$$\begin{aligned} g_Z &= \gamma_0 + \gamma_1 Z^2 + \gamma_2 Z^4 \\ h_Z &= \eta_0 + \eta_1 Z^2 + \eta_2 Z^4 + \eta_3 Z^6 \end{aligned}$$

Dans le même papier, il a été démontré que seule la généralisation du paramètre  $h$  est intéressante pour avoir de meilleurs ajustements et qu'à partir de quatre coefficients pour  $h$  on n'a plus de gain en adéquation. Plus loin dans l'étude on étudiera le cas où  $g$  est constant et  $h$  est polynomiale et l'on discutera le nombre de paramètres optimal pour  $h$ .

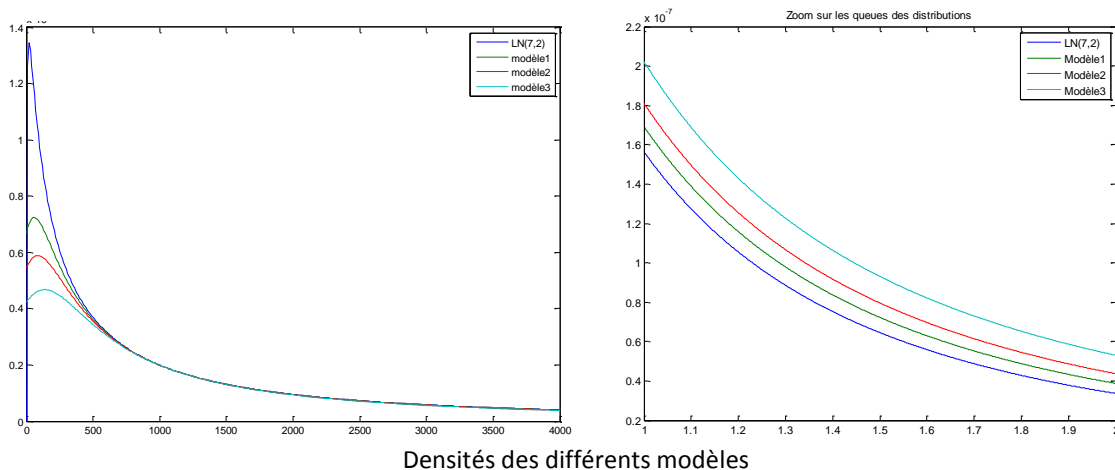
Remarque : Aux propriétés souhaitées évoquées dans la partie (I-5-1) s'ajoute un 5<sup>ème</sup> point important, signalé par A. Chernobai dans [10] : Le pourcentage de perte en-dessous du seuil de collecte, qui est lui aussi estimé par le modèle par la valeur de la fonction de répartition avec les paramètres  $\theta$  estimés, évaluée au seuil de collecte  $H$  ( $F_{\theta}(H) = P(\text{Perte} \leq H)$ ). Cette valeur doit être acceptable pour la banque, par exemple le cas  $F_{\theta}(H) \geq 0.6$  signifie que le modèle a considéré que les données manquantes correspondent à au moins 60% de la totalité de la distribution, ce qui est, d'un point de vue métier, incorrect.

## 2.6 Propriétés théoriques des méthodes d'estimation

### 2.6.1 Sur données non-tronquées

Dans un premier temps, nous comparons les estimateurs IQ, QD, ML et GMM dans un environnement simulé. Nous simulons des échantillons selon une loi  $\mathcal{G}\&\mathcal{H}(A, B, g, h)$  pour différentes valeurs de  $\theta = (A, B, g, h)$ . Les estimations de paramètres sont moyennées sur  $N = 1000$  simulations. Afin de représenter les cas de figures les plus courants en risque opérationnel. Nous choisissons  $(A, B, g)$  à partir de paramètres  $(\mu, \sigma)$ , de la loi  $\mathcal{LN}(\mu, \sigma)$ , communément observés et on prendra un paramètre  $h$  différent pour chaque modèle afin de représenter l'effet queue épaisse qui caractérise cette catégorie de risques. Nous allons utiliser les jeux de paramètres ci-dessous et chaque échantillon est constitué de  $n = 5000$  pertes :

Modèle	$A$	$B$	$g$	$h$
1	$\exp(7)$	$2 * \exp(7)$	2	0.05
2	$\exp(7)$	$2 * \exp(7)$	2	0.1
3	$\exp(7)$	$2 * \exp(7)$	2	0.2



On voit bien sur le zoom que l'on effectue sur les queues de distributions que des pertes de l'ordre de  $10^5$  sont de plus en plus probables en allant du modèle log-normal au modèle 3 ( $h = 0.2$ ).

### Biais relatif

Le biais de l'estimateur  $\hat{\theta}_n$  de  $\theta$  est défini par :

$$\mathbb{E}(\hat{\theta}_n) - \theta$$

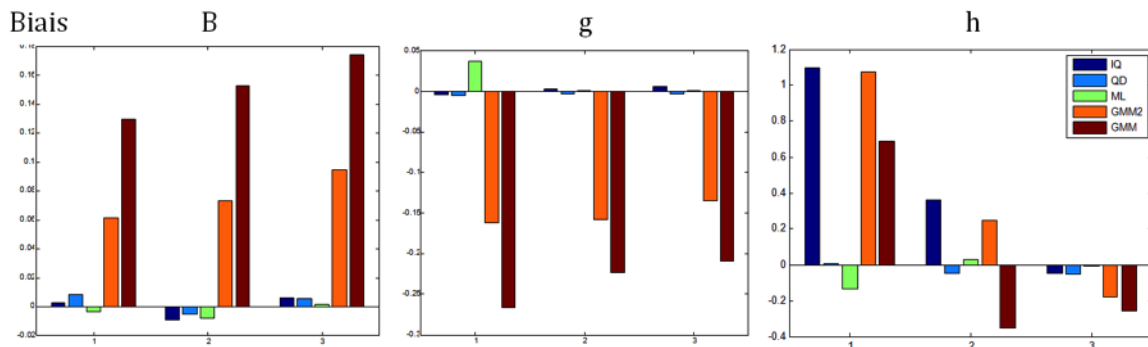
Cette espérance sera calculée en moyennant les estimateurs  $\hat{\theta}_n^1, \dots, \hat{\theta}_n^N$  obtenus sur  $N$  estimations distinctes :

$$b(\hat{\theta}_n) = \frac{1}{N} \sum_{i=1}^N \hat{\theta}_n^i - \theta$$

Les valeurs de biais sont ainsi déterminées pour chacune des méthodes et pour les trois paramètres  $B, g$  et  $h$ , car le paramètre  $A$  est estimé par la médiane de l'échantillon et donc est le même quelle que soit la méthode.

Nous appelons la méthode GMM<sup>2</sup> une méthode des moments généralisée utilisant les deux premiers moments seulement. C'est-à-dire que pour estimer nos trois paramètres, nous n'utilisons que la moyenne et la variance.

Les figures ci-dessus représentent les biais relatifs  $b(\hat{\theta}_n)/\theta$  pour chacun des paramètres  $B, g$  et  $h$ .



L'existence de contraintes difficiles à respecter sur les moments théoriques fait que les moments ne sont utilisables que pour certaines valeurs des paramètres, ceci entrave donc la maximisation car pendant l'optimisation, des valeurs « fausses » de la fonction objectif seront comparées.

Les résultats ci-dessus le montrent, en effet les biais sont plus grands en utilisant les trois premiers moments (GMM) qu'en utilisant seulement les deux premiers (GMM<sup>2</sup>). Cela prouve que les moments d'ordres supérieurs à deux ne sont pas utilisables et par conséquent, la méthode sera encore moins précise pour des données tronquées (impliquant la nécessité d'un quatrième moment).

Nous constatons les biais relativement faibles des méthodes ML et QD, ainsi que le comportement stable de cette dernière, car elle n'est pas affectée par le manque de données en queue de distribution (modèle1) contrairement à la ML qui affiche un biais plus élevé pour la même distribution.

## Précision

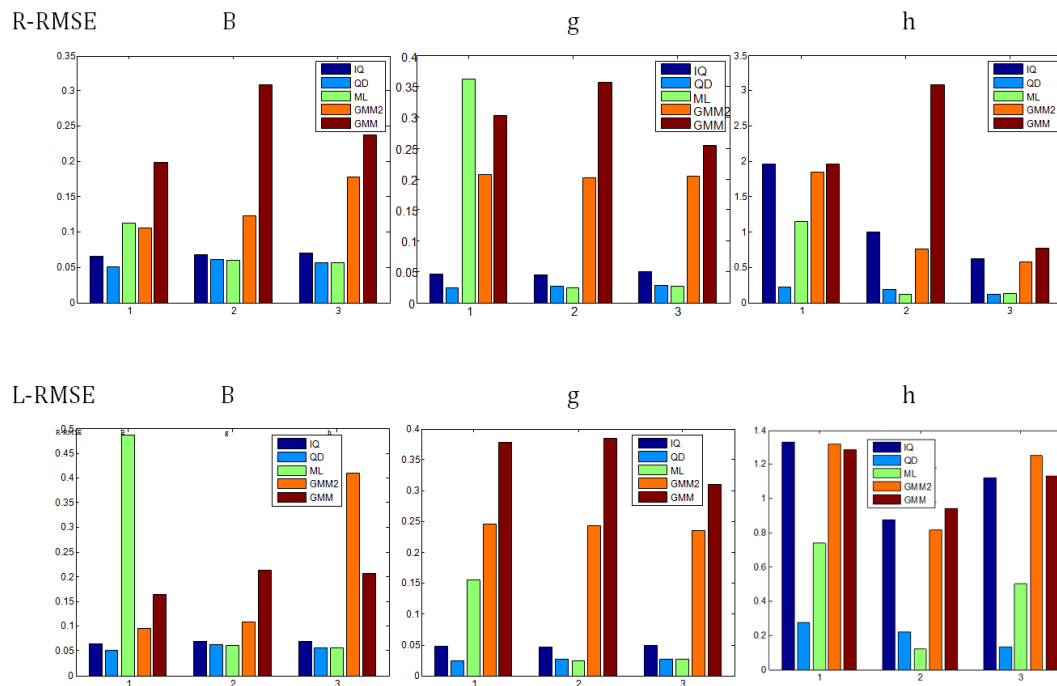
Cette partie présente deux indicateurs utilisés pour évaluer la précision des estimations obtenues par différentes méthodes. Le premier des deux indicateurs déterminé pour chacun des 3 modèles simulés, est la racine de l'erreur quadratique moyenne relative (*relative root mean square error*, R-RMSE) :

$$R - RMSE(\hat{\theta}_n) = \frac{1}{\theta} \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_n^i - \theta)^2}$$

Le second indicateur reprend l'idée du premier en utilisant une échelle logarithmique :

$$L - RMSE(\hat{\theta}_n) = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \ln \frac{\hat{\theta}_n^i}{\theta} \right)^2}$$

Cette échelle logarithmique octroie plus de poids aux erreurs de sous-estimation qu'aux erreurs de surestimation (contrairement au précédent indicateur où les poids sont symétriques)



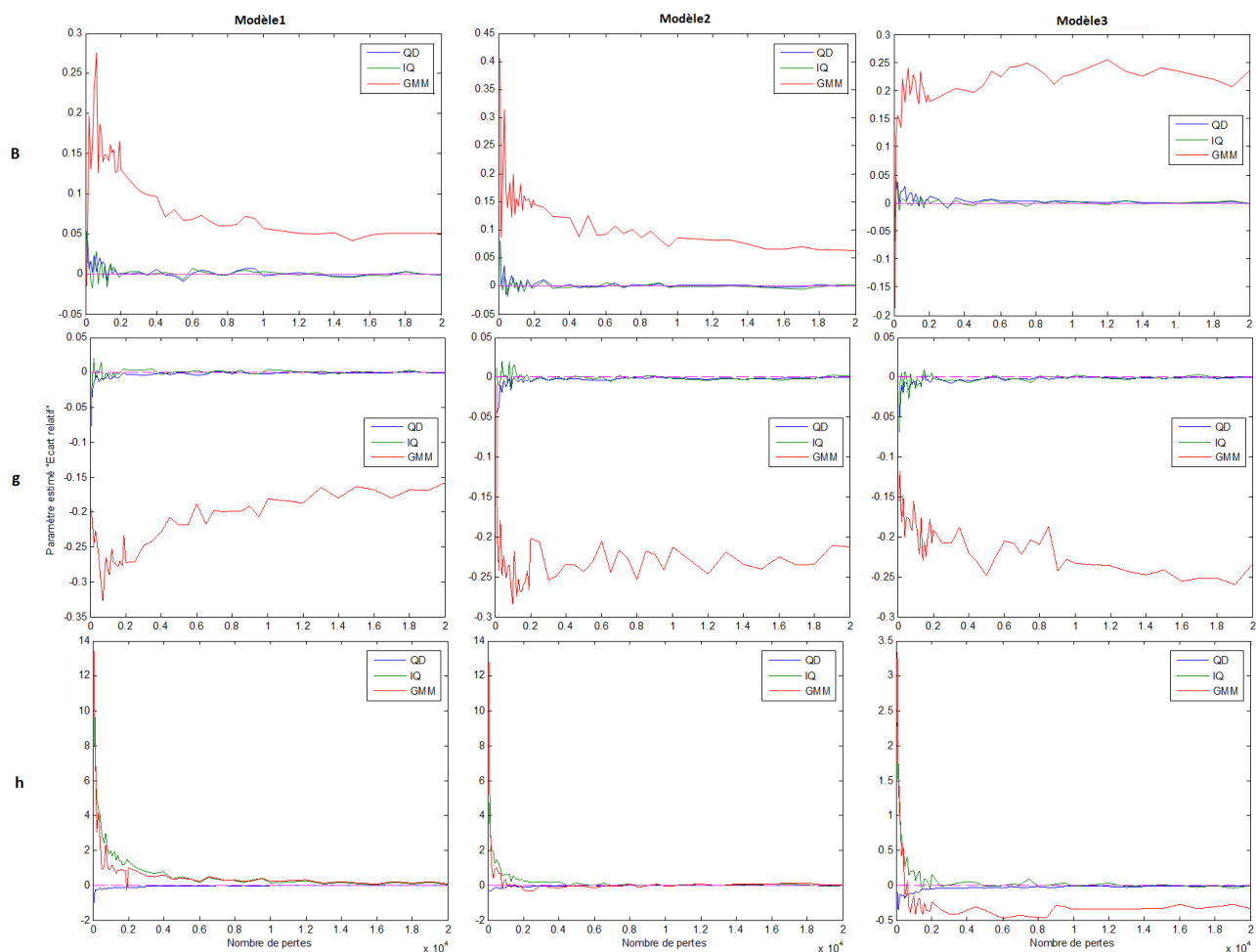
Les résultats de ces deux indicateurs appuient les remarques précédentes et nous encourage plus à privilégier la méthode QD qui montre une précision incomparable à celle des autres méthodes d'estimation.

### Convergence

Nous présentons dans cette partie le comportement des estimations, pour chacun des modèles traités précédemment, en fonction du nombre de pertes de l'échantillon.

Pour ce faire, nous constituons des échantillons comportant un nombre de pertes allant de 50 à 20 000, puis nous estimons les paramètres de ces échantillons avec les différentes méthodes comparées.

Nous obtenons pour les différents modèles les graphiques ci-dessous : (ils nous a été impossible de joindre la méthode ML à la comparaison à cause de sa complexité temporelle)



Les résultats de ces graphes valident ceux de la partie précédente. En effet la méthode GMM converge très lentement pour les paramètres  $B$  et  $g$  et ne converge pas du tout pour le paramètre  $h$ , ce qui rend légitime le rejet de la GMM en tant que méthode d'estimation pour les lois  $\mathcal{G}\&\mathcal{H}$ . Il faut également noter que pour des échantillons de petite taille, seule la QD se comporte de façon acceptable.

## Conclusion

La méthode des moments généralisée est très rapide, car elle ne fait intervenir à chaque itération de la minimisation de la distance entre moments empiriques et moments théoriques que le calcul de ces derniers qui se fait à l'aide d'une formule fermée. Cependant, à cause des raisons évoquées précédemment, elle ne converge pas. Ce résultat est validé par les résultats de la convergence qui exhibent l'instabilité de la méthode.

Les méthodes dites à quantiles donnent quant à elles des résultats intéressants que ce soit en termes de biais, de précision ou de convergence. De plus ces méthodes utilisent la fonction quantile de la distribution  $\mathcal{G}\&\mathcal{H}$ , qui n'est qu'une transformation de la fonction quantile de la loi normale centrée réduite. Elles sont donc très adaptées à l'estimation des paramètres des lois  $\mathcal{G}\&\mathcal{H}$ .

Seule la méthode QD sera retenue par la suite. Dans la prochaine partie, nous la développons de façon plus approfondie et nous la testons sur données simulées et tronquées, puis sur données réelles. Pour des fins de comparaison, nous utilisons la méthode ML, malgré sa complexité temporelle.

### 2.6.2 Sur données tronquées

Dans cette partie nous comparons les performances des méthodes ML et QD, sur données tronquées à gauche à 20%. Nous gardons les mêmes modèles que dans la partie précédente. En revanche nous diminuons le nombre de pertes par échantillon à 1000 (800 après troncature) pour s'approcher du cas des données réelles. Nous moyennons cette fois-ci sur  $N = 100$  simulations afin de pouvoir obtenir des résultats pour la méthode du maximum de vraisemblance qui est extrêmement lente. Nous commençons par introduire quelques précisions sur la manière dont les procédures numériques seront utilisées.

#### Maximum de vraisemblance

La fonction à maximiser est :

$$L_X(A, B, g, h|x) = \prod_{i=1}^n \frac{\varphi \left[ Y^{-1} \left( \frac{x_i - A}{B} \right) \right]}{B \cdot Y' \left[ Y^{-1} \left( \frac{x_i - A}{B} \right) \right] \cdot \left[ 1 - \Phi \left[ Y^{-1} \left( \frac{H - A}{B} \right) \right] \right]} \mathbb{1}_{\{x \geq H\}}$$

Il est plus judicieux de minimiser :

$$L = -\log( L_X(A, B, g, h|x) )$$

$$L = \sum_{i=1}^n \left[ \log \left( \varphi \left[ Y^{-1} \left( \frac{x_i - A}{B} \right) \right] \right) - \log \left( B \cdot Y' \left[ Y^{-1} \left( \frac{x_i - A}{B} \right) \right] \right) - \log \left( 1 - \Phi \left[ Y^{-1} \left( \frac{H - A}{B} \right) \right] \right) \right]$$

L'inverse  $Y^{-1}$  doit être calculé  $n$  fois à chaque itération, nous avons donc pris le soin de détailler, en annexe, les méthodes numériques utilisées pour effectuer cette opération. L'algorithme choisi pour cette optimisation est le Trust-Region-Modified-Dogleg Algorithm (TRMDA). Cependant ce dernier ne garantit pas que le minimum trouvé soit un minimum global. Il est possible de combiner cet algorithme à une méthode de recuit simulé afin d'augmenter les chances de tomber sur un minimum global. Des travaux, présentés en annexe, ont été effectués concernant ce sujet.

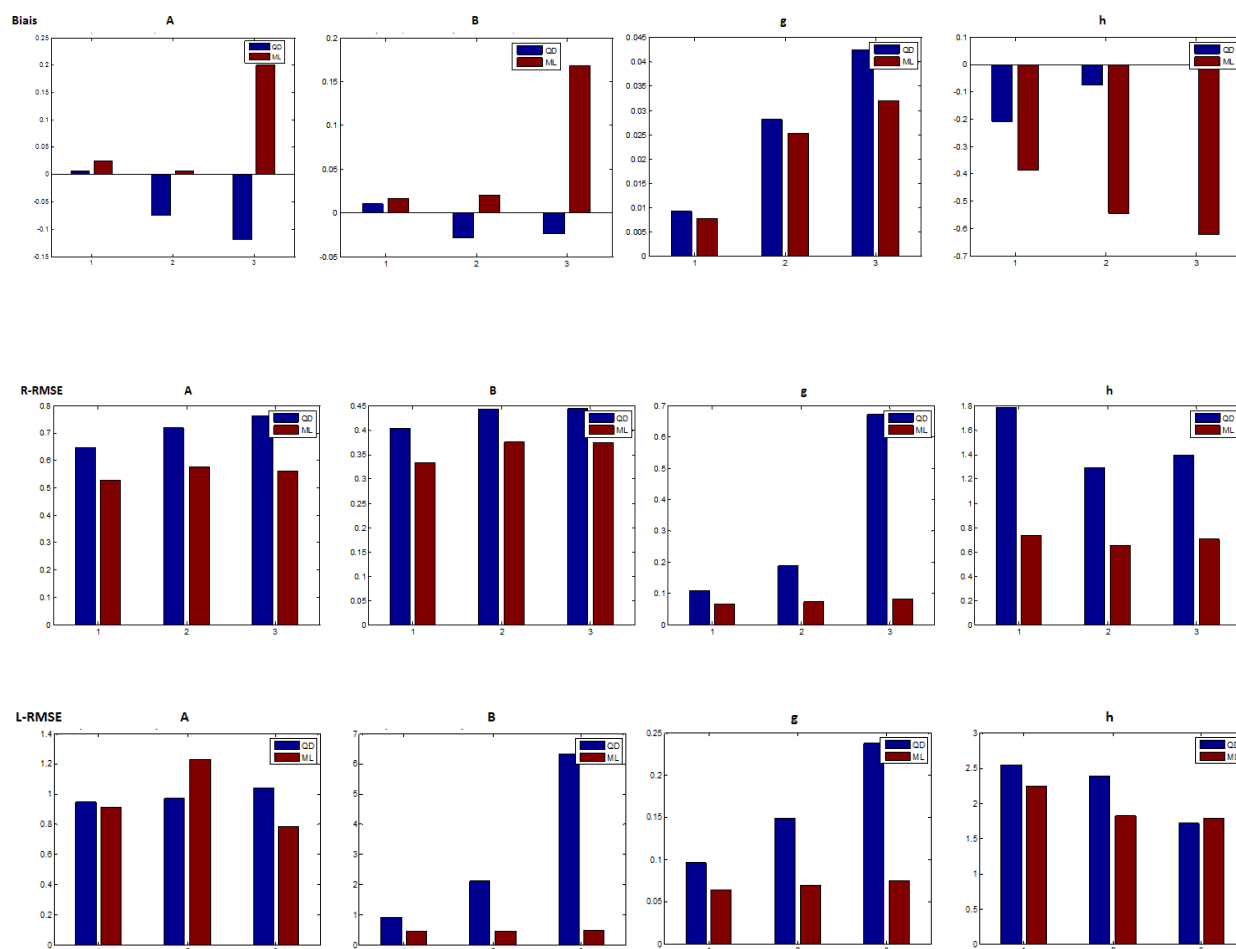
#### Distance-quantile

Cette procédure se fait en deux étapes :

- A l'aide de l'algorithme du simplexe de Nelder-Mead on fait une première estimation de  $\theta$ , en approchant la valeur de la fonction de répartition au point  $H$  (le seuil de collecte) par  $(\hat{q}(p_i) + (1 - p_i)H)$  comme précisé dans la section traitant de cette méthode d'estimation.
- On utilise le résultat précédent pour approcher  $F_\theta(H)$  et on minimise, toujours par le même algorithme, la distance :

$$\hat{\theta}_{QD} = \arg \min_{\theta} \sum_{i=1}^k \frac{1}{\hat{q}(p_i)^2} \left( \hat{q}(p_i) - F_\theta^{-1} \left( p_i + (1 - p_i) F_{\hat{\theta}}(H) \right) \right)^2$$

## Résultat



Ces résultats confirment ce que l'on a affirmé à propos de méthode QD dans le cadre de données non tronquées, elle est en effet plus stable et moins biaisée que la méthode ML, en plus elle présente un temps de calcul négligeable devant celui de cette dernière.

### 3. Déploiement du modèle en environnement réel

Nous appliquons le modèle sur données réelles de pertes opérationnelles d'un groupe bancaire européen. Les données sont d'une profondeur de 5 ans, présentent des distributions positivement asymétriques ( $Skewness > 0$ ) et à queue épaisse ( $Kurtosis > 3$ ) et contiennent une matrice de pertes pour les entités Ent1, Ent2 et Ent3 et les sept catégories de risques introduites dans Bâle II.

**Pour des raisons de confidentialité, toutes les données traitées ont été modifiées par le biais de constantes multiplicatives.**

### 3.1 Statistiques descriptives des pertes

		Nombre de lignes	Seuil	Max	Moyenne	Ecart type	Skewness	Kurtosis
ENT1	EL1	111	1000	496459	28065	70009	5	28
	EL2M	1027	1	76475	1313	4681	11	140
	EL2HM	1054	1000	1031159	12947	46659	14	259
	EL3	529	1000	1585394	25303	88432	15	238
	EL4	1570	1000	4353721	22884	162601	20	467
	EL5	356	1000	421225	14414	40312	7	58
	EL6	202	1000	875166	9648	62522	14	187
	EL7	5503	1000	8102283	11212	168183	37	1533
ENT2	EL1	183	1000	1821612	68385	225229	5	33
	EL2M	29389	1	312000	1315	3059	48	4077
	EL2HM	2664	1000	3627614	22469	106989	21	598
	EL3	374	1000	100500352	4571057	8751742	5	43
	EL4	4146	1000	14630616	47650	360562	27	926
	EL5	3988	1000	1877857	5824	36275	39	1849
	EL6	646	1000	652817	10339	38631	11	151
	EL7	11536	1000	9613226	17137	134715	39	2412
ENT3	EL2M	89	1000	22974	3921	3415	2	13
	EL2HM	55	1000	233414	19631	47530	3	13
	EL3	89	1000	4013108	333290	636494	4	19
	EL4	182	1000	10160150	295472	1462323	6	38
	EL6	1307	1000	3591510	15092	113683	26	770
	EL7	7476	1000	44703038	32414	617068	56	3811

### 3.2 Estimation des paramètres

Nous estimons les paramètres de la loi  $\mathcal{G}\&\mathcal{H}$  en utilisant la méthode QD avec une distribution uniforme des quantiles choisis, car étant donné que la loi  $\mathcal{G}\&\mathcal{H}$  capte à l'aide de son paramètre d'aplatissement  $h$  l'effet de queue, il n'est pas nécessaire de garder la distribution exponentielle de ces quantiles utilisée pour la loi  $\mathcal{LN}$ .

Il nous a été difficile d'estimer les paramètres du modèle  $\mathcal{G}\&\mathcal{H}$  par la méthode du maximum de vraisemblance à cause de sa complexité temporelle sur échantillon de taille relativement grande. Ainsi nous comparons un modèle  $\mathcal{G}\&\mathcal{H}$  estimé par (QD) à un modèle log-normal estimé par (QD) et enfin à un modèle log-normal estimé par (GMM) car c'est le modèle utilisé actuellement en interne. L'inhomogénéité des données et leur taille limitée ont fait que la méthode de maximum de vraisemblance ne converge pas pour la loi log-normale et donne des valeurs aberrantes pour le paramètre  $\mu$  (valeurs négatives).

$\mathcal{G}\&\mathcal{H}$ (QD)					$\mathcal{LN}$ (GMM)			$\mathcal{LN}$ (QD)		
$\log A$	$\log B$	$g$	$h$	$F_{\theta}(H)$	$\mu$	$\sigma$	$F_{\theta}(H)$	$\mu$	$\sigma$	$F_{\theta}(H)$



ENT1	EL1	7,79	8,54	2,80	-0,45	22%	9,14	1,43	6%	7,93	2,16	32%
	EL2M	5,99	6,39	1,26	-0,05	2%	5,85	1,14	0%	5,69	1,70	0%
	EL2HM	7,73	8,05	1,67	0,13	26%	7,33	1,81	41%	7,65	1,69	33%
	EL3	8,42	9,27	2,18	-0,66	22%	8,61	1,65	15%	8,66	1,48	12%
	EL4	7,34	7,50	2,02	0,32	33%	7,09	2,16	47%	6,92	2,15	50%
	EL5	7,42	7,71	2,95	-0,62	24%	8,05	1,57	23%	7,46	1,91	39%
	EL6	7,24	7,32	1,10	0,46	38%	5,09	2,28	79%	7,05	1,51	46%
ENT2	EL7	7,12	7,05	0,64	0,75	41%	2,77	2,82	93%	6,80	1,48	53%
	EL1	7,89	8,66	3,06	-0,29	20%	9,84	1,58	3%	7,85	2,59	36%
	EL2M	6,53	6,79	1,07	-0,04	1%	2,53	2,65	17%	6,12	1,39	5,16E-06
	EL2HM	7,49	7,89	3,01	-0,54	24%	7,93	1,87	29%	7,52	2,07	38%
	EL3	13,42	14,15	1,98	-0,02	2%	14,57	1,24	3,31E-10	13,28	2,01	0%
	EL4	7,34	7,35	3,47	-0,34	22%	8,30	2,09	25%	6,79	2,66	52%
	EL5	7,21	7,02	0,98	0,50	36%	2,58	2,52	96%	6,89	1,42	50%
ENT3	EL6	7,29	7,26	1,92	0,18	31%	6,93	1,82	49%	6,93	1,84	50%
	EL7	7,14	6,72	2,19	0,29	31%	6,33	2,24	60%	6,41	2,06	60%
	EL2M	7,45	7,58	1,63	-0,55	24%	7,85	0,80	12%	7,64	0,90	21%
	EL2HM	6,95	4,30	7,42	-2,65	38%	8,49	1,48	14%	5,80	2,94	65%
	EL3	11,35	11,99	1,72	-0,08	6%	11,95	1,24	2,23E-05	11,39	1,71	0%
	EL4	7,91	8,37	2,62	0,63	22%	10,96	1,80	1%	6,95	3,06	49%
	EL6	7,54	7,68	1,44	0,42	29%	6,15	2,24	63%	7,32	1,72	41%
	EL7	7,55	7,64	1,22	0,72	30%	5,78	2,64	66%	7,26	1,77	42%

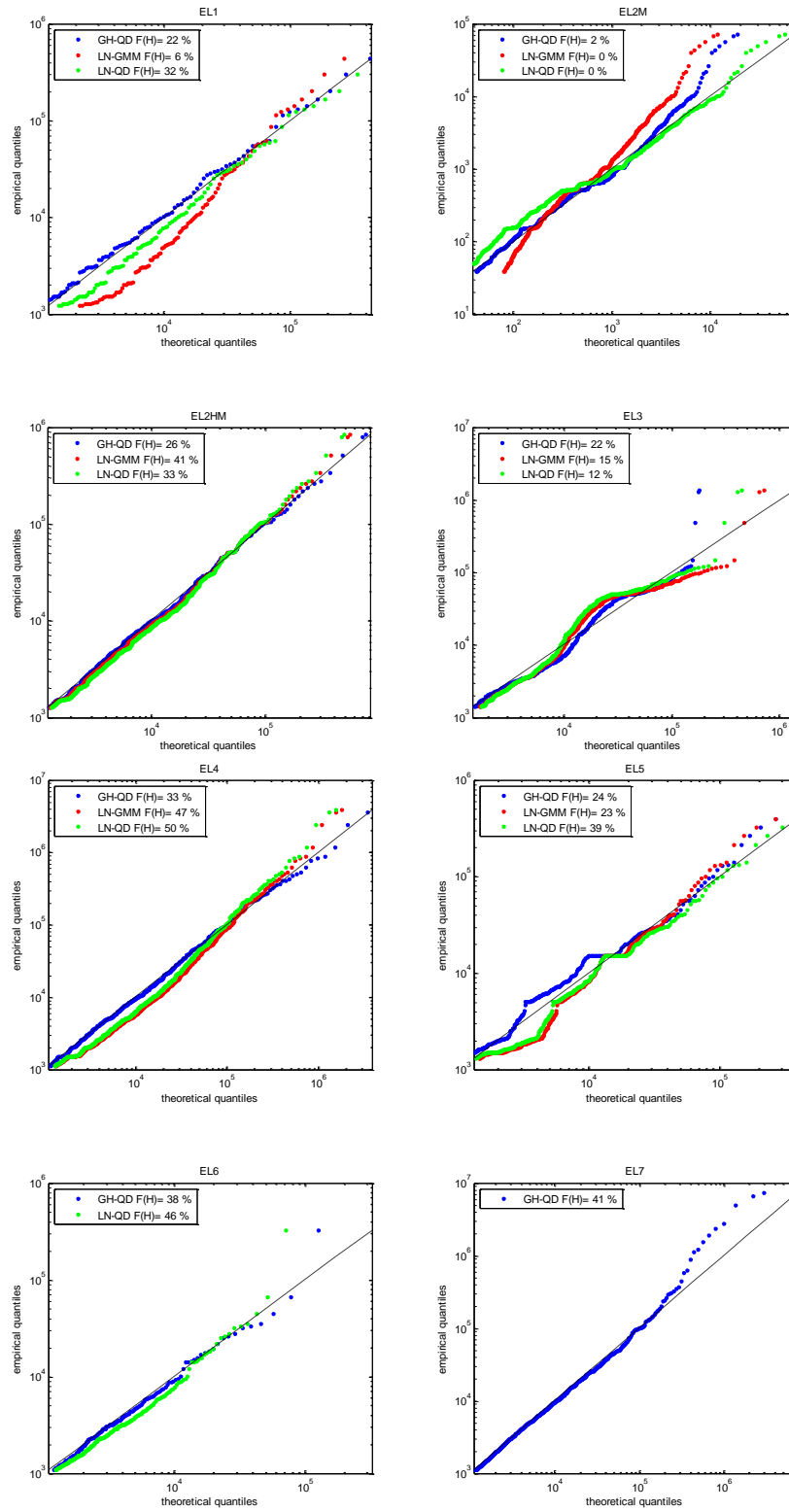
Nous marquons en rouge les valeurs de  $F_\theta(H) \geq 50\%$  car ce terme représente le pourcentage de donnée en dessous du seuil de collecte, et on estime qu'au-delà de 50% le modèle n'est pas réaliste. Nous remarquons que ce cas de figure se produit plus souvent lorsque l'on utilise une log-normale alors qu'avec la distribution  $\mathcal{G}\&\mathcal{H}$  cela n'arrive jamais.

### 3.3 Critères d'ajustement

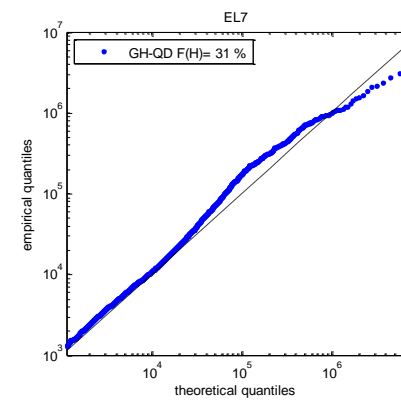
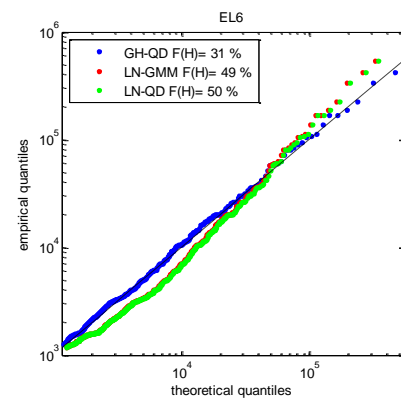
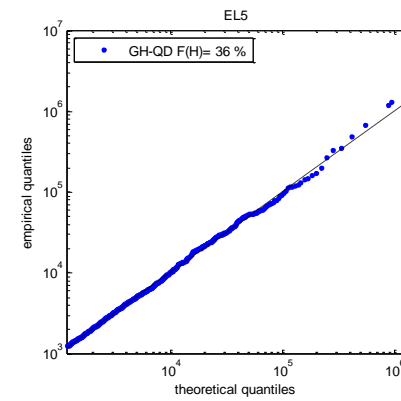
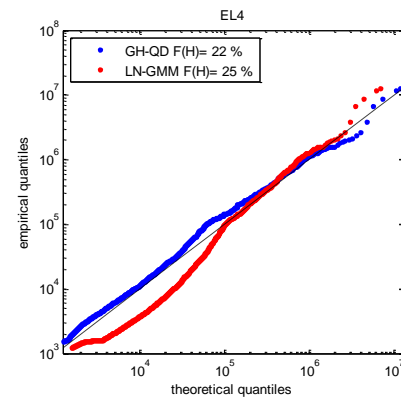
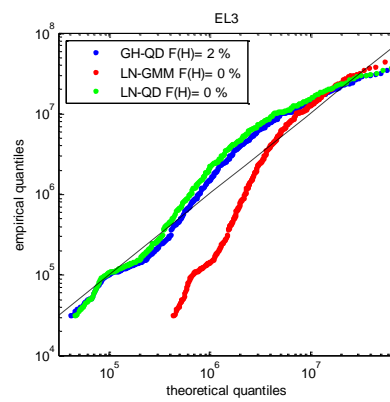
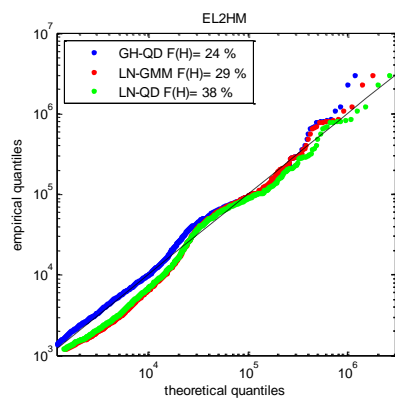
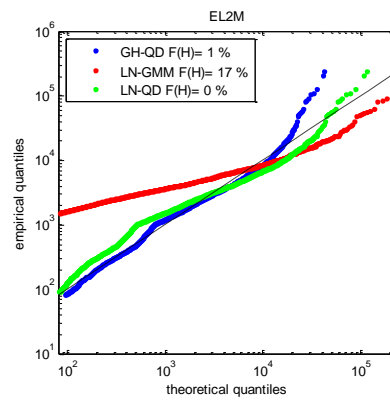
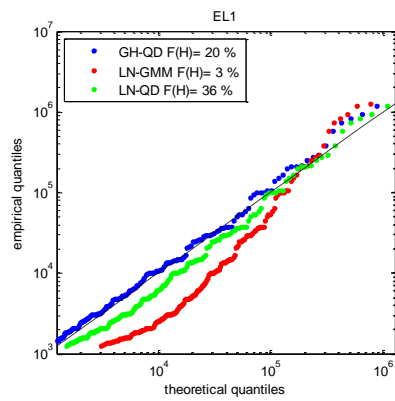
#### QQ-Plots

Seuls les QQ-plots obtenus avec des paramètres permettant d'observer un  $F_\theta(H)$  inférieur à 50% seront représentés.

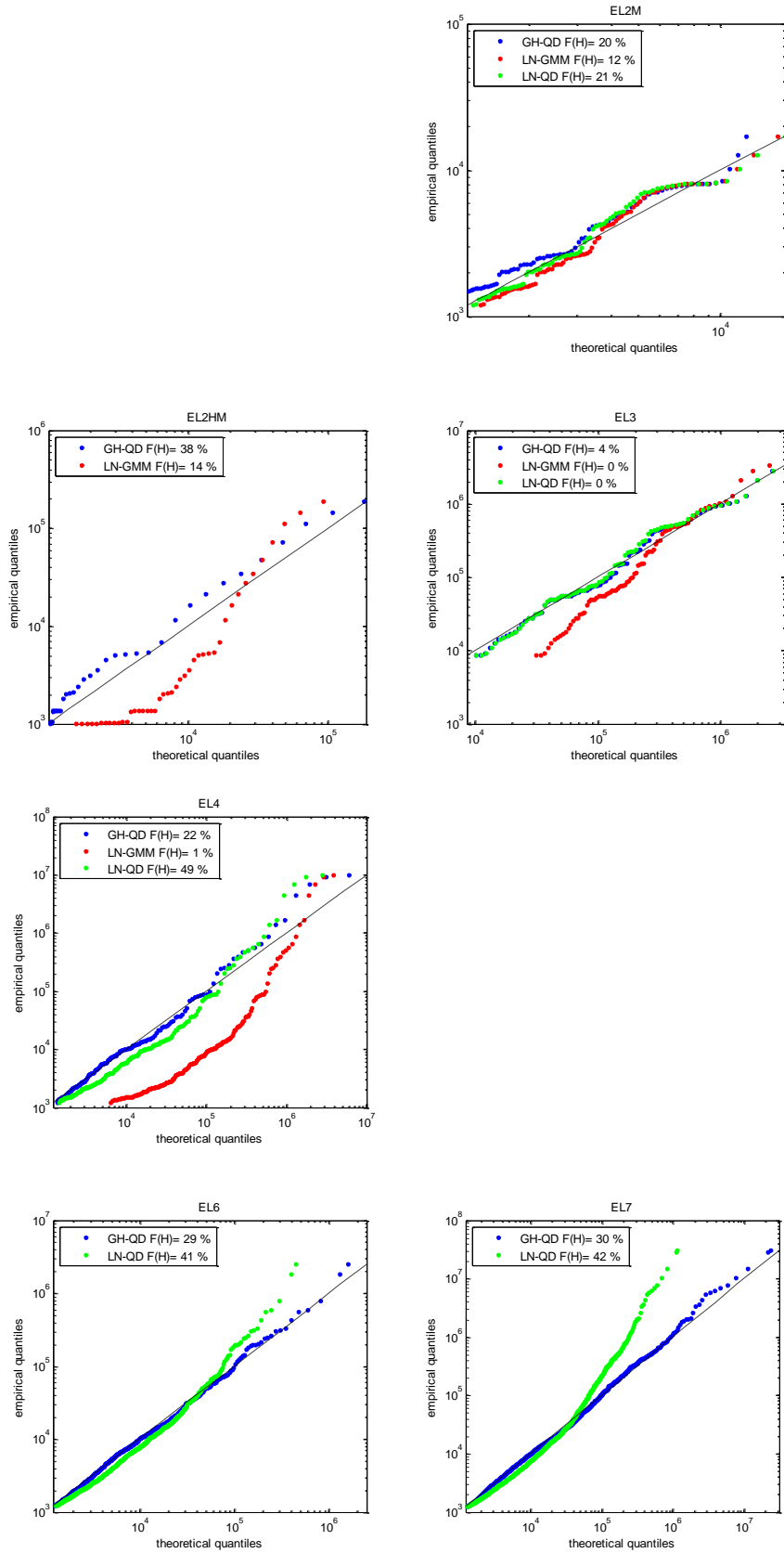
ENT1 :



ENT3 :



ENT3 :



On remarque que, dans la majorité des cas, les QQ-plots obtenus en modélisant les données de sévérités par une loi  $\mathcal{GH}$  se rapprochent plus de la première bissectrice que ceux obtenus par une

loi log-normale.

Pour des données, à queue de distribution « très » épaisse (EL7 par exemple), la modélisation obtenue par log-normale n'est pas utilisable et les QQ-plots correspondant ne peuvent être interprétés. Tandis qu'avec la  $\mathcal{G}\&\mathcal{H}$ , on obtient des QQ-plots se rapprochant clairement de la première bissectrice.

### Tests d'adéquation

Les quatre tests d'adéquation utilisés (et détaillés dans la première partie) sont ceux couramment déployés dans le cadre du calcul des exigences de fonds propres :

- Test de Kolmogorov-Smirnov (KS)
- Test de Cramer-von-Mises (CVM)
- Test d'Anderson-Darling (AD)
- Test d'Anderson-Darling modifié ( $AD_{up}$ )

L'algorithme de ces tests décrit dans la partie I-5-4 fait appel à la fonction de répartition de la distribution étudiée un grand nombre de fois. Dans le cas de la  $\mathcal{G}\&\mathcal{H}$ , il nécessite l'inversion de la fonction  $Y$  de Tukey, cela rend le calcul fastidieux.

Pour y remédier, lors de la génération d'un échantillon de loi  $\mathcal{G}\&\mathcal{H}$  on génère d'abord un échantillon de loi normale centrée réduite, ce dernier contient les valeurs cherchées en inversant la fonction de Tukey. Il suffit alors d'effectuer les mêmes manipulations sur les vecteurs d'observations pour garder une trace des inverses de Tukey. Nous effectuons les changements nécessaires sur la procédure qui génère des échantillons de loi  $\mathcal{G}\&\mathcal{H}$  pour qu'elle renvoie, sans calculs supplémentaires, les valeurs de la fonction de répartition correspondant au vecteur généré.

Résultats :

		$\mathcal{G}\&\mathcal{H}$ par QD				$\mathcal{LN}$ par GMM				$\mathcal{LN}$ par QD			
		KS	CVM	AD	Adup	KS	CVM	AD	Adup	KS	CVM	AD	Adup
ENT1	EL1	0,89	0,80	0,86	0,89	0,00	0,00	0,00	0,02	0,01	0,01	0,00	0,45
	EL2M	0,03	0,01	0,01	0,03	0,03	0,03	0,01	0,00	0,00	0,00	0,00	0,02
	EL2HM	0,11	0,05	0,14	0,11	0,00	0,00	0,00	0,04	0,00	0,00	0,00	0,17
	EL3	0,80	0,79	0,46	0,80	0,28	0,33	0,29	0,00	0,00	0,00	0,00	0,01
	EL4	0,18	0,12	0,26	0,18	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,06
	EL5	0,99	0,97	0,99	0,99	0,00	0,00	0,00	0,03	0,00	0,00	0,00	0,24
	EL6	0,12	0,21	0,17	0,12	0,79	0,63	0,55	0,12	0,04	0,00	0,00	0,01
	EL7	0,01	0,02	0,02	0,0520	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ENT2	EL1	0,88	0,80	0,85	0,88	0,00	0,00	0,00	0,02	0,01	0,01	0,00	0,49
	EL2M	0,02	0,01	0,01	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02
	EL2HM	0,10	0,05	0,14	0,10	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,17
	EL3	0,82	0,80	0,43	0,82	0,00	0,00	0,00	0,06	0,00	0,00	0,00	0,01
	EL4	0,18	0,12	0,26	0,18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,07
	EL5	0,1270	0,2060	0,1960	0,1270	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,27
	EL6	0,12	0,22	0,17	0,12	0,00	0,00	0,00	0,01	0,04	0,00	0,00	0,00
	EL7	0,02	0,02	0,03	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ENT3	EL2M	0,25	0,33	0,24	0,25	0,02	0,03	0,04	0,07	0,18	0,10	0,05	0,11
	EL2HM	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,21
	EL3	0,39	0,20	0,26	0,39	0,00	0,00	0,00	0,09	0,15	0,27	0,26	0,42

EL4	0,54	0,70	0,60	0,54	0,00	0,00	0,00	0,00	0,03	0,00	0,00	0,13
EL6	0,22	0,20	0,21	0,22	0,60	0,58	0,59	0,21	0,00	0,00	0,00	0,02
EL7	0,00	0,00	0,00	0,0930	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

P-values des quatre tests d'adéquation pour la  $\mathcal{G}\&\mathcal{H}$  estimé par QD, la  $\mathcal{LN}$  par GMM telle qu'elle est utilisée en production

Nous confirmons l'information apportée par les QQ-plots. En effet, en donnant plus de flexibilité au modèle à l'aide des deux paramètres additionnels de la loi  $\mathcal{G}\&\mathcal{H}$ , nous avons un meilleur ajustement.

### 3.4 Charges en capital

Le tableau ci-dessous regroupe les charges en capital ( $Var_{99,9}$ ) correspondant aux données réelles étudiées.

			$\mathcal{G}\&\mathcal{H}$ (QD)		$\mathcal{LN}$ (GMM)		$\mathcal{LN}$ (QD)	
		$\lambda$ Exp	$\lambda$ Corr	CaR ( $\lambda$ Corr)	$\lambda$ Corr	CaR ( $\lambda$ Corr)	$\lambda$ Corr	CaR ( $\lambda$ Corr)
ENT1	EL1	12	17	14 291 215 €	13	2 481 763 €	18	12 369 068 €
	EL2M	13500	13662	15 013 218 €	13500	9 433 542 €	13503	21 998 996 €
	EL2HM	180	242	13 348 690 €	304	7 539 183 €	268	6 521 684 €
	EL3	200	245	9 659 585 €	236	14 544 415 €	227	8 281 903 €
	EL4	390	584	303 172 948 €	732	38 690 198 €	777	32 338 113 €
	EL5	120	173	15 850 738 €	156	4 744 681 €	195	10 297 433 €
	EL6	40	64	7 210 165 €	189	4 099 174 €	74	966 858 €
	EL7	1050	1789	315 069 308 €	14719	58 593 406 €	2228	7 595 356 €
Total				693 615 867 €		140 126 361 €		100 369 411 €
ENT2	EL1	37	49	139 390 809 €	38	14 074 629 €	58	118 699 802 €
	EL2M	28500	28721	42 904 802 €	34330	34 413 390 €	28500	53 378 288 €
	EL2HM	490	701	66 014 765 €	693	28 384 890 €	794	43 426 746 €
	EL3	73	74	3 906 813 726 €	73	771 048 163 €	73	3 014 975 758 €
	EL4	740	921	668 349 114 €	990	112 555 629 €	1536	362 579 665 €
	EL5	710	1107	42 206 293 €	16529	13 600 726 €	1434	4 982 099 €
	EL6	130	188	21 917 250 €	257	4 986 763 €	257	5 327 674 €
	EL7	2100	3047	806 814 943 €	5270	78 250 075 €	5199	44 988 051 €
Total				5 694 411 702 €		1 057 314 265 €		3 648 358 084 €
ENT3	EL2M	12	17	134 470 €	14	124 155 €	15	130 830 €
	EL2HM	12	28	65 124 773 €	14	1 592 874 €	34	45 100 092 €
	EL3	25	25	99 317 711 €	25	30 169 780 €	25	85 810 496 €
	EL4	55	71	24 886 917 223 €	56	116 736 744 €	109	533 561 435 €
	EL6	300	423	109 885 559 €	816	22 715 271 €	505	7 488 434 €
	EL7	1900	2697	4 615 941 095 €	5670	292 407 366 €	3287	33 133 209 €
Total				29 777 320 831 €		463 746 190 €		705 224 496 €

Nous relevons une hausse considérable des charges en capital sur la plupart des catégories qui, lorsqu'elles sont modélisées par une log-normale, ne valident pas les tests d'adéquation.

Il faut noter que certaines catégories contiennent des pertes représentant des scénarios internalisés. La méthode d'internalisation a ses limites, étant donné qu'elle ajoute à l'échantillon en question de grandes pertes censées représenter des scénarios catastrophiques, ces pertes ne sont pas forcément compatibles avec les données collectées et peuvent avoir un impact colossal sur l'estimation des paramètres, en particuliers sur des échantillons de petites tailles. Dans ce cas-là, une charge en capital jugée très grande ne l'est pas vraiment car, sur un 'petit' échantillon, elle modélise bien le fait qu'une perte 'extrême' survient avec une fréquence non négligeable et ne fait pas d'elle un élément rare. Il est donc indispensable de traiter avec un plus grand soin ces pertes car nous avons désormais un paramètre ( $h$ ) très sensible et qui a un impact direct sur la charge en capital.

## IV- Conclusion

---

Nous avons montré par le présent mémoire l'excellente qualité d'ajustement qu'apportent les distributions  $G\&H$  par leur flexibilité, néanmoins, pour avoir des résultats plus robustes, une plus grande quantité de données est nécessaire car avec peu de données en queue de distribution le paramètre  $h$  qui mesure l'aplatissement n'est plus fiable et peut causer une charge en capital aléatoirement explosive. Il faut donc être très vigilant vis-à-vis des données utilisées, en traitant des pertes aberrantes et en mettant des critères d'homogénéité plus sévères sur la méthodologie d'internalisation des scénarios qui peuvent contaminer les échantillons.

Nous avons bien noté la divergence de quelques méthodes d'estimation et l'instabilité d'autres dans un contexte de données réelles très spécifiques et présentant de mauvaises propriétés statistiques, la méthode d'estimation des paramètres de sévérité adoptée se doit avant tout d'être robuste. C'est ainsi que la méthode basée sur une distance entre quantiles présentée dans le papier de V. Leherissé et A. Renaudin [13] présente des résultats très encourageants, paraît bien adaptée à la problématique d'estimation de sévérité du risque opérationnel et facile à mettre en place pour des distributions transformées (plus complexes à priori) de lois usuelles comme la  $G\&H$ .

L'utilisation de lois généralisées telles que la  $G\&H$  peut être étendue à d'autres risques. En effet, les données, présentant des structures d'asymétrie et d'aplatissement complexes, sont fréquemment traitées dans la littérature financière. Dans leur article, Badrinath et Chatterjee [18] réalisent une étude des propriétés de la loi de rendements boursiers en utilisant les lois  $G\&H$ . Ils montrent que, pour des périodes de temps suffisamment longues, l'indice boursier peut être modélisé efficacement par une loi  $G\&H$ . Ils décrivent une procédure d'ajustement simple et robuste aux valeurs aberrantes, et illustrent leur approche au moyen de données quotidiennes et mensuelles provenant du CRSP (Center for Research in Security Prices). Ensuite, dans l'article, [19] les mêmes auteurs utilisent les lois  $G\&H$  pour explorer la nature de l'asymétrie et de l'aplatissement des distributions des rendements boursiers quotidiens de plusieurs entreprises individuelles. Une fois de plus, l'ajustement de la loi  $G\&H$  est meilleur que celui de la loi log-normale utilisée jusqu'à présent. Les distributions  $G\&H$  ont également eu un excellent succès dans la modélisation des taux d'intérêt à court terme. En effet Dutta et Babbel [4] ont observé que les lois des taux LIBOR 1 mois et 3 mois peuvent être décrites de façon précise à l'aide de ces distributions. Les mêmes auteurs appliquent ces résultats dans [20] pour obtenir une formule explicite d'évaluation des options européennes, comparable au modèle de Black-Scholes en simplicité et supérieure en ajustement au marché.

Finalement, certains papiers ont récemment abordé la question de la modélisation des risques opérationnels. Plusieurs nouvelles distributions ont été testées, telles que la  $G\&H$  la théorie des valeurs extrêmes (souvent à moyenne infinie) et la NIG (introduite en annexe). Dutta et Perry [6] en 2007 effectuent une évaluation détaillée des méthodes couramment employées et de nouvelles techniques de mesure de ce risque basées sur la théorie des valeurs extrêmes, l'échantillonnage empirique ou l'ajustement d'une loi paramétrique. Ces différentes approches sont ensuite appliquées à des données de pertes internes de sept établissements financiers américains et comparées au moyen de diverses mesures de performance, ce qui permet aux auteurs de conclure que la loi  $G\&H$  offre un meilleur rendement que les autres modèles évalués. Notre étude vient confirmer les résultats de Dutta et Perry sur des données européennes.



## V- Bibliographie

---

- [1] Basel Committee, 2011, Operational risk- Supervisory guidelines for the advanced measurement approaches.
- [2] J. KARAMATA sur un mode croissance régulière, *Mathematica (Cluj)* Vol.4 (1930) 38-53
- [3] Julie DROUIN, août 2008, Les lois g-and-h: définition, propriétés, généralisations et applications.
- [4] Dutta, Kabir K. and David F. Babbel, "On Measuring Skewness and Kurtosis in Short Rate Distributions: The Case of the US Dollar London Inter-Bank Offered Rates," 2002. Wharton Financial Institutions Center Working Paper.
- [5] Hoaglin, D. C., "Summarizing Shape Numerically: The g-and-h Distributions." Chapter 11 in *Exploring Data Tables Trends, and Shapes*. Eds. Hoaglin, Mosteller, and Tukey. New York,
- [6] Kabir Dutta and Jason Perry A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital 2007
- [7] Matthias Degen, Paul Embrechts Dominik & D. Lambrigger (April 2007): The Quantitative Modeling of Operational Risk: Between g-and-h and EVT.
- [8] Martinez, Jorge and Boris Iglewicz, "Some Properties of the Tukey g and h Family of Distributions," *Communications in Statistics - Theory and Methods*, 1984, 13 (3), 353{369.
- [9] Rayner, G. D. and H. L. MacGillivray, "Numerical Maximum Likelihood Estimation for the g-and-h and Generalized g-and-h Distributions," *Statistics and Computing*, 2002,12, 57{75.
- [10] Chernobai, Rachev, Bertocchi, Consigli, and Giacometti: Heavy-Tailed Distributional Model for Operational Losses March 16, 2007
- [11] Tukey, John W., *Exploratory Data Analysis*, Reading, MA: Addison-Wesley, 1977a.
- [12] MacGillivray, H. L. and W. Cannon (1997). Generalizations of the g-and-h distributions and their uses. *Unpublished*.
- [13] Vincent Leherissé, Alexis Renaudin, Quantile distance estimation for operational risk, Working paper 2012
- [14] CHERNOBAI A., RACHEV S., FABOZZI F., 2005, *Composite Goodness-of-Fit Tests for Left-Truncated Loss Samples*, Department of Statistics and Applied Probability, University of California, USA
- [15] Claude LONGCHAMP 2009, Optimisation numérique : Méthode de la région de confiance Chapitre 4.1 et 4.2
- [16] A.R. Conn, N.I.M. Gould, Ph.L. Toint (2000). Trust-Region Methods. MPS/SIAM Series on Optimization 1. SIAM and MPS, Philadelphia.
- [17] Eriksson, A., L. Forsberg, and E. Ghysels (2004): "Approximating the Probability Distribution of Functions of Random Variables: A New Approach," Discussion paper CIRANO.
- [18] Badrinath, S. G. and S. Chatterjee, " On Measuring Skewness and Elongation in Common Stock Return Distributions: The Case of the Market Index." *Journal of Business*, Vol.61,No.4,1988.
- [19] Badrinath, S. G. and S. Chatterjee, "A Data-Analytic Look at Skewness and Elongation in Common-Stock-Return Distributions." *Journal of Business & Economic Statistics*, Vol. 9,No.2, 1991.
- [20] Kabir K. Dutta and David F. Babbel 2005 *Extracting Probabilistic Information from the Prices of Interest Rate Options: Tests of Distributional Assumptions*

## VI- Annexes :

---

### 1. Réciproque de la fonction $Y$ de Tukey

La majorité des résultats de la modélisation par des lois  $\mathcal{G}\&\mathcal{H}$  se composent de la réciproque  $Y^{-1}$  de la fonction de Tukey ci-dessous :

$$Y(z) = \left( \frac{e^{gz} - 1}{g} \right) e^{hz^2/2}$$

Cette fonction n'admet pas de réciproque sous forme analytique, nous avons donc testé plusieurs méthodes numériques afin de choisir la façon la plus optimale en temps de calcul et en précision de la calculer.

Rayner et MacGillivray (2002a) utilisent la méthode du simplexe de Nelder-Mead. Nous comparons cette méthode à l'algorithme de la recherche par la section d'or (Golden Search Algorithm), et enfin à un algorithme de région de confiance (Trust-Region-Dogleg Algorithm).

#### 1.1 Présentations des algorithmes :

Dans cette partie nous donnons un bref aperçu de ces algorithmes utilisés pour approcher la réciproque de  $Y$ . Pour plus de détail consulter l'aide de Matlab. Le but de l'opération est de trouver  $z$  qui minimise la fonction objectif :

$$f(z) = \left| Y - \left( \frac{e^{gz} - 1}{g} \right) e^{hz^2/2} \right|$$

On cherche évidemment à que ce minimum soit le plus possible proche de zéro.

*Les définitions ci-dessous sont extraites en partie du travail de Claude LONGONCHAMP [15], Wikipédia, et le manuel de Matlab.*

##### 1.1.1 Simplexe de Nelder-Mead (NMA)

###### Principe

La méthode de Nelder-Mead est un algorithme d'optimisation non-linéaire qui a été publiée par Nelder et Mead en 1965. C'est une méthode numérique heuristique qui cherche à minimiser une fonction continue dans un espace à plusieurs dimensions. L'algorithme exploite le concept de simplexe qui est un polytope de  $N + 1$  sommets dans un espace à  $N$  dimensions. Partant initialement d'un tel simplexe, celui-ci subit des transformations simples au cours des itérations : il se déforme, se déplace et se réduit progressivement jusqu'à ce ses sommets se rapprochent d'un point où la fonction est localement minimale.

###### Algorithme

Soit  $N$  la dimension de l'espace où la fonction objectif  $f$  prend ses valeurs. L'algorithme débute par la définition d'un simplexe non dégénéré choisi dans cet espace. Par itérations successives, le processus consiste à déterminer le point du simplexe où la fonction est maximale afin de le

substituer par la réflexion de ce point par rapport au centre de gravité des  $N$  points restants. Si la valeur de la fonction en ce nouveau point est inférieure aux valeurs prises sur les autres points, le simplexe est étiré dans cette direction. Sinon, il est supposé que l'allure locale de la fonction est une vallée, et le simplexe est réduit par une similitude centrée sur le point du simplexe où la fonction est minimale.

Plus précisément :

#### Algorithme

1. Choix de  $N + 1$  points de l'espace à  $N$  dimensions des inconnues, formant un simplexe :  $\{x_1, x_2, \dots, x_{N+1}\}$ ,
2. Calcul des valeurs de la fonction  $f$  en ces points, réindexation des points de façon à avoir  $f(x_1) \leq f(x_2) \leq \dots \leq f(x_{N+1})$ . Il suffit en fait de connaître le premier et les deux derniers.
3. Calcul de  $x_0$ , centre de gravité de tous les points sauf  $x_{N+1}$ .
4. Calcul de  $x_r = x_0 + (x_0 - x_{N+1})$  (réflexion de  $x_{N+1}$  par rapport à  $x_0$ ).
5. Si  $f(x_r) < f(x_N)$ , calcul de  $x_e = x_0 + 2(x_0 - x_{N+1})$  (étirement du simplexe). Si  $f(x_e) < f(x_r)$ , remplacement de  $x_{N+1}$  par  $x_e$ , sinon, remplacement de  $x_{N+1}$  par  $x_r$ . Retour à l'étape 2.
6. Si  $f(x_N) < f(x_r)$ , calcul de  $x_c = x_{N+1} + \frac{1}{2}(x_0 - x_{N+1})$  (contraction du simplexe). Si  $f(x_c) < f(x_N)$ , remplacement de  $x_{N+1}$  par  $x_c$  et retour à l'étape 2, sinon aller à l'étape 7.
7. Similitude de rapport  $\frac{1}{2}$  et de centre  $x_1$  : remplacement de  $x_i$  par  $x_1 + \frac{1}{2}(x_i - x_1)$  pour  $i \geq 2$ . Retour à l'étape 2.

#### Avantages

- La généralité : une fonction continue (sans évaluer ses dérivées).
- La simplicité de la mise en œuvre.
- L'efficacité pour une fonction non dérivable.
- L'interprétation géométrique sous-jacente.
- L'assurance d'obtenir une série décroissante de valeurs.

#### Inconvénients

- S'applique mal (ou difficilement) lorsque le domaine de définition de la fonction est complexe ou que le minimum recherché se situe dans un voisinage de la frontière.
- La donnée « arbitraire » d'un simplexe de départ.
- Une dégradation des performances lorsque la dimension  $N$  augmente.
- Le risque que les simplexes obtenus successivement aient tendance à dégénérer (bien que l'expérience montre que ce soit rarement le cas).

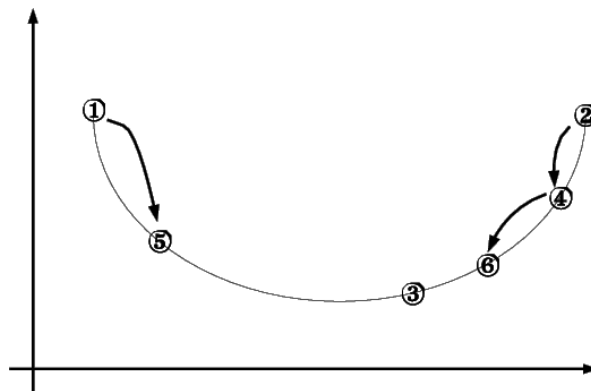
Cet algorithme va être utilisé par le biais de la fonction *fminsearch* de matlab.

#### 1.1.2 Recherche par la section d'or, Golden Search Algorithm (GSA)

## Principe

Pour obtenir une valeur approchée d'une racine d'une fonction il suffit de l'encadrer dans un intervalle  $[a, b]$ . Ceci est insuffisant pour caractériser un minimum. Nous aurons besoin dans ce cas de trois points  $(a, b, c)$ . Un intervalle tel que  $a < b < c$  et  $f(b) < f(a), f(b) < f(c)$  caractérise un minimum dans l'intervalle  $[a, c]$ . En d'autres termes pour encadrer un minimum nous avons besoin d'un triplé de point tel que le point central présente une valeur de  $f$  inférieure à celles des bornes de l'intervalle.

Le principe de cette recherche est résumé sur la figure ci-dessous. Le minimum est encadré à l'origine par  $(①, ③, ②)$ . La fonction est évaluée en  $④$ , qui remplace  $②$ . Ensuite en  $⑤$  qui remplace  $①$  puis en  $⑥$  qui remplace  $④$ . La règle est de garder un point central pour lequel la valeur de  $f$  est inférieure à celles des bornes. Après cette série d'itérations, le minimum est alors encadré par  $(⑤, ③, ⑥)$ . Le principe est analogue à celui utilisé lors d'une recherche de racine par dichotomie.



**Figure:** Encadrements successifs d'un minimum d'une fonction  $f$ .

Le seul point délicat, est de définir une méthode pour choisir un nouveau point d'encadrement dans l'intervalle  $(a, b, c)$  initial. Supposons que  $b$  soit une fraction  $w$  du segment  $[a, c]$ , alors

$$\frac{b-a}{c-a} = w \Leftrightarrow \frac{c-b}{c-a} = 1-w$$

Si l'on choisit un nouveau point  $x$  éloigné d'une fraction  $z$  par rapport à  $b$

$$\frac{x-b}{c-a} = z$$

On montre alors simplement que le nouveau point  $x$  est le symétrique de  $b$  dans son intervalle d'origine, c'est à dire:  $|b-a| = |x-c|$ . Cela signifie que le point  $x$  se trouve dans le plus grand des segments  $[a, b]$  et  $[b, c]$ . Reste maintenant à définir la position du point  $x$  à l'intérieur de ce segment. La valeur  $w$  provient d'une étape précédente de calcul, et si l'on suppose qu'elle est optimale, alors  $z$  doit être choisie de la même manière. Cette similarité d'échelle implique que  $x$  doit être situé à la même fraction du segment  $[b, c]$  (si  $[b, c]$  est le segment le plus long) que  $b$  l'était du segment  $[a, c]$ . Cela conduit à la relation

$$\frac{z}{1-w} = w$$

en combinant cette équation avec la définition de  $z$  on arrive à l'équation quadratique

$$w^2 - 3w \Leftrightarrow w = \frac{3 - \sqrt{5}}{2} \approx 0.38197$$

En d'autres termes, ce résultat signifie que l'intervalle d'encadrement  $(a, b, c)$  a son point central  $b$  est situé à une distance telle que

$$\frac{b - a}{c - a} = w \approx 0.38197$$

ou

$$\frac{c - b}{c - a} = 1 - w \approx 0.61803$$

Ces fractions sont celles de la moyenne d'or ou de la section d'or supposées avoir certaines propriétés esthétiques par les pythagoriciens du monde antique. C'est la raison pour laquelle cette méthode optimale de détermination du minimum d'une fonction est appelée recherche par section d'or (golden search section).

L'algorithme `fminbnd` décrit une implémentation optimisée de cette méthode. Et comme la valeur cherchée est censé représenter une réalisation d'une variable aléatoire normale centrée réduite ayant une probabilité de  $(1 - 9.8 * e^{-10})$  d'être dans l'intervalle  $[-6, 6]$ , il n'est donc pas nécessaire de chercher des valeurs en dehors de cet intervalle-là.

### Avantages

- La généralité : une fonction continue (sans évaluer ses dérivées).
- La simplicité de la mise en œuvre.
- L'efficacité pour une fonction non dérivable.
- L'assurance d'obtenir une série décroissante de valeurs.
- Bonne performances lorsque la dimension  $N$  augmente.

### Inconvénients

- S'applique mal (ou difficilement) lorsque le domaine de définition de la fonction est complexe ou que le minimum recherché se situe dans un voisinage de la frontière.
- Convergence lente quand la solution se situe au voisinage de la frontière

### 1.1.3 Région de confiance, Trust-Region-Dogleg Algorithm (TRDA)

La fonction objectif  $f$  est remplacée par une fonction modèle quadratique  $m_k$  dans une certaine région autour d'un point  $x_k$  donnée. Les deux premiers termes intervenant dans la fonction modèle  $m_k$  de chaque itéré  $x_k$  sont identiques aux deux premiers termes de la série de Taylor de  $f$  autour de  $x_k$ . On a donc :

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p$$

où  $f_k = f(x_k)$ ,  $\nabla f_k$  est le gradient de  $f$  au point  $x_k$  et  $B_k$  est une matrice symétrique représentant la hessienne de  $f$  à l'itéré  $x_k$  ou une approximation de ce dernier. Les régions de confiance ajoutent au problème d'optimisation initiale une contrainte sur la longueur du pas. On cherche la solution de chaque sous-problème de la forme :

$$\min_{p \in \mathbb{R}^n} m_k(p) \text{ sous la contrainte } \|p\| \leq \Delta_k \quad (*)$$

où  $\Delta_k > 0$  est le rayon de la région de confiance.

## Aperçu de l'algorithme

Pour un point  $x_k$  et un rayon  $\Delta_k$  donnés, on détermine l'efficacité de la fonction modèle par le rapport suivant:

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p)}$$

Ce rapport est utilisé comme critère d'actualisation du rayon  $\Delta_k$  de la région de confiance. On remarque que le dénominateur est forcément positif, car  $p_k$  est la solution de (\*), ce qui veut dire que  $m_k$  décroît. Si  $\rho_k < 0$ ,  $f(x_k + p_k) > f(x_k)$ , le pas doit alors être rejeté. Si  $\rho_k$  est positif, on accepte le pas et le rayon de la région de confiance est actualisé selon les valeurs prise par  $\rho_k$ . Si  $\rho_k$  est proche de 1, les fonctions modèles et objectifs sont en bon accord au cours de ce pas. Il est alors plus intéressant d'augmenter le rayon de la région de confiance pour la prochaine itération afin d'obtenir des pas plus performants. Pour finir, si  $\rho_k$  est proche de 0,  $m_k$  ne représente pas correctement  $f$ . On diminue donc le rayon de confiance.

## Le point de Cauchy

De manière similaire aux méthodes de recherches par lignes, la détermination de pas optimaux n'est pas une condition nécessaire pour obtenir une convergence globale. Bien que, en principe, on recherche une solution optimale du sous-problème (\*), il suffit de trouver une solution approchée  $p_k$  à l'intérieur de la région de confiance, qui produise une « réduction suffisante » de la fonction modèle. Cette réduction peut être obtenue par la méthode du point de Cauchy  $p_k^c$  :

### Algorithme :

1. Trouver un vecteur  $p_s^k$ , solution du problème (\*) linéarisé, soit  

$$p_s^k = \arg \min_{p \in \mathbb{R}^n} (f_k + \nabla f_k^T p) \text{ sous la contrainte } \|p\| \leq \Delta_k$$
2. Calculer le scalaire  $\tau_k > 0$  qui minimise  $m_k(\tau p_s^k)$ , soit  

$$\tau_k = \arg \min_{\tau > 0} m_k(\tau p_s^k) \text{ sous la contrainte } \|\tau p_s^k\| \leq \Delta_k$$
3. Finalement, on obtient le point de Cauchy, soit  

$$p_k^s = \tau_k p_s^k$$

## Amélioration du point de Cauchy par la méthode Dogleg

Bien que le point de Cauchy  $p_k^s$  fournisse une réduction suffisante de la fonction modèle  $m_k$  pour produire une convergence globale et que le coût du calcul est assez petit, il est intéressant de chercher une meilleure solution approximative de (\*). En effet, le point de Cauchy est défini comme le point minimisant  $m_k$  le long de la plus grande pente  $-\nabla f_k$ . C'est tout simplement la mise en œuvre de la méthode de la plus forte descente avec un choix particulier de la longueur de chaque pas. On considère trois méthodes pour trouver une solution approchée de (\*). Dans ce chapitre, on va se focaliser sur le travail interne d'une seule interaction. On enlève alors l'indice  $k$  de  $\Delta_k$ ,  $p_k$  et  $m_k$  pour simplifier la notation. Le sous-problème (\*) devient donc :

$$\min_{p \in \mathbb{R}^n} m(p) = f + g^T p + \frac{1}{2} p^T B_p p \text{ sous la contrainte } \|p\| \leq \Delta \quad (**)$$

où  $g$  est le gradient de  $f$ . On dénote la solution par  $p^*(\Delta)$  pour montrer la dépendance de  $\Delta$ .

Dans la méthode du Dogleg on examine l'effet du rayon  $\Delta$  de la région de confiance de la solution  $p^*(\Delta)$  du sous problème (\*\*). Si  $B$  est définie positive, on a toujours noté que le minimiseur sans contrainte de  $m$  est le pas intégral  $p^B = -B^{-1}g$ . Quand ce point est admissible pour (\*\*), c'est évidemment une solution, on a donc:

$$p^*(\Delta) = p^B$$

Quand  $\Delta$  est minuscule, la restriction  $\|p\| \leq \Delta$  garantit que le terme quadratique de  $m$  a un petit effet sur la solution de (\*\*). La solution  $p(\Delta)$  est approximativement la même solution que nous obtiendrions en minimisant la fonction linéaire  $f + g^T p$  sur  $\|p\| \leq \Delta$ , on a donc:

$$p^*(\Delta) \cong -\Delta \frac{g}{\|g\|} \text{ quand } \Delta \text{ est petit.}$$

Pour des valeurs intermédiaires de  $\Delta$ , la solution  $p^*(\Delta)$  suit une trajectoire courbe. La méthode de Dogleg trouve une solution approchée en remplaçant cette trajectoire courbe par un chemin constitué de deux segments. Le premier va de l'origine vers le minimiseur sans contrainte le long de la direction de la plus grande pente définie par:

$$p^U = -\frac{(g^T g)}{g^T B g}$$

Le second segment va de  $p^U$  vers  $p^B$ . Formellement, on écrit cette trajectoire par  $\tilde{p}(\tau)$  pour  $\tau \in [0, 2]$ :

$$\tilde{p}(\tau) = \begin{cases} \tau p^U & \text{si } 0 \leq \tau \leq 1 \\ p^U + (\tau - 1)(p^B - p^U) & \text{si } 1 \leq \tau \leq 2 \end{cases}$$

### Trust-Region-Modified-Dogleg Algorithm (TRMDA)

L'algorithme (TRDA) décrit précédemment est implémenté dans la fonction `Fsolve` de matlab. Néanmoins cette fonction prend en entrée un vecteur de fonction objectif à minimiser et, d'après les tests que nous avons effectués, la taille de ce vecteur affecte de façon directe le temps de calcul ainsi que la convergence de l'optimisation. Nous avons donc décidé de chercher la taille optimale pour que la procédure soit la plus performante. Nos tests ont abouti à une taille de vecteur de 50. L'algorithme que nous avons donc implémenté divise le vecteur initial en petits vecteurs de taille 50 et appelle la fonction `Fsolve` sur chacun d'eux. Ces optimisations sont mutuellement indépendantes, nous les effectuerons parallèlement à l'aide de la boucle `For` optimisée de Matlab (*Parfor*).

### Avantages

- La simplicité de la mise en œuvre.
- L'assurance d'obtenir une série décroissante de valeurs.
- Bonnes performances lorsque la dimension  $N$  augmente.
- Convergence quadratique.
- Convergence non affectée dans le cas d'une solution au voisinage de la frontière.
- Coût (temps de calcul) réduit.

### Inconvénients

- S'applique mal (ou difficilement) lorsque le domaine de définition de la fonction est complexe ou que le minimum recherché se situe dans un voisinage de la frontière.
- Evaluations des dérivées de la fonction objectif.

## 1.2 Méthodologie

Au lieu de calculer l'erreur sur la fonction  $Y$ , nous calculons la calculons sur la fonction  $X$ , avec :

$$X(z) = A + B * Y(z)$$

Cette opération a pour but de ne pas sous-estimer l'erreur de l'inversion de  $Y$ , car une erreur  $e$  sur  $Y$  est équivalente à une erreur  $B * e$  sur  $X$ . On procède alors de la façon suivante :

### procédure

- Simuler un échantillon  $X^0$  de taille  $n$  selon une loi  $\mathcal{G\&H}(A, B, g, h)$ . Nous prenons  $(A, B, g, h) = (e^7; 2 * e^7; 2; 0.4)$
- Calculer  $Y^0(z) = \frac{X^0(z) - A}{B}$
- Trouver  $z^i$  par la méthode  $M_i$
- Calculer  $X^i(z) = A + B * Y(z^i)$
- Calculer l'erreur quadratique  $EQ_i = \sum_{k=1}^n (X_k^i - X_k^0)^2$
- Refaire les opérations  $N$  fois et calculer l'erreur quadratique moyenne  $EQM_i$  (la moyenne des  $EQ_i$ )

Dans le tableau ci-dessous nous rapportons les erreurs calculés ainsi que les temps de calcul (TC) en seconds, en prenant  $N = 10000$  et plusieurs valeurs pour  $n$ .

$n$	50		100		1000		5000		10000	
	EQM	TC	EQM	TC	EQM	TC	EQM	TC	EQM	TC
<b>TRDA</b>	2,12E-22	0,09	3,42E-21	0,20	1,4E-17	3,45	4,8E+14	74,21	5,9E+17	$\infty$
<b>TRMDA</b>	2,12E-22	0,10	3,43E-21	0,18	1,4E-17	1,27	2,6E-18	5,65	1,4E-14	14,14
<b>NMA</b>	6,30E-06	0,13	3,13E-05	0,26	2,8E-04	2,59	2,0E-03	13,25	3,7-E03	26,74
<b>GSA</b>	5,14E-09	0,05	3,26E-05	0,10	1,1E-02	0,99	1,1E+00	5,11	4,4E+01	10,20

On effectue maintenant la même méthodologie pour  $h < 0$  ( $h = -0,4$ )

$n$	50		100		1000		5000		10000	
	EQM	TC	EQM	TC	EQM	TC	EQM	TC	EQM	TC
<b>TRDA</b>	4,0E-20	0,07	3,0E-22	0,14	3,1E-21	4,12	5,0E+02	52,91	$\infty$	$\infty$
<b>TRMDA</b>	4,0E-20	0,07	2,5E-21	0,18	1,1E-13	1,29	2,0E-12	8,5	1,4E-10	14,18
<b>NMA</b>	2,7E-05	0,13	2,4E-05	0,26	3,1E-04	2,54	1,4E-03	13,00	3,7-E03	27,05
<b>GSA</b>	4,0E+09	0,07	6,4E+09	0,14	5,6E+10	1,38	2,5E+11	7,28	$\infty$	12



A l'issue de ces tests on retient l'instabilité du *GSA* lorsque  $h$  est négatif et la divergence de *TRDA* lorsque  $n$  devient grand, on choisit alors d'utiliser *TRMDA* qui est stable et qui donne les meilleurs résultats que ça soit en terme de temps de calcul ou en précision.

## 2. Loi Normal-Inverse-Gaussian (NIG)

Dans cette partie nous explorons brièvement une loi généralisée autre que la distribution  $\mathcal{G}\&\mathcal{H}$ , la loi de Wald. Cette loi est connue par son utilisation répandue en modélisation de la sévérité des risques opérationnels.

### 2.1 Loi Inverse-Gaussian (IG)

#### 2.1.1 Définition

Nous commençons dans un premier temps par définir la loi inverse-gaussienne ou loi de Wald, une loi à deux paramètres dont la NIG en est une transformation. En théorie des probabilités et en statistique, la loi inverse-gaussienne est une loi de probabilité continue à deux paramètres dont le support est  $[0, \infty[$ . Le terme « inverse » ne doit pas être mal interprété, la loi est inverse dans le sens suivant : la valeur du mouvement brownien à un temps fixé est de loi normale, à l'inverse, le temps en lequel le mouvement brownien avec une dérive positive (drifté) atteint une valeur fixée est de loi inverse-gaussienne.

Sa densité de probabilité est donnée par :

$$f(x, \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp \frac{-\lambda(x - \mu)^2}{2\mu^2 x} \mathbb{I}_{[0, \infty)}(x)$$

Où  $\mu > 0$  est son espérance et  $\lambda > 0$  est un paramètre de forme.

Lorsque  $\lambda$  tend vers l'infini, la loi inverse-gaussienne se comporte comme une loi normale, elle possède plusieurs propriétés similaires avec cette dernière.

La fonction génératrice des cumulants (logarithme de la fonction caractéristique) de la loi inverse-gaussienne est l'inverse de celle de la loi normale.

Pour indiquer qu'une variable aléatoire  $X$  est de loi inverse-gaussienne de paramètres  $\mu$  et  $\lambda$ , on utilise la notation  $X \sim IG(\mu, \lambda)$

#### 2.1.2 Maximum de vraisemblance

Considérons le modèle donné par

$$X_i \sim IG(\mu, \lambda), \quad i = 1, 2, \dots, n$$

La fonction de vraisemblance s'écrit :

$$\mathcal{L}(x; \mu, \lambda) = \left(\frac{\lambda}{2\pi}\right)^{\frac{n}{2}} \sqrt{\prod_{i=1}^n \frac{1}{x_i^3}} \exp \left( n * \frac{\lambda}{\mu} - \frac{\lambda}{2\mu^2} \sum_{i=1}^n x_i - \frac{\lambda}{2} \sum_{i=1}^n \frac{1}{x_i} \right)$$

En résolvant l'équation de vraisemblance, on obtient les estimateurs suivant :

$$\hat{\mu} = \bar{X}_n, \hat{\lambda} = \left( \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{x_i} - \frac{1}{\hat{\mu}} \right) \right)^{-1}$$

Nous remarquons que :

$$\hat{\mu} \sim IG(\mu, n\lambda), \quad \frac{n}{\hat{\lambda}} \sim \frac{1}{\lambda} \chi_{n-1}^2$$

### 2.1.3 Intégration du seuil de collecte

On considère maintenant la vraisemblance conditionnelle afin de prendre en compte le seuil de collecte  $H$  :

$$\mathcal{L}_{X|H}(x|\mu, \lambda) = \frac{\left(\frac{\lambda}{2\pi}\right)^{\frac{n}{2}} \sqrt{\prod_{i=1}^n \frac{1}{x_i^3}} \exp\left(n * \frac{\lambda}{\mu} - \frac{\lambda}{2\mu^2} \sum_{i=1}^n x_i - \frac{\lambda}{2} \sum_{i=1}^n \frac{1}{x_i}\right)}{1 - F(H; \mu, \lambda)}$$

On remplace la valeur de la fonction de répartition par sa valeur :

$$\mathcal{L}_{X|H}(x|\mu, \lambda) = \frac{\left(\frac{\lambda}{2\pi}\right)^{\frac{n}{2}} \sqrt{\prod_{i=1}^n \frac{1}{x_i^3}} \exp\left(n * \frac{\lambda}{\mu} - \frac{\lambda}{2\mu^2} \sum_{i=1}^n x_i - \frac{\lambda}{2} \sum_{i=1}^n \frac{1}{x_i}\right)}{1 - \Phi\left(\sqrt{\frac{\lambda}{H}}\left(\frac{H}{\mu} - 1\right)\right) - \exp\left(\frac{2\lambda}{H}\right) \Phi\left(-\sqrt{\frac{\lambda}{H}}\left(\frac{H}{\mu} + 1\right)\right)}$$

### 2.1.4 Convergence de la méthode du maximum de vraisemblance

	Paramètres connus		Estimation sans seuil		Estimation avec un seuil de 1000			
	$\log(\mu)$	$\log(\lambda)$	$\log(\hat{\mu})$	$\log(\hat{\lambda})$	$\log(\hat{\mu}) H$	$\log(\hat{\lambda}) H$	$F_{\mu,\lambda}(H)$	$F_{\hat{\mu},\hat{\lambda}}(H)$
<b>M1</b>	10	5	10,00	4,99	10,08	5,21	0,70	1
<b>M2</b>	11	6	11,00	6,00	11,00	6,03	0,52	0
<b>M3</b>	12	7	12,02	7,00	12,03	7,02	0,30	0
<b>M4</b>	12,5	7	12,53	7,50	12,53	7,51	0,18	0
<b>M5</b>	13	8	13,00	8,00	12,98	7,99	0,08	0

## 2.2 Loi NIG

### 2.2.1 Définition

Si  $Y$  suit une loi  $IG(\delta, \sqrt{\alpha^2 - \beta^2})$  et  $X|Y$  suit une loi  $N(\mu + \beta Y, Y)$  alors  $X$  est de loi  $NIG(\alpha, \beta, \mu, \delta)$  de densité :

$$f_{NIG}(x; \alpha, \beta, \mu, \delta) = \frac{\alpha}{\pi} \exp\left(\delta\sqrt{\alpha^2 - \beta^2} - \beta\mu\right) \frac{K_1\left(\alpha\delta\sqrt{1 + \left(\frac{x-\mu}{\delta}\right)^2}\right)}{\sqrt{1 + \left(\frac{x-\mu}{\delta}\right)^2}} \exp(\beta x)$$

où  $\mu \in \mathbb{R}, \alpha > 0, \delta > 0, \mu \in \mathbb{R}, 0 < |\beta| < \alpha$  et  $K_1$  la fonction de Bessel modifiée de troisième espèce avec l'index 1. Le cas limite  $\alpha \rightarrow \infty$  correspond à la loi normale.

La distribution *NIG* a deux propriétés intéressantes :

-Propriété de l'échelle :

$$X \sim NIG(\alpha, \beta, \mu, \delta) \Leftrightarrow cX \sim NIG\left(\frac{\alpha}{c}, \frac{\beta}{c}, c\mu, c\delta\right)$$

-Propriété de convolution :

$$NIG(\alpha, \beta, \mu_1, \delta_1) * NIG(\alpha, \beta, \mu_2, \delta_2) \Leftrightarrow NIG(\alpha, \beta, \mu_1 + \mu_2, \delta_1 + \delta_2)$$

### 2.2.2 Méthode des moments

Les estimateurs par la méthode des moments ont été calculés dans Eriksson, Forsberg, et Ghysels [17]:

Soit  $X \sim NIG(\alpha, \beta, \mu, \delta)$ , notant  $\bar{X}_n, \bar{V}_n^2, \bar{S}_n, \bar{K}_n$  la moyenne, la variance, l'asymétrie et le Kurtosis de l'échantillon, on a alors :

$$\hat{\alpha} = 3 \sqrt{\rho} (\rho - 1)^{-1} \bar{V}_n |\bar{S}_n^{-1}|$$

$$\hat{\beta} = 3(\rho - 1)^{-1} \bar{V}_n \bar{S}_n^{-1}$$

$$\hat{\mu} = \bar{X}_n - 3\rho^{-1} \bar{V}_n \bar{S}_n^{-1}$$

$$\hat{\delta} = 3 \rho^{-1} (\rho - 1)^{1/2} \bar{V}_n |\bar{S}_n^{-1}|$$

### 3. Optimisation par recuit simulé

Contrairement aux algorithmes d'optimisation utilisés précédemment, l'algorithme du recuit simulé permet la recherche d'un minimum global. C'est son principal avantage et le prix à payer est un temps de calcul beaucoup plus grand que celui des méthodes itératives (non-probabilistes).

#### 3.1 Principe

Cette méthode provient d'un principe de la thermodynamique utilisé en métallurgie pour améliorer la qualité d'un solide, elle s'inspire de l'évolution de ce dernier vers une position d'équilibre lors de son refroidissement.

On suppose qu'on a un système physique à la température  $T$ . On fait l'hypothèse que  $S$  peut avoir un nombre dénombrable d'états  $i \in \mathbb{N}$ . A chaque état  $i$  on associe un niveau d'énergie  $E_i$ . On note par  $X$  l'état du système. On a alors, la distribution qui caractérise l'équilibre thermique (Loi de Boltzmann) :

$$P_T(X = i) = \frac{1}{Z(T)} e^{-\frac{E_i}{k_B T}}$$

Où  $k_B$  est la constante de Boltzmann et  $Z$  est une fonction de normalisation donnée par :

$$Z(T) = \sum_{i \in \mathbb{N}} \exp\left(-\frac{E_i}{k_B T}\right)$$

Si  $i$  et  $j$  sont deux états. On pose  $\Delta E$  leur différence d'énergie. Il vient alors

$$\frac{P_T(X = i)}{P_T(X = j)} = e^{-\frac{\Delta E}{k_B T}}$$

Nous en déduisons que si  $\Delta E > 0$ , l'état  $j$  est plus probable que le l'état  $i$  et inversement. Néanmoins le rapport de probabilités dépend aussi de  $k_B T$ , et si ce terme est grand devant  $\Delta E$ , les états  $i$  et  $j$  sont presque équiprobables.

Nous faisons à présent l'analogie suivante entre système physique et problème de minimisation :

- Etat physique  $\leftrightarrow$  solutions admissibles
- Energie du système  $\leftrightarrow$  coût d'une solution
- $k_B T \leftrightarrow$  paramètre de contrôle noté  $T$

C'est sur cette analogie que nous allons introduire une implémentation de la méthode du recuit simulé, l'algorithme de Métropolis. L'idée est d'effectuer un mouvement selon une distribution de probabilité qui dépend de la qualité des différents voisins :

- Les meilleurs voisins ont une probabilité plus élevée.
- Les moins bons ont une probabilité plus faible.

Le paramètre  $T$  (température) varie au cours de la recherche :  $T$  est élevée au début, puis diminue et finit par tendre vers 0.

### 3.2 L'algorithme de Métropolis

Si  $(S, f)$  est un problème d'optimisation, et  $i, j$  deux solutions, nous introduisons le critère d'acceptation de Métropolis :

$$P_T(\text{accepter } j) = \begin{cases} 1 & \text{si } \Delta f \geq 0 \\ e^{-\frac{\Delta f}{T}} & \text{sinon} \end{cases}$$

Où  $T$  est notre paramètre de contrôle et  $\Delta f = f(j) - f(i)$ .

L'algorithme est défini à partir d'un état initial  $i = i_0$  donné, et de la répétition de deux étapes. De plus on va faire baisser la température à fréquence régulière.

La première étape est appelée déplacement. Il s'agit de générer une solution admissible  $j$  à partir de  $i$ . Dans une seconde étape, on met en jeu le critère d'acceptation afin de décider si  $j$  sera retenue ou non.

Plusieurs règles de déplacement existent, nous en donnons deux exemples connues :

-**La règle de Cerny** : on construit un repère  $\{e_1, \dots, e_n\} \in \mathbb{R}^n$ , avec  $e_1$  pointant dans la direction du meilleur résultat obtenu. On choisit une direction  $d$  aléatoirement parmi  $e_1, \dots, e_n$  avec une probabilité plus grande pour  $e_1$ , et on se déplace aléatoirement suivant  $d$ .

-**La règle uniforme** : on tire aléatoirement un vecteur  $u$  dont chacune des composantes suit une loi uniforme sur  $[-1, 1]$ . On se déplace de  $\delta j = q \cdot u$ , où  $q$  est un pas fixé.

En ce qui concerne le refroidissement, on peut considérer par exemple l'une des règles suivantes :

-Réduire  $T$  à  $(1 - \varepsilon)T$  tous les  $m$  déplacements,  $\varepsilon$  et  $m$  choisis par expérience, jusqu'à  $T \simeq 0$ .

-On se fixe un nombre total  $K$  (très grand) de déplacement et on réduit  $T$  chaque  $\left(\frac{K}{N}\right)$  déplacements ( $N$  fois) en posant :  $T_k = T_0(1 - \frac{k}{N})^\alpha$  pour  $k = 1 \dots N$ , on choisit en général  $\alpha = 1, 2$  ou  $4$ .

#### Algorithme :

##### Début

Engendrer une configuration initiale  $x_0$ ;  $x = x_0$

$T := T_0$

##### Répéter

Nb\_Moves := 0

Pour  $i := 1$  à Nb\_Iter

Engendrer un voisin  $x'$  de  $x$

Calculer  $\Delta = f(x') - f(x)$

Si CritMetropolis( $\Delta, T$ ), alors

$x := x'$  ;

Nb\_Moves++ ;

Acceptance\_Rate := i / (Nb\_Moves)

$T := \text{DecreaseTemp}(T)$ ;

Jusqu'à <CritStop >

Retourner la meilleure configuration trouvée

##### Fin.

CritStop est un critère d'arrêt :

- Acceptance\_Rate devient très petit
- $f$  cesse d'évoluer.

L'inconvénient de cet algorithme est qu'il est très lent, donc pour l'optimisation d'une fonction à quatre variables il peut s'avérer inutilisable en pratique. De plus, le choix des paramètres doit être effectué à chaque appel à la fonction, ce qui rend difficile l'automatisation du processus d'estimation des paramètres. Néanmoins l'outil *simulannealbnd* de la Toolbox 'Global optimisation' de Matlab, est une version optimisée de cet algorithme, et son principale avantage est qu'il permet de choisir automatiquement les bonnes valeurs pour les paramètres de température, nombre d'itérations...etc.

### 3.3 Optimisation itérative- probabiliste

L'idée ici est d'implémenter un algorithme qui combine une méthode d'optimisation itérative à une autre probabiliste. L'intérêt de cette opération est d'augmenter la probabilité de trouver un minimum globale et ainsi permettre d'éviter les valeurs aberrantes lors des estimations de paramètres.

Pour les lois  $\mathcal{G}\&\mathcal{H}$  on a utilisé Nelder-Mead pour minimiser les différentes fonctions objectif, cet algorithme ne garantit pas que le minimum trouvé soit un minimum global.

Le résultat dépend donc des 'starting values' que l'on fixe pour pouvoir effectuer l'optimisation.

On va donc procéder de la façon suivante : sur un problème d'optimisation de la fonction *Func* sur  $\Omega$ .

#### Algorithme

```

-FuncValue = Inf
-Simuler N 'starting values' (Sv1, ..., SvN), selon une loi DistSV qui a  $\Omega$  comme support
- for i := 1 à N
    -x := Nelder_Mead(Func, Svi)
    -if Func(x) < FuncValue
        xbest := x; FuncValue := Func(x);
    Fin if
Fin for
-return xbest
Fin

```

Cet algorithme garantit un minimum global si *DistSV* couvre toutes les valeurs de  $\Omega$ , ce cas de figure est bien sûr impossible à atteindre en pratique car il nécessite un nombre infini d'appels à la fonction Nelder-Mead.

*DistSV* doit être choisie selon le comportement de la fonction *Func*. Pour le cas de l'estimation des paramètres de lois  $\mathcal{G}\&\mathcal{H}$  tronquées par la méthode QD on choisit *DistSV* selon une des règles suivantes :

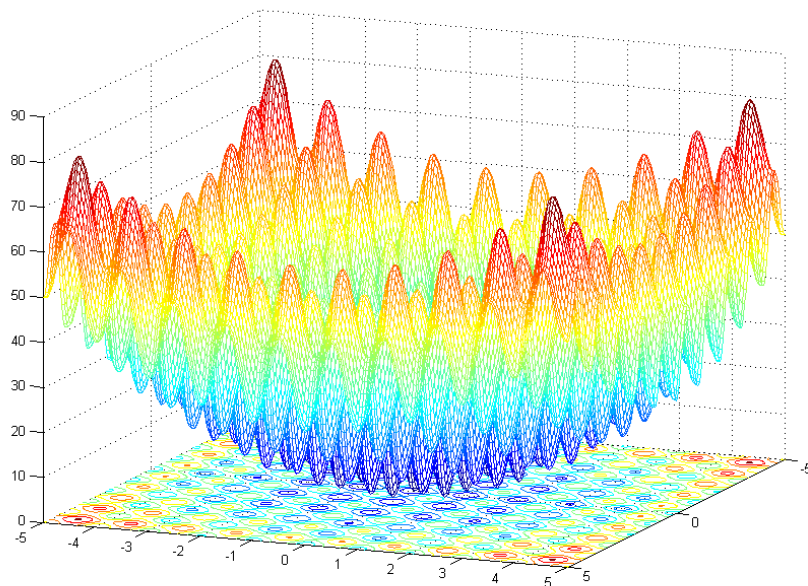
- une distribution uniforme de support [*Sv*<sub>0</sub>, *Sv*<sub>max</sub>].
- une répartition *Sv*<sub>0</sub>. *Exp*( $\lambda$ ) de support [*Sv*<sub>0</sub>, *Sv*<sub>max</sub>]. Le paramètre de la distribution sera choisi en fonction de la fonction à minimiser.

Où *Sv*<sub>0</sub> est le résultat d'une estimation par la méthode QD sans prendre en compte le seuil de collecte et *Sv*<sub>max</sub> le maximum de  $\Omega$ .

Application :

Nous testons cet algorithme sur la fonction de Rastrigin :

$$f(x_1, x_2) = 20 + x_1^2 + x_2^2 - 10 * (\cos(2\pi x_1) + \cos(2\pi x_2))$$



Cette fonction possède plusieurs minimums locaux et admet un minimum global à son origine. Nous lançons fminsearch (Nelder\_Mead) avec différentes 'starting values' :

Fonction appelée	Starting values		Résultats	
	$x_0$	$y_0$	$x$	$y$
testfmins	2	2	1.9899	1.9899
testfmins	1	1	0.9950	0.9950
testfmins	0,1	0,1	-0.000000001	-0.000000001

Le programme n'aboutit à un minimum global que si l'on lui fournit des starting values proches de la solution.

Avec l'algorithme décrit précédemment, on lui fournit seulement l'intervalle sur lequel il doit chercher la solution, et un nombre de simulation qui présente le nombre de starting values aléatoires à générer. Pour 50 simulations on obtient :

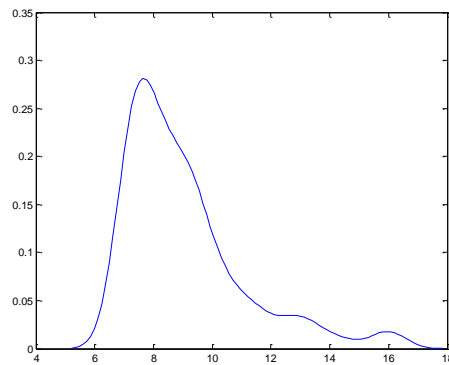
Fonction appelée	Intervalles de recherche		Résultats	
	Pour $x$	Pour $y$	$x$	$y$
testfmins_global(50)	$[-10,10]$	$[-10,10]$	-0.000000002	-0.000000001

## 4. Traitement des pertes aberrantes

### 4.1 Réduction de la variance

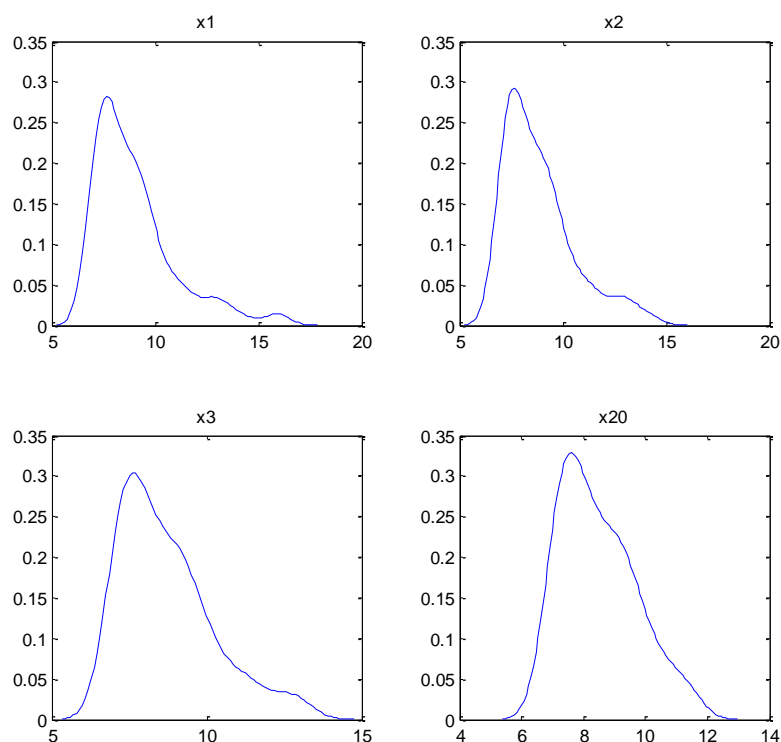
Dans ce qui précède nous remarquons que la catégorie EL4 de ENT3 donne une valeur de charge en capital de 24,8 milliards d'euros, cette valeur est bien évidemment aberrante et peut être expliquée par une distribution empirique de pertes très hétérogènes. Nous traçons, afin de mieux nous en convaincre, la densité de la distribution estimée par la méthode du noyau du log des pertes de cette catégorie.





On remarque la présence d'un deuxième pique d'amplitude conséquente en queue de distribution (autour de  $e^{16}$ ), c'est à cause de ceci que la loi log-normale peine à modéliser cette catégorie et que la loi  $\mathcal{G}\&\mathcal{H}$  y arrive un peu mieux en choisissant un paramètre  $g$  grand pour présenter la variance des données et un paramètre  $h$  important pour caractériser leur queue épaisse. Bien que cette modélisation soit meilleure que celle obtenue avec une loi log-normale en termes d'adéquation, elle ne reflète pas le niveau de risque réel de la banque et ne peut être utilisée en pratique. Il est donc indispensable de lisser les données de sévérité avant de les modéliser, les pertes ainsi écartées seront traitées en tant que scénarios.

Nous essayons dans un premier temps d'écarter les pertes aberrantes selon un critère simpliste celui de la réduction de la variance, et nous évaluerons l'impact de cette opération sur la valeur de la CaR. La méthode évalue les variances des échantillons obtenus en enlevant une perte différente à chaque itération, et choisie l'échantillon minimise la variance. Nous présentons ci-dessous les résultats de cette opération en enlevant 1, 5, 10 et 20 pertes de l'échantillon d'origine.

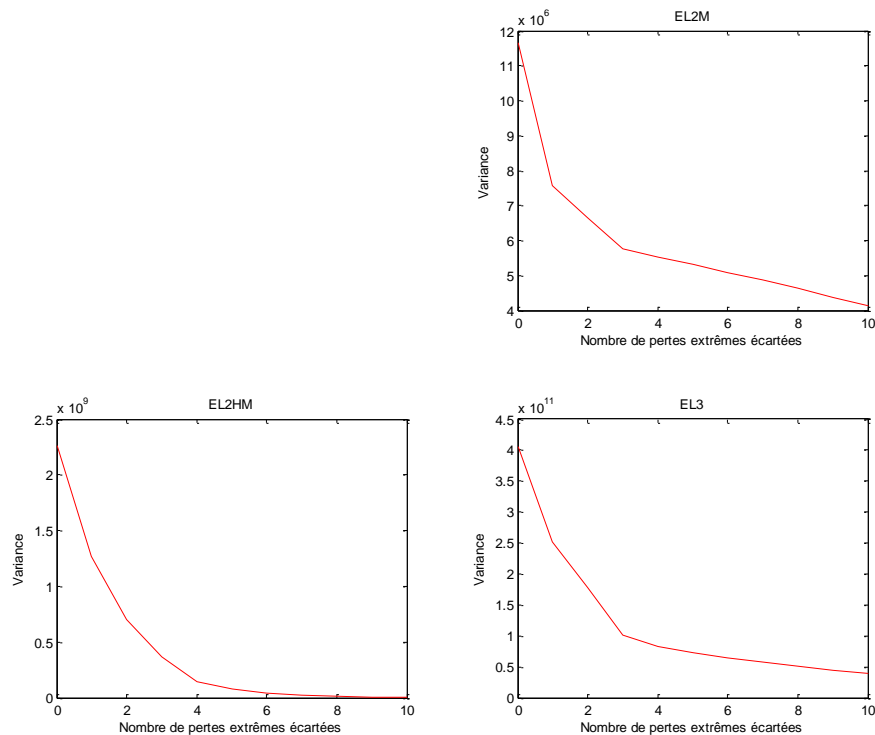


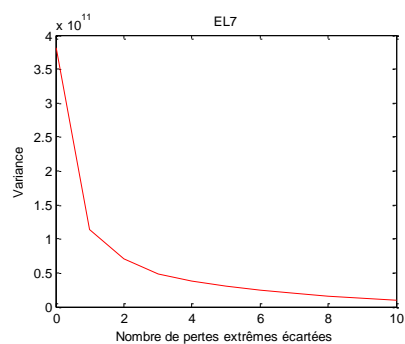
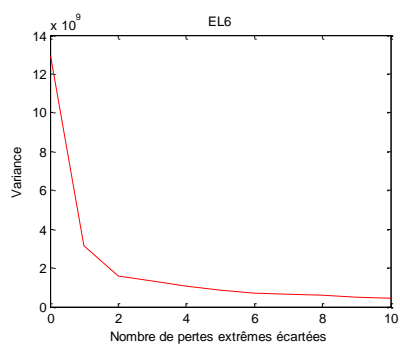
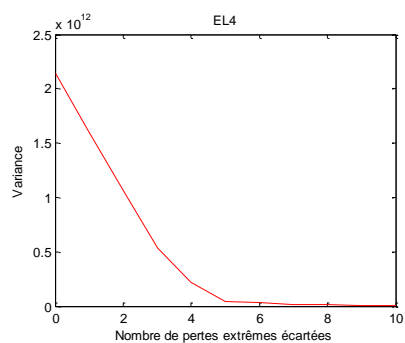
Nous calculons maintenant, après estimation des paramètres, les CaR correspondants à chaque cas.

	$X_1$	$X_5$	$X_{10}$	$X_{20}$
CaR	28 769 917 730 €	285 834 311 €	18 286 814 €	1 158 456 €

## 4.2 Choix du nombre de pertes à externaliser

On remarque que dans la plupart des cas les pertes enlevées sont des pertes extrêmes, afin de choisir les valeurs qui s'éloignent 'beaucoup' du profil de risque nous traçons l'évolution du critère (la variance dans un premier temps) de l'échantillon en fonction du nombre de pertes écartées pour ENT3.





On remarque que sur certaines catégories (EL6 et EL7) les deux plus grandes pertes (0.4% et 0.07% de l'échantillon) ne sont pas compatibles avec le reste des données, dans le sens où elles s'écartent relativement beaucoup de l'échantillon. Le nombre de ces valeurs est de cinq (3% de l'échantillon) pour la catégorie EL4