

**Mémoire présenté pour l'obtention du  
Diplôme de Statisticien mention Actuariat de l'ISUP  
et l'admission à l'Institut des Actuaire**

Le 1<sup>er</sup> octobre 2014

Par : Vincent Leherissé

Sujet : **Modèles robustes pour la sévérité des risques opérationnels**

Confidentialité : oui. Durée : 1 an.

*Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus.*

**Membres présents du jury de  
l'Institut des Actuaire**

**Entreprise**

Crédit Agricole S.A.

**Directeur de mémoire en entreprise**

Sophie Lavaud

**Membres présents du jury de  
l'ISUP**

**Autorisation de publication et de mise  
en ligne sur un site de diffusion de do-  
cuments actuariels**

*(après expiration du délai de confidentialité)*

Signature du responsable entreprise

Signature du candidat

# Sommaire

---

<b>Sommaire</b>	<b>iii</b>
<b>Résumé</b>	<b>iv</b>
<b>Introduction</b>	<b>1</b>
<b>I Contexte de l'étude</b>	<b>2</b>
<b>1 Risques bancaires et réglementation</b>	<b>3</b>
1.1 Risques bancaires	3
1.1.1 Panorama des risques	3
1.1.2 Fondements de la gestion des risques	3
1.1.3 Fonds propres d'une banque	5
1.2 La réglementation prudentielle	6
1.2.1 Évolution des réglementations	6
1.2.2 La norme Bâle II	7
1.2.3 Accentuation avec la norme Bâle III	8
1.3 Parallèle avec Solvency II	10
<b>2 Risque opérationnel</b>	<b>12</b>
2.1 Classification des risques opérationnels	12
2.2 Dispositif réglementaire	13
2.3 Approche par distributions de pertes : méthode LDA	14
2.3.1 Estimation des paramètres	15
2.3.2 Calculs de charges en capital	16
<b>3 Modélisation de la sévérité des pertes</b>	<b>19</b>
3.1 Contraintes spécifiques	19
3.2 Lois de probabilité	20
3.3 Méthodes d'estimation standards	20
3.4 Tests d'adéquation	22
3.5 Objectifs du mémoire	24
<b>II Modèles robustes</b>	<b>25</b>
<b>4 Statistique robuste</b>	<b>26</b>
4.1 Fonction d'influence et B-robustesse	26
4.1.1 Fonction d'influence	27
4.1.2 B-robustesse	28
4.1.3 Robustesse qualitative	30
4.2 M-estimateurs et maximum de vraisemblance	30
4.2.1 M-estimateurs	30
4.2.2 Fonction d'influence de l'estimateur par maximum de vraisemblance	31

<b>5</b>	<b>Étude d'estimateurs robustes</b>	<b>37</b>
5.1	Estimateur OBRE	39
5.1.1	Définition	39
5.1.2	Algorithme	40
5.1.3	Calculs pour une loi log-normale tronquée	41
5.2	Estimateur de Cramér-von-Mises	42
5.2.1	Prise en compte normale du seuil	43
5.2.2	Prise en compte des données sous le seuil	44
5.3	Estimateur Quantile-Distance	45
5.3.1	Spécification des paramètres	47
5.3.2	Version par optimisation directe	49
5.3.3	Version par optimisation scindée	49
<b>III</b>	<b>Application au calcul du capital réglementaire</b>	<b>53</b>
<b>6</b>	<b>Tests de robustesse</b>	<b>54</b>
6.1	Robustesse locale	54
6.2	Robustesse globale	58
<b>7</b>	<b>Calcul de capital sur données réelles</b>	<b>66</b>
7.1	Résultats généraux	66
7.2	Focus sur trois échantillons	67
	<b>Conclusions</b>	<b>71</b>
	<b>Références bibliographiques</b>	<b>73</b>
	<b>Annexes</b>	<b>75</b>
<b>A</b>	<b>Calculs de fonctions d'influence</b>	<b>75</b>
<b>B</b>	<b>Problématiques numériques liées à l'OBRE</b>	<b>84</b>
<b>C</b>	<b>Résultats de robustesse complémentaires</b>	<b>91</b>
<b>D</b>	<b>Résultats sur données réelles</b>	<b>102</b>

## Résumé

Dans le cadre de la réglementation Bâle II, les banques ont l'obligation de mobiliser une partie de leurs fonds propres à des fins de couverture de leurs risques opérationnels. Chaque établissement peut faire le choix d'utiliser l'approche avancée, conditionnée par la validation d'un modèle interne, la Loss distribution approach (LDA). La seule exigence est de construire le profil de risque de pertes sur une année et de déterminer le quantile à 99,9 % de cette distribution. La construction de ce profil de risque se fait en découpant de façon matricielle (lignes d'activités  $\times$  catégories de risques) les différentes sources potentielles de pertes opérationnelles. Au sein de chacune de ces cellules, l'estimation s'effectue sur deux dimensions, la fréquence et la sévérité, afin de pouvoir construire la distribution de perte annuelle.

C'est la modélisation de la dimension sévérité qui soulève le plus de questions, en raison des spécificités des pertes opérationnelles. Ce mémoire explore le recours à des méthodes d'estimation robustes, alors que les méthodes classiques le plus souvent utilisées (maximum de vraisemblance, méthode des moments généralisée) ne présentent pas cette caractéristique, et sont très sensibles à l'hétérogénéité des données.

En lien avec les exigences du Comité de Bâle, l'objectif de ce mémoire est donc d'étudier la notion d'estimation robuste. Différentes familles d'estimateurs (OBRE, Cramér-von-Mises, quantile-distance) sont ainsi présentées et leurs propriétés sont analysées dans le contexte du calcul de capital au titre du risque opérationnel.

**Mots-clés :** risque opérationnel, Bâle II, charge en capital, sévérité, statistique robuste, méthodes d'estimation.

## Abstract

In the Basel II framework, banks are required to raise a part of their capital to hedge their operational risks. Each institution may choose to use the advanced approach, conditioned by the validation of an internal model, the loss distribution approach (LDA). The only requirement is to build a risk profile over a year and determine the 99.9% quantile of this distribution. The construction of this risk profile is done by cutting different potential sources of operational losses into a matrix (business lines *times* risk categories). Within each of these cells, the estimation is performed in two dimensions, the frequency and severity, to build the distribution of annual loss.

Within the LDA framework, the modelling of the severity is the most difficult task because of the specificity of operational risk data. This thesis explores the use of robust estimation methods, while the most commonly used methods (maximum likelihood, generalized method of moments) did not show this effect, and are very sensitive to the heterogeneity of the data. In line with the requirements of the Basel Committee, the objective of this thesis is to study the concept of robust estimation. Different families of estimators (OBRE, Cramér-von-Mises, quantile-distance) are thus presented and their properties are analyzed in the context of computing capital charge for operational risk.

**Keywords :** operational risk, Basel II, capital charge, severity, robust statistics, estimation methods.

# Introduction

---

Le rôle clé d'intermédiation joué par les institutions bancaires dans le financement de l'économie est tel qu'il est impossible d'envisager une stabilité du système financier mondial sans un fonctionnement régulé des banques. Les diverses activités conduites par une banque l'amènent de fait à encourir une grande variété de risques. Or une appréhension erronée des risques encourus peut mener à des pertes financières considérables, qui peuvent elles-mêmes engendrer la faillite et ainsi entraver la stabilité de l'économie mondiale.

L'enjeu de la gestion des risques est donc bien de parvenir à prévenir l'occurrence de pertes financières importantes pouvant mettre en péril la pérennité d'un établissement. Si elle garantit un fonctionnement sain à court terme, une gestion efficace des risques concourt également, à plus long terme, à des gains substantiels pour la banque, liés à la stabilité de son activité.

C'est à cette fin qu'ont été mis en place les Accords de Bâle, contraignant les banques à instituer des mesures visant à quantifier de manière appropriée leur exposition au risque et à assurer son pilotage. Dans ce contexte, le risque opérationnel a commencé à être considéré comme un risque majeur. La grande difficulté inhérente à la gestion des risques opérationnels est la capacité à anticiper des événements rares ayant des conséquences extrêmement lourdes pour les banques.

Et c'est dans l'objectif de couvrir ces pertes à caractère imprévisible que les banques ont développé de nombreuses méthodologies et modèles statistiques. Le modèle le plus couramment utilisé se fonde sur l'ajustement des distributions de pertes (*Loss Distribution Approach*). Le capital réglementaire correspond alors à une Value-at-Risk à 99,9% de la perte annuelle (couverture d'un risque millénaire), calculée à partir de deux distributions : la fréquence des pertes (nombre annuel des pertes) et la sévérité (montant des pertes).

La modélisation de la sévérité de ces risques divers est la partie du modèle la plus sensible et a fait l'objet de nombreuses publications proposant nombre d'options méthodologiques envisageables, en termes de distributions paramétriques ou de méthodes d'estimation de leurs paramètres. La problématique développée dans ce mémoire est d'aborder la question des méthodes d'estimation en se focalisant sur celles présentant un caractère robuste. Il s'agit à la fois d'une réponse directe aux textes réglementaires qui prônent l'usage de méthodes robustes – au sens large – et d'une ouverture à une branche (trop) peu utilisée, la statistique robuste.

En conséquence, ce mémoire s'articule autour de trois parties. La première est consacrée à la présentation du contexte de l'étude, sur les plans réglementaire et méthodologique. La deuxième partie présente les bases de la statistique robuste et différents estimateurs robustes. Enfin, la dernière partie est consacrée à l'application de ces méthodes dans le cadre du calcul de capital réglementaire.

## Partie I

---

# **Contexte de l'étude**

# 1 Risques bancaires et réglementation

---

## 1.1 Risques bancaires

### 1.1.1 Panorama des risques

Les risques bancaires sont liés à la survenance d'événements non prévisibles pouvant avoir des conséquences importantes sur le bilan de la banque. Par exemple, une fraude peut conduire à une perte importante, affectant le résultat net de la banque.

Toutes les activités menées par la banque l'exposent potentiellement à un risque, il existe par conséquent un large spectre de risques. Néanmoins, on distingue généralement trois grands types de risque :

- le **risque de marché**, qui correspond au risque que la valeur des actifs détenus par la banque varie en raison de l'évolution des prix sur les marchés financiers ;
- le **risque de crédit**, qui est lié au risque qu'une contrepartie ne parvienne pas à remplir ses obligations vis-à-vis de la banque ;
- le **risque opérationnel**, qui résulte d'inadéquations ou de défaillances liées à des procédures, des agents, des systèmes internes ou des événements extérieurs.

Pour une banque universelle, on estime que le risque le plus important est le risque de crédit, suivi par le risque opérationnel, puis par le risque de marché.

Le tableau suivant présente quelques exemples liés aux trois grands risques :

Risque de crédit	Risque de marché	Risque opérationnel
Défaillance	Change	Fraude
Dégradation de créance	Taux d'intérêt	Désastre
	Volatilité	Risque juridique

Il est souvent facile de distinguer ces trois risques, mais la séparation est parfois ténue. Certains risques sont à la frontière entre les risques de crédit, de marché et le risque opérationnel. C'est par exemple le cas du risque de modèle dans le cadre de la valorisation de produits exotiques, qui peut être considéré comme un risque de marché ou un risque opérationnel.

Par ailleurs, d'autres risques ne correspondent pas à des pertes avérées. Si la banque prend de mauvaises décisions qui peuvent conduire à la perte de parts de marché, les conséquences ne sont pas visibles immédiatement. On parle de risque stratégique.

### 1.1.2 Fondements de la gestion des risques

Pour faire face à la diversité des risques auxquels elles s'exposent, les banques ont progressivement déployé des moyens considérables. Ces évolutions (méthodologies, matériel, équipes, etc.) s'expliquent en grande partie par des facteurs historiques concernant la théorie financière, les marchés financiers ou encore les désastres financiers.

## La théorie financière

La naissance de la théorie financière est généralement associée aux travaux fondateurs de Louis Bachelier. Les années trente marquent le début des recherches empiriques sur les prix des actifs, et notamment la formation des prix, l'efficacité des marchés et l'anticipation des cours des actions. Ce n'est que dans les années cinquante que des chercheurs (Markowitz, Sharpe, ...) entreprennent les premiers travaux conséquents sur le risque, qui aboutissent à la théorie moderne du choix de portefeuille basée sur le modèle CAPM (*Capital Asset Pricing Model*).

L'année 1973 est marquée par la parution de l'article de Black et Scholes pour valoriser une option européenne, point de départ des recherches concernant la valorisation des produits dérivés. Ces travaux sont d'autant plus importants qu'ils abordent le problème de la valorisation du point de vue du risque, en considérant le prix de l'option comme le prix de la couverture mise en place pour insensibiliser le portefeuille. Quarante ans plus tard, ces résultats font toujours figure de référence pour la gestion des produits optionnels. Durant les années quatre-vingt-dix, la couverture des produits optionnels devient de plus en plus complexe. Si la méthodologie Raroc publiée en 1996 par Bank of America a trait au risque de crédit, c'est le risque de marché qui fait l'objet des plus grandes avancées avec la publication en 1994 de RiskMetrics par JP Morgan, qui permet une large diffusion des méthodes de valeur en risque. Voici les grandes dates de la théorie financière :

1900	Thèse de Louis Bachelier « Théorie de la spéculation ».
1952	Parution de l'article « Portfolio selection » (Markowitz).
1964	Modèle CAPM (Sharpe).
1970	Synthèse des travaux sur l'efficacité des marchés par Fama.
1973	Formule de valorisation d'une option européenne (Black et Scholes).
1974	Étude de l'obligation risquée (Merton).
1977	Modèles de taux de Vasicek et de Cox, Ingersoll et Ross.
1994	Publication de RiskMetrics par JP Morgan.
1996	Publication de la méthodologie Raroc par Bank of America.

## Le développement des marchés financiers

On peut considérer que les activités premières d'une banque l'exposent principalement au risque de défaillance d'emprunteurs à ne pas rembourser leurs dettes et au risque de volatilité des prix des actifs, c'est-à-dire au risque de crédit et au risque de marché. À partir des années quatre-vingt, la globalisation des marchés financiers entraîne une concurrence accrue entre les banques, associée à un besoin de financement plus important de la part des entreprises. C'est pourquoi le risque de crédit revêt un fort enjeu, d'autant qu'il touche tous les types de contreparties, y compris les États.

La succession de crises mettant en difficulté le système financier international amène les banques et les autorités à réfléchir pour mieux gérer le risque de crédit. Parallèlement, il apparaît que le risque de marché doit aussi être mieux appréhendé au vu des positions de hors-bilan et de la complexité des produits traités. L'innovation financière connaît en effet un développement accéléré, d'abord sur le marché des changes (*FX market*), en raison de la crise pétrolière de 1973, puis sur les marchés d'actions (*equity market*) et de taux d'intérêt et, dans une moindre mesure, sur le marché des matières premières (*commodities*).

## Les crises financières

La stabilité du système financier international est une condition importante pour la mise en place des politiques économiques, notamment pour garantir la croissance. La régulation du système financier international et la mise en place de mécanismes de surveillance des établissements financiers par des autorités de supervision sont donc une logique économique afin d'éviter les crises systémiques.

La mise en place d'une réglementation vise donc dans un premier temps à limiter le risque systémique, et dans un deuxième temps à éviter les défaillances individuelles des établissements financiers, dans la



mesure où la défaillance d'un seul établissement peut engendrer une contagion aux autres établissements. Voici quelques points de repère sur les désastres financiers :

1974	Herstatt Bank : 620 millions de dollars (change)
1995	Barings : 1,33 milliard de dollars (stock index futures)
1997	Natwest : 127 millions de dollars (swaptions)
1998	LTCM : 2 milliards de dollars (crise de liquidité)
2006	Amaranth : 6,5 milliards de dollars (contrats à terme sur gaz)
2007	Société Générale : 7,1 milliards de dollars (fraude)
2008	Madoff : 65 milliards de dollars (fraude)

Au-delà de ces désastres émanant d'acteurs relativement isolés, la crise financière de 2007-2008 a eu un impact considérable sur l'économie. À l'origine de cette crise, des prêts immobiliers avec des taux d'emprunt variables aux modalités complexes, sont accordés à des ménages "subprimes". Ces prêts servent en fait de sous-jacents à des produits structurés bénéficiant de très bons ratings car mêlés à des obligations d'États, et placés dans des hedge funds. En période de hausse de taux, les ménages ne parviennent plus à rembourser, et les actifs titrisés s'en retrouvent très mal valorisés. Les investisseurs se détournent alors de ces actifs à risque pour des instruments financiers plus transparents, ce qui entraîne l'illiquidité de nombreux fonds, déclenchant ainsi une crise de liquidité globale. L'interdépendance du système bancaire engendre par suite une crise de liquidité sans précédent. L'intervention des Banques Centrales sur le refinancement et le soutien massif des États permet néanmoins d'éviter de nombreuses faillites.

### 1.1.3 Fonds propres d'une banque

Les risques encourus par une banque impactent son bilan à travers ses fonds propres. Le bilan de la banque peut en effet se présenter, de façon simplifiée, sous la forme suivante :

Actif	Passif
Actifs immobilisés	<i>Fonds propres</i>
Crédits et prêts	Dettes
Titres	Dépôts
Trésorerie	

Les fonds propres sont donc un des éléments du passif. Ils regroupent notamment les actions ordinaires, les certificats d'investissement, les réserves et le résultat non distribué. Les autres éléments du passif d'une banque sont les dépôts, l'épargne des ménages, ainsi que les dettes. À l'actif, on trouve les crédits et les prêts aux ménages et aux entreprises, les services, le portefeuille de titres (portefeuilles de négociation et d'investissement).

Trois caractéristiques sont attribuables aux fonds propres :

- (1) Les fonds propres sont le moteur de l'activité de la banque : en raison de contraintes externes (principalement la réglementation) et internes (imposées par exemple par les actionnaires), ils déterminent le risque pris par la banque, et donc son activité. À ce titre, ils dimensionnent également la croissance de l'établissement, en fonction de l'évolution de son capital.
- (2) Les fonds propres sont une garantie vis-à-vis des créanciers dans la mesure où ils permettent d'absorber d'éventuelles fortes pertes dues à des éléments exogènes et/ou inattendus. Ainsi, plus leur niveau est élevé, plus la banque présente des gages de solidité. Ces fonds propres sont d'ailleurs un des éléments utilisés pour la notation de la banque (*rating*), note qui impacte le coût des ressources (de trésorerie et de long terme).
- (3) Les fonds propres permettent de couvrir les risques, et sont par conséquent rémunérés, par le ROE (*Return on Equity*). L'objectif de la banque est donc d'offrir le ROE le plus élevé à ses actionnaires en minimisant le risque encouru pour maximiser le profit. Celui-ci a une influence très importante sur la valeur de marché de la banque et conditionne la croissance externe de l'établissement.

## 1.2 La réglementation prudentielle

### 1.2.1 Évolution des réglementations

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). Même si celui-ci n'a aucun pouvoir décisionnel, ses recommandations sont reprises par les autorités de tutelle des différents pays industrialisés. Le Comité de Bâle est un organisme de réflexion et de proposition sur la supervision bancaire, situé à la Banque des Règlements Internationaux (*Bank of International Settlements*, BIS).

La mise en place d'une réglementation est un processus assez long dont les différentes étapes sont les suivantes :

- publication d'un document de travail par le Comité de Bâle ;
- discussions entre le Comité de Bâle et l'industrie bancaire ;
- rédaction d'un « Accord » définitif par le Comité de Bâle ;
- traduction juridique de cet accord au niveau national.

En Europe, c'est la Commission Européenne qui est chargée de définir la CRD (*Capital Requirement Directive*), dont la mise en œuvre revient aux différentes autorités nationales de tutelle. En France, c'est l'Autorité de Contrôle Prudentiel et de Résolution<sup>1</sup> (ACPR, dont le Secrétariat Général est rattaché à la Banque de France) qui contrôle l'application de la CRD, une fois que celle-ci a été adaptée à la législation française par le Comité de la Réglementation Bancaire et Financière (CRBF). Le tableau suivant répertorie les grandes étapes de la réglementation prudentielle :

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 réglementation Bâle III.
2013	Début de mise en œuvre de la norme Bâle III.

### 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  $FP_1$  (*Tier one*), qui correspondent au capital et aux réserves ;
- (2) les fonds propres complémentaires  $FP_2$  (*Tier two*), principalement constitués d'emprunts subordonnés ;
- (3) les fonds propres surcomplémentaires  $FP_3$  (*Tier three*).

Selon l'Accord de Bâle, les établissements financiers doivent respecter les contraintes  $FP_2 \leq FP_1$  et  $FP_1/EPC \geq 4\%$ , mais surtout le ratio de solvabilité, appelé ratio Cooke :

$$\frac{FP_1 + FP_2}{EPC} \geq 8\%$$

1. L'ACPR résulte de la fusion intervenue en janvier 2010 des deux organes de contrôle des banques (la Commission Bancaire) et des assurances (l'Autorité de Contrôle des Assurances et des Mutuelles). Initialement dénommée, ACP, elle prend le nom d'ACPR en 2014.

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 fiables pour calculer les risques de marché et donc d'obtenir des exigences de fonds propres beaucoup 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.

### 1.2.2 La norme Bâle II

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 intègre désormais le risque opérationnel, et la définition d'un nouveau ratio de solvabilité, le ratio McDonough :

$$\frac{FP_1 + FP_2}{E_{RC} + E_{RO} + E_{RM}} \geq 8\%$$

où  $E_{RC}$ ,  $E_{RO}$  et  $E_{RM}$  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 rapprocher la réglementation des pratiques en vigueur dans l'industrie pour le pilotage des risques, afin que l'exigence de fonds propres soit plus sensible au risque réel de la banque. L'idée est d'autoriser les banques, sous certaines conditions, à utiliser des modèles internes pour mesurer le risque de crédit et le risque opérationnel, de la même façon que pour le risque de marché.

Le nouveau dispositif s'articule autour de trois piliers :

- (1) le premier pilier correspond aux exigences minimales de fonds propres et vise à définir le ratio McDonough ;
- (2) le deuxième pilier concerne le processus de surveillance et donne un pouvoir accru aux autorités de contrôle (Autorité de Contrôle Prudentiel) ;
- (3) le troisième pilier porte sur la communication financière et la discipline de marché.

#### Pilier 1 : exigences minimales de fonds propres

En ce qui concerne le premier pilier, le traitement du risque de marché est inchangé. Le traitement du risque de crédit est revu en profondeur. Trois méthodes sont désormais possibles pour mesurer le risque de crédit :

- l'approche standard (SA, pour *Standard Approach*) reste une méthode forfaitaire, mais désormais les pondérations en risque dépendent de la notation des agences externes ;
- l'approche IRB (*Internal Ratings Based*) « simple » est basée sur des mesures internes des probabilités de défaillance (PD, pour *Probability of Default*) et des mesures exogènes des autres paramètres du modèle ;
- dans l'approche IRB « avancée », la banque estime d'autres paramètres comme le taux de perte (LGD, pour *Loss Given Default*) et l'exposition en cas de défaut (EAD, pour *Exposure at Default*).

Les méthodes de calcul sont standardisées et imposées par le Comité de Bâle. Que ce soit pour l'approche SA ou les approches IRB, le point central de ces approches est la notation, externe ou interne. Les approches IRB ne sont donc pas des modèles internes, mais correspondent plutôt à un modèle externe avec des paramètres internes. C'est la différence fondamentale avec l'approche précédente.

En outre, le risque opérationnel est intégré à l'assiette des risques. Trois méthodes sont proposées pour le mesurer :

- La méthode *Basic Indicator Approach* (BIA) : le risque opérationnel est appréhendé à partir du produit annuel brut et la charge en capital est alors une fonction linéaire de cet indicateur.
- La méthode *Standardized Approach* (TSA) : l'établissement est divisé en lignes métiers (*business lines*) et la méthode BIA est déclinée pour les différentes lignes métiers avec une pondération spécifique ; la charge en capital correspond alors à la somme des charges en capital spécifiques.
- Les méthodes *Advanced Measurement Approach* (AMA) : la banque peut alors choisir entre différents « modèles internes » pour quantifier son exigence de fonds propres.

## Pilier 2 : surveillance prudentielle

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.

## Pilier 3 : discipline de marché

Le troisième pilier concerne la discipline de marché. Il vise à améliorer la transparence financière des banques, en leur imposant de communiquer les informations nécessaires (structure du capital, de l'allocation de fonds propres, de l'exposition aux risques et des pertes) pour permettre à des tiers d'apprécier l'adéquation de leurs fonds propres. Une meilleure discipline de marché en est espérée.

### 1.2.3 Accentuation avec la norme Bâle III

Sous l'impulsion du G20, le Comité de Bâle a engagé une révision complète du dispositif prudentiel, les Accords de Bâle III, publiés en décembre 2010 et (en partie<sup>2</sup>) applicables depuis début 2013.

Cette réforme part du constat que la sévérité de la crise s'explique en grande partie par la croissance excessive des bilans et hors-bilan bancaires (via, par exemple, les produits dérivés), tandis que dans le même temps le niveau et la qualité des fonds propres destinés à couvrir les risques se dégradent. En outre, de nombreuses institutions ne disposent pas non plus de réserves suffisantes pour faire face à une crise de liquidité. Dans ce contexte, le système bancaire s'est révélé incapable d'absorber les pertes intervenues d'abord sur les produits structurés de titrisation et d'assumer ensuite la ré-intermédiation d'une partie des expositions de hors-bilan. Au pire de la crise, les incertitudes pesant sur la qualité des bilans, la solvabilité des banques et les risques liés à leur interdépendance ont provoqué une crise de liquidité généralisée.

---

2. La mise en œuvre intégrale du texte doit s'étaler jusqu'en 2019.

Les enjeux de cette réforme sont considérables, non seulement pour le secteur bancaire et la stabilité des marchés financiers, mais aussi pour l'économie dans son ensemble. Le nouvel accord se concentre sur quatre points d'attention majeurs.

## Renforcement de la qualité et de la quantité des fonds propres

**Définition plus restrictive des éléments de fonds propres.** Le premier objectif est d'accroître l'homogénéité, la qualité et la transparence des fonds propres, c'est-à-dire s'assurer qu'ils sont constitués de capitaux capables d'absorber des montants de pertes importants et non d'instruments financiers complexes et sujets à fluctuations. En particulier, les établissements devront constituer un capital *Common Equity Tier 1* (CET 1) ne pouvant plus contenir que des actions ordinaires (comme les parts sociales pour les groupes mutualistes) et des bénéfices mis en réserve. Parallèlement, les participations dans les entreprises financières (sociétés d'assurances par exemple) seront plus sévèrement déduits.

### Constitution de coussins de fonds propres complémentaires.

- *Coussin additionnel de conservation des fonds propres*  
Coussin de protection des fonds propres de 2,5% composé uniquement d'éléments CET 1, avec mise en réserve obligatoire de résultat lorsque le coussin baisse en-dessous du minimum requis.
- *Coussin contra-cyclique*  
Coussin additionnel destiné à limiter une progression excessive du crédit, compris entre 0% et 2,5% et composé uniquement d'éléments CET 1.
- *Coussin supplémentaire pour les institutions systémiques*  
Coussin supplémentaire de fonds propres en fonction du caractère plus ou moins systémique de la banque, les institutions dites systémiques (*systematically important banks*, SIB) étant celles qui, par leur taille ou leur activité, font peser un risque systémique sur l'économie

## Renforcement de la couverture des risques

- Renforcement des exigences associées au calcul de l'*Expected Positive Exposure* (EPE) au titre du risque de contrepartie sur instruments dérivés de gré à gré (*over-the-counter*, OTC), ainsi que la prise en compte des risques de corrélation défavorable<sup>3</sup> (*wrong-way risk*) ;
- Mise en place du *Credit Value Adjustment* (CVA), nouvelle charge en fonds propres destinée à couvrir les variations défavorables du prix de marché des risques de contrepartie sur toutes les opérations sur dérivés OTC.
- Prise en compte du risque lié aux institutions financières systémiques par une modification de la formule de calcul des RWA (coefficient de corrélation de la formule IRB multiplié par 1.25) pour les établissements de grande taille (bilan supérieur à 70 MEUR) et les établissements non régulés ;
- Renforcement du cadre prudentiel des activités de marché (Bâle 2.5, depuis 2009) :  
Mise en place d'une exigence additionnelle en fonds propres (IRC, *incremental risk charge*) visant à couvrir les risques de dégradation des dérivés de crédit du *trading book*, d'une exigence supplémentaire au titre de la VaR stressée (destinée à réduire la procyclicité), d'une *comprehensive risk measure* (CRM) vouée à couvrir le risque de corrélation entre les institutions financières, ainsi que l'augmentation des pondérations relatives aux titrisations.
- Instauration d'une charge en fonds propres pour les expositions sur les chambres de compensation centrales<sup>4</sup> : pondération des expositions de trading à 2%.

3. Risque que l'exposition à une contrepartie soit inversement corrélée à sa qualité de crédit, comme c'est le cas pour une position acheteuse de CDS sur une entité dont la qualité de crédit est positivement corrélée à celle de la contrepartie.

4. Une banque est exposée à une chambre de compensation de deux manières : au titre des transactions qui y transitent (appels de marge, collatéral déposé auprès de la chambre), et au titre du fonds commun (*default fund*), qui permet de mutualiser le coût du défaut d'un intervenant.

## Instauration de ratios de liquidité

Le Comité de Bâle propose la mise en place de deux ratios de liquidité :

- (1) Le LCR (*liquidity coverage ratio*) impose aux banques de détenir un stock d'actifs sans risque, qui leur permettrait de résister pendant trente jours à une crise qui se traduirait par des retraits massifs (en vigueur en 2015) :

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

- (2) Le NSFR (*net stable funding ratio*) vise le même objectif sur un an. Plus contraignant, il doit inciter les établissements de crédit à renforcer leur profil de financement en faveur de ressources de long terme, supposées plus stables (en vigueur en 2018) :

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

## Mise en place d'un ratio de levier

Le Comité envisage d'instituer un ratio de levier (*leverage ratio*) destiné à plafonner les expositions en risques indépendamment de la qualité des risques encourus, et ce afin de limiter l'endettement des banques. Précisément, les établissements devront vérifier, à partir de 2019 :

$$\frac{\text{FP}_1^*}{\text{Actifs non pondérés}} \geq 3\%.$$

## 1.3 Parallèle avec Solvency II

En termes de structures, les réglementations Bâle II et Solvency II sont très similaires. En effet, les deux cadres réglementaires s'organisent selon trois piliers dont les objectifs s'avèrent très comparables :

- (1) Pilier 1 : définit les modalités de l'exigence minimale en fonds propres.
- (2) Pilier 2 : fixe des exigences en matière de contrôle interne et d'exhaustivité de la couverture de la gestion des risques
- (3) Pilier 3 : donne les exigences en matière de communication financière.

Néanmoins, ces deux réglementations diffèrent très largement l'une de l'autre du fait des grandes différences d'activités entre les banques et les compagnies assurances et des objectifs du contrôle prudentiel inhérent. Le contrôle de l'industrie bancaire se concentre avant tout sur la stabilité d'un système singulièrement interconnecté et soumis à un risque systémique très important. À l'inverse, Solvency II a pour principal objectif de protéger les assurés contre le risque de faillite de leur compagnie d'assurance. En effet, les risques liés à l'activité des compagnies d'assurance sont de natures bien différentes.

Contrairement à l'industrie bancaire, ils portent davantage sur le passif du bilan et sur des risques en grande partie non-financiers comme les catastrophes climatiques qui sont souvent moins corrélés et spécifiques. En outre, l'inversion du cycle de production permet un financement sur le long terme et les compagnies d'assurances ne sont donc pas aussi exposées que les banques au risque de liquidité.

Ainsi, bien que leur structures soient comparables, Bâle II et Solvency II poursuivent des objectifs très différents. Néanmoins, ces deux réglementations n'évoluent pas indépendamment l'une de l'autre, dans la mesure où l'Union Européenne souhaite harmoniser le traitement réglementaire des types d'actifs communs aux domaines assurantiel et bancaire afin de limiter les possibilités d'arbitrages entre ces deux industries.

Concernant le premier pilier, Bâle II distingue principalement trois types de risques (risque de crédit, opérationnel et de marché) alors que Solvency II en considère six (risque non-vie, vie, santé, de marché,

de contrepartie et opérationnel). Alors que le capital réglementaire pour les assurances se fonde sur le capital économique nécessaire aux paiements des prestations aux assurés sous une certaine probabilité, le capital réglementaire bâlois est destiné à pouvoir absorber les pertes au sein de chacune des trois familles de risque définies précédemment.

Par ailleurs, Solvency II définit un niveau « souhaitable » de capital SCR (Solvency Capital Requirement) en-dessous duquel les compagnies d'assurance ne peuvent rester que pour une période de six à neuf mois et un niveau de capital minimum MCR (Minimum Requirement Capital) situé dans une fourchette comprise entre 25% et 45% du SCR et en-dessous duquel les autorités de contrôle peuvent retirer l'agrément.

Le prochain chapitre décrit le cadre et les approches réglementaires pour la mesure du risque opérationnel (bancaire).

## 2 Risque opérationnel

---

### 2.1 Classification des risques opérationnels

Même s'ils sont connus depuis longtemps, la gestion des risques opérationnels est un exercice encore récent pour les banques, qui a largement évolué grâce à la réforme Bâle II. Le Comité de Bâle définit les risques opérationnels « *comme le risque de pertes dues à une inadéquation ou à une défaillance des procédures, personnels, systèmes internes ou à des événements extérieurs* ».

Pour répondre à la diversité des événements assimilables à des risques opérationnels, le Comité de Bâle présente une classification des différentes catégories d'événement (*event line*, EL) :

- (EL 1) **Fraude interne** : pertes liées à des actes commis à l'intérieur de l'entreprise visant à commettre une fraude ou un détournement d'actif ou à enfreindre une disposition législative ou réglementaire, ou des règles de l'entreprise, à l'exclusion des cas de pratiques discriminatoires ou contraires aux règles en matière d'égalité professionnelle, et impliquant au moins un membre de l'entreprise.
- (EL 2) **Fraude externe** : pertes liées à des actes de tiers visant à commettre une fraude ou un détournement d'actif ou à enfreindre une disposition législative ou réglementaire.
- (EL 3) **Pratiques en matière d'emploi et sécurité du travail** : pertes liées à des actes contraires aux dispositions législatives ou réglementaires, ou aux conventions en matière d'emploi, de santé ou de sécurité, à la réparation de préjudices personnels ou à des pratiques discriminatoires ou contraires aux règles en matière d'égalité professionnelle.
- (EL 4) **Clients, produits et pratiques commerciales** : pertes liées à un manquement, délibéré ou non, à une obligation professionnelle envers un client (y compris les exigences en matière de confiance et d'adéquation du service), à la nature ou aux caractéristiques d'un produit.
- (EL 5) **Dommages occasionnés aux actifs physiques** : pertes liées à la perte ou à l'endommagement d'actifs physiques résultant d'une catastrophe naturelle ou d'autres événements.
- (EL 6) **Interruption de l'activité et dysfonctionnements des systèmes** : pertes liées à une interruption de l'activité ou au dysfonctionnement d'un système.
- (EL 7) **Exécution, livraison et gestion des processus** : pertes liées aux lacunes du traitement des transactions ou de la gestion des processus et aux relations avec les contreparties commerciales et les fournisseurs.

Ces différentes catégories d'événements sont observées suivant une répartition des lignes métier (*business lines*, BL) de la banque, proposée par le Comité de Bâle :

- (BL 1) Financement d'entreprise (*corporate finance*) : placement d'instruments financiers avec engagement ferme, services financiers aux entreprises, conseil et services en matière de fusions et de rachat d'entreprises, recherche en investissements et analyse financière, financement des collectivités locales et de l'administration, etc.
- (BL 2) Activités de marché (*trading & sales*) : négociation pour compte propre, intermédiation sur les marchés interbancaires, réception et transmission d'ordres sur instruments financiers, etc.
- (BL 3) Banque de détail (*retail banking*) : réception de dépôts ; prêts ; octroi de garanties et souscription d'engagements, etc.
- (BL 4) Banque commerciale (*commercial banking*) : financement de projets, immobilier, financement d'exportations, affacturage, crédit-bail, prêts, garanties, etc.



- (BL 5) Paiements et règlements (*payment and settlement*) : émission et gestion de moyens de paiement, paiements et recouvrements, transferts de fonds, etc.
- (BL 6) Fonctions d'agent (*agency services*) : conservation, dépôts fiduciaires, certificats de titres en dépôt, prêts de titres, etc.
- (BL 7) Gestion d'actifs (*asset management*) : gestion de portefeuille discrétionnaire ou non discrétionnaire.
- (BL 8) Courtage de détail (*retail brokerage*).

Cette répartition en deux dimensions permet d'obtenir une vision matricielle des risques opérationnels que peut subir un établissement bancaire :

	EL 1    ...    EL 7
BL 1	
⋮	
BL 8	

## 2.2 Dispositif réglementaire

Trois approches sont proposées pour le calcul de l'exigence de fonds propres réglementaires. Les fonds propres sont déterminés de manière à couvrir à la fois les pertes exceptionnelles (*unexpected loss*) et les pertes attendues (*expected loss*), la mesure du risque opérationnel correspondant à une valeur en risque, c'est-à-dire à la somme de l'*expected loss* et de l'*unexpected loss*.

### Approche indicateur de base (*Basic Indicator Approach, BIA*)

Il s'agit de la méthode la plus simple : le calcul du capital se fait à partir d'un indicateur d'exposition fondé sur le produit annuel brut (PAB), qui intègre le produit net bancaire et d'autres éléments d'exploitation. Au titre de l'année  $n$ , l'indicateur retenu est la moyenne des parties positives des produits annuels bruts (PAB) des trois dernières années :

$$IE_n = \frac{1}{3} \sum_{i=1}^3 \max(\text{PAB}_{n-i}, 0).$$

L'exigence en fonds propres  $FP_n$  vaut alors  $FP_n = 15\% \times IE_n$ .

Cette mesure est donc reliée aux résultats de la banque plutôt qu'à son exposition au risque opérationnel (et à la qualité de sa mesure). Par ailleurs, cette méthode fournit le plus souvent des estimations de fonds propres surévaluées au regard d'autres méthodes basées sur un modèle interne. À l'inverse, en période de crise, alors que la banque est plus exposée, les fonds propres mobilisés sont moindres, ce qui accroît la vulnérabilité de la banque. Les établissements d'envergure internationale ont par conséquent opté pour des méthodes plus complexes mais plus réalistes pour évaluer leur exposition aux risques opérationnels.

### Approche standardisée (*Standardized Approach, TSA*)

Il s'agit d'un prolongement plus fin de l'approche précédente qui décline ce calcul par type d'activité. Dans cette approche, les fonds propres pour l'année  $n$  sont égaux à la somme des fonds propres  $FP_n^{(i)}$  de chaque catégorie d'activité  $i$  :

$$FP_n = \sum_{i=1}^8 FP_n^{(i)} = \sum_{i=1}^8 \beta_i \times IE_n^{(i)}$$

Les coefficients  $\beta_i$  proposés par le Comité de Bâle sont les suivants :

$i$	Ligne métier $i$	$\beta_i$
1	Financement d'entreprise	18%
2	Activités de marché	18%
3	Banque de détail	12%
4	Banque commerciale	15%
5	Paiement et règlement	18%
6	Fonctions d'agent	15%
7	Gestion d'actifs	12%
8	Courtage de détail	12%

Pour utiliser la méthode TSA, une banque doit notamment être en mesure de définir un produit annuel brut pour chaque catégorie d'activité, en lien avec les processus de gestion des risques de l'établissement.

### Approche avancée (*Advanced Measurement Approach, AMA*)

En ce qui concerne l'approche avancée, le Comité de Bâle ne précise pas la méthode à employer pour estimer la charge en capital à mobiliser au titre du risque opérationnel. La banque doit cependant être à même de démontrer que son modèle prend en compte des « événements exceptionnels générateurs de pertes potentiellement sévères » et correspond à une période de détention d'un an pour un niveau de confiance  $\alpha = 99,9\%$ , ce qui correspond à la couverture d'un risque millénaire.

Bien que la méthode de calcul ne soit pas imposée dans le texte actuel, trois méthodes ont été envisagées par le passé :

- l'*Internal Measurement Approach* consiste à reproduire l'approche IRB utilisée en risque de crédit en considérant notamment les probabilité d'occurrence et perte moyenne en cas d'événement ;
- l'approche *Scorecard* consiste à déterminer les facteurs de risque auxquels est exposé la banque, puis à évaluer les niveaux d'exposition à ces facteurs sous forme de scores.
- l'approche par distribution de pertes (LDA, pour *Loss Distribution Approach*) est la méthode la plus couramment utilisée actuellement par les banques et est décrite en détail dans la partie suivante.

## 2.3 Approche par distributions de pertes : méthode LDA

L'objectif de la méthode LDA est de modéliser la perte liée au risque opérationnel pour une période d'une année afin d'en déduire la valeur en risque. La particularité de cette perte annuelle est qu'elle résulte de plusieurs pertes (individuelles) successives. Elle se détermine autour de deux dimensions :

- la **fréquence** des pertes, qui correspond au nombre de pertes individuelles, a priori inconnu pour l'année à venir, et que l'on modélise par un processus de comptage ;
- la **sévérité** des pertes, qui modélise le montant de chacune de ces pertes individuelles.

La perte annuelle est donc la somme (aléatoire) de ces pertes individuelles, sa distribution est donc déterminée à partir des distributions de fréquence et de sévérité. L'objectif est alors d'estimer avec précision l'exigence en fonds propres réglementaires correspondant à la couverture d'un risque millénaire, la période de détention  $h$  étant fixée à un an et le niveau de confiance  $\alpha$  à 99,9%.

En tant qu'approche par distributions de pertes, elle s'appuie sur l'ensemble des pertes collectées dans les bases de données internes. Par mesure de fiabilité, les banques collectent les pertes à partir d'un certain seuil  $H$  : aucune perte d'un montant inférieur à  $H$  ne figure dans les bases de données et ceci constitue, on va le voir, une difficulté supplémentaire pour l'estimation des paramètres de sévérité.

## Formalisation mathématique

On fait l'hypothèse que les pertes individuelles sont indépendantes et identiquement distribuées. On considère  $I$  catégories d'activité différentes ( $i = 1, \dots, I$ ) et  $J$  types de risque ( $j = 1, \dots, J$ ).  $X_{(i,j)}$  est la variable aléatoire représentant le montant d'une perte pour l'activité  $i$  et le type de risque  $j$ . La distribution de la sévérité des pertes est notée  $F_{(i,j)}$ . Le nombre d'occurrences entre les dates  $t$  et  $t + h$  est modélisé par une variable de comptage  $N_{(i,j)}$ , de fonction de probabilité  $p_{(i,j)}$ . La distribution de la fréquence des pertes  $P_{(i,j)}$  correspond alors à  $P_{(i,j)}(n) = \sum_{k=1}^n p_{(i,j)}(k)$ .

La perte pour la ligne de métier  $i$  et le type de risque  $j$  entre les dates  $t$  et  $t + h$  vaut donc

$$L_{(i,j)} = \sum_{k=1}^{N_{(i,j)}} X_{(i,j)}^k$$

et sa fonction de répartition est donc la distribution composée suivante :

$$\Lambda_{(i,j)}(x) = \mathbb{P}[L_{(i,j)} \leq x] = \begin{cases} \sum_{k=0}^{+\infty} p_{(i,j)}(k) F_{(i,j)}^{k*}(x) & \text{si } x > 0 \\ p_{(i,j)}(0) & \text{si } x = 0 \end{cases}$$

où  $F_{(i,j)}^{k*}$  est la fonction de répartition de  $X_{(i,j)}^1 + \dots + X_{(i,j)}^k$ .

La charge en capital (CaR, pour *Capital-at-Risk*) pour l'activité  $i$  et le type de risque  $j$  correspond alors au quantile  $\alpha$  de  $\Lambda_{(i,j)}$  :

$$\text{CaR}_{(i,j)}(\alpha) = \Lambda_{(i,j)}^{-1}(\alpha) = \inf \{x \mid \Lambda_{(i,j)}(x) \geq \alpha\}$$

Au sein d'une unité de mesure  $(i, j)$ , la méthodologie LDA utilisée pour le calcul du risque opérationnel se décompose en quatre étapes :

- (1) estimation de la distribution de sévérité ;
- (2) estimation de la distribution de fréquence ;
- (3) construction de la distribution annuelle des pertes ;
- (4) calcul de la charge en capital (quantile à 99,9%).

### 2.3.1 Estimation des paramètres

#### Estimation de la distribution de sévérité

La troncature des données liée au seuil de collecte  $H$  affecte l'estimation des paramètres de sévérité puisque la distribution empirique (pertes effectivement collectées) est différente de la vraie distribution (celle que l'on doit obtenir si aucune perte n'est négligée). Il faut donc relier la vraie distribution à sa version empirique, en considérant la densité conditionnelle :

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

On suppose que l'on a collecté un échantillon de  $n$  pertes de montants  $x_1, \dots, x_n$  (supposés i.i.d.) au-delà du seuil  $H$ , et on cherche à déterminer le "vrai" vecteur de paramètres  $\theta_0$ .

Les méthodes du maximum de vraisemblance (*Maximum Likelihood*, ML) et des moments généralisée (*Generalized Method of Moments*, GMM) sont plus usuellement appliquées pour l'estimation des paramètres de distribution de sévérité.

Ces méthodes sont décrites en détail dans la partie 3.3.

## Estimation de la distribution de fréquence

La distribution de la fréquence est le plus souvent modélisée par une loi de Poisson  $\mathcal{P}(\lambda)$ , de fonction de probabilité  $p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ . Cette distribution présente l'avantage de ne nécessiter qu'un seul paramètre  $\lambda$ , qui s'estime aisément par la méthode du maximum de vraisemblance comme la moyenne empirique du nombre annuel de pertes. Elle est par ailleurs largement utilisée dans le domaine de l'assurance pour modéliser des phénomènes similaires.

On calibre la distribution de la fréquence après celle de la sévérité afin de prendre en compte la présence du seuil de collecte. En effet, la probabilité qu'une perte soit supérieure au seuil de collecte est égale au ratio du nombre de pertes collectées sur le nombre total "réel" de pertes. On corrige donc l'estimation du paramètre de fréquence en le divisant par la proportion de pertes supérieures au seuil  $H$  :

$$\hat{\lambda} = \frac{\hat{\lambda}_H}{\mathbb{P}[X > H]} = \frac{\hat{\lambda}_H}{1 - F_{\hat{\theta}}(H)}$$

En pratique, il suffit de calculer la moyenne empirique du nombre de pertes annuel (qui est une estimation de  $\lambda_H$ ) et d'utiliser l'estimateur de  $\hat{\theta}$  pour obtenir la vraie valeur du paramètre de la distribution de la fréquence.

### 2.3.2 Calculs de charges en capital

Une fois que les paramètres  $\hat{\theta}$  et  $\hat{\lambda}$  sont calibrés, on a déterminé les distributions de fréquence  $N$  et de sévérité  $X$  et le calcul de la charge en capital est assez simple. Il s'agit essentiellement de construire (au moyen de méthodes numériques) la distribution de la perte annuelle

$$L = \sum_{k=1}^N X_k.$$

On détaille ci-après la construction de la perte annuelle dans le cas log-normal, avec  $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ .

**Méthode de Monte Carlo.** Cette construction peut-être menée grâce à la méthode de Monte Carlo, en simulant un grand nombre (noté  $S$ ) de réalisations de la perte annuelle.

Pour chaque simulation  $s \in \llbracket 1, S \rrbracket$  :

- (1) on effectue un tirage  $n^s$  de la variable de comptage  $N \sim \mathcal{P}(\hat{\lambda})$  ;
- (2) on simule un échantillon log-normal  $\mathcal{LN}(\hat{\mu}, \hat{\sigma})$  de taille  $n^s$ ,  $(x_1^s, \dots, x_{n^s}^s)$  ;
- (3) on obtient une simulation de la perte annuelle en écrivant

$$l^s = \sum_{k=1}^{n^s} x_k^s.$$

On obtient ainsi un vecteur des pertes annuelles  $l = (l_1, \dots, l_S)'$  et la charge en capital (CaR) est alors le quantile à  $\alpha = 99,9\%$  de  $l$ , c'est-à-dire la  $(\alpha \times S)^{\text{ème}}$  plus grande valeur de  $l$ .

**Fast Fourier Transform (FFT).** La perte annuelle étant une convoluée des distributions de fréquence et de sévérité, il peut être intéressant de recourir aux transformées de Fourier pour la déterminer. La transformée de la densité de la sévérité fournit sa fonction caractéristique.

$$\mathcal{F}(f_X)(v) = g(v) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(u) e^{-ivu} du$$

L'algorithme FFT est employé pour le calcul de transformées de Fourier de façon rapide basé sur la discrétisation des fonctions à transformer. La fonction caractéristique d'une somme de variables aléatoires indépendantes étant le produit des fonctions caractéristiques de ces variables, on peut écrire

$$\phi_{X_1+\dots+X_n}(u) = \prod_{k=1}^n \phi_{X_k}(u) = (\phi_X(u))^n$$

La fonction caractéristique de la distribution des pertes annuelles peut s'écrire ( $N$  étant indépendante de  $X$ ) :

$$\phi_L(u) = g_N \circ \phi_X(u) = \exp(\lambda(\phi_X(u) - 1))$$

avec  $g_N(t) = \sum_{n=0}^{+\infty} \mathbb{P}[N = n]t^n = \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} e^{-\lambda} t^n = e^{\lambda(t-1)}$  la fonction génératrice de  $N \sim \mathcal{P}(\lambda)$ .

La densité de la distribution des pertes annuelles  $f_L$  se déduit par transformée de Fourier inverse :

$$f_L = \mathcal{F}^{-1}(\phi_L)$$

La fonction de répartition  $F_L$  est construite par intégration et on détermine ainsi la charge en capital :

$$\text{CaR}(\alpha) = F_L^{-1}(\alpha)$$

**Prise en compte des assurances.** Le Comité de Bâle autorise les banques « à prendre en compte la couverture d'assurance comme technique d'atténuation des risques » dans la mesure où cette prise en compte n'excède pas à 20 % de l'exigence de fonds propres totale. Pour intégrer une couverture d'assurance de montant  $A$  comprenant une franchise  $F$ , on définit chaque perte unitaire avec prise en compte des assurances, soit  $\bar{x}_k^s = x_k^s \times \mathbb{1}_{\{x_k^s < F\}} + \max(F, x_k^s - A) \times \mathbb{1}_{\{x_k^s \geq F\}}$ , puis on calcule la charge en capital à partir de la perte totale ainsi obtenue par la méthode de Monte Carlo.

## Agrégation des charges en capital

Le Comité de Bâle indique que la banque est autorisée à « appliquer des corrélations déterminées en interne à condition de démontrer que ses systèmes de détermination des corrélations sont sains, mis en œuvre avec intégrité et tiennent compte des incertitudes inhérentes à toute estimation de corrélations ». Il est raisonnable de penser que les pertes annuelles  $L_{i,j}$  ne sont pas parfaitement corrélées et faire la somme de toutes les charges en capital est donc sûrement très conservateur. Le modèle LDA impose l'hypothèse d'indépendance entre les pertes d'un même type de risque mais il est difficile de supposer à la fois cette indépendance et de supposer une corrélation des pertes entre types de risque. On peut donc faire l'hypothèse que la corrélation entre les pertes annuelles est une corrélation des fréquences. Pour deux pertes annuelles  $L_1 = \sum_{n=1}^{N_1} X_n^1$  et  $L_2 = \sum_{n=1}^{N_2} X_n^2$ , en supposant que les deux fréquences sont parfaitement corrélées ( $N_1 = N_2 = N$ ) et de loi  $\mathcal{P}(\lambda)$  :

$$\begin{aligned} \text{Cov}[L_1, L_2] &= \mathbb{E}[N^2 \mathbb{E}[X^1] \mathbb{E}[X^2]] - \lambda^2 \mathbb{E}[X^1] \mathbb{E}[X^2] \\ &= (\mathbb{V}[N] + \mathbb{E}^2[N] - \lambda^2) \mathbb{E}[X^1] \mathbb{E}[X^2] \\ &= \lambda \mathbb{E}[X^1] \mathbb{E}[X^2] \end{aligned}$$

On en déduit que la corrélation entre  $L_1$  et  $L_2$  ne dépend pas de  $\lambda$  et vaut :

$$\text{Cor}[L_1, L_2] = \frac{\mathbb{E}[X^1] \mathbb{E}[X^2]}{\sqrt{(\mathbb{V}[X^1] + \mathbb{E}^2[X^1])(\mathbb{V}[X^2] + \mathbb{E}^2[X^2])}},$$

soit, pour des sévérités log-normales :  $\text{Cor}[L_1, L_2] = e^{-\frac{1}{2}(\sigma_1^2 + \sigma_2^2)}$ .

On peut alors établir des corrélations entre les différentes lignes métier et catégories de risque et en déduire un calcul de la charge en capital global, en faisant par exemple l'hypothèse que le vecteur des pertes annuelles a la structure d'un vecteur gaussien. On peut dans ce cas écrire :

$$\text{CaR}(\alpha) = \sum_{k=1}^K \text{EL}_k + \sqrt{\sum_{k=1}^K \sum_{k'=1}^K \rho_{k,k'} (\text{CaR}_k(\alpha) - \text{EL}_k) (\text{CaR}_{k'}(\alpha) - \text{EL}_{k'})}$$

avec  $\text{CaR}_k(\alpha) = F_{L_k}^{-1}(\alpha)$ ,  $\text{EL}_k = \mathbb{E}[L_k] = \hat{\lambda}_k \exp(\hat{\mu}_k + \hat{\sigma}_k^2/2)$ , où  $k$  et  $k'$  représentent deux couples (catégorie de risque)  $\times$  (ligne métier) et  $\rho_{k,k'}$  est la corrélation entre les pertes annuelles  $L_k$  et  $L_{k'}$  établie selon la méthode décrite ci-dessus.

Il est par ailleurs possible de recourir à l'usage de copules pour mettre en œuvre l'agrégation des charges en capital. Cette méthode présente l'avantage de ne nécessiter qu'une hypothèse sur la forme de la structure de dépendance et pas sur les distributions marginales, au contraire de l'hypothèse de vecteur gaussien.

## 3 Modélisation de la 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  $\theta$ . 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 note  $(x_1, \dots, x_n)$  l'échantillon de pertes à modéliser.

### 3.1 Contraintes spécifiques

Les données de pertes opérationnelles possèdent des profils très atypiques. Malgré la classification des risques et l'amélioration continue 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.

#### Le seuil de collecte

Comme l'exige 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  $f_\theta$  de la distribution sévérité lors de l'inférence, on considère la densité conditionnelle par rapport à  $H$  :

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

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 :

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

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

#### 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"  $X^\kappa = \left(\frac{x_1}{\kappa_1}, \dots, \frac{x_n}{\kappa_n}\right)$ .

## 3.2 Loïs de probabilité

De nombreuses lois de probabilités peuvent être utilisées pour modéliser la sévérité des pertes opérationnelles. Ces lois doivent être à support positif et asymétriques avec une queue à droite plus épaisse. En probabilités, il est d'usage de classer les lois à support positif en trois catégories : les lois "light-tailed" (à queues fines voire finies), les lois "medium-tailed" (à queues moyennes) et les lois "heavy-tailed" (à queues épaisses). Par exemple, la loi Beta appartient à la première catégorie, la loi Exponentielle à la deuxième catégorie et la loi de Pareto translatée à la troisième catégorie. Par conséquent, on peut dresser une liste des lois les plus couramment utilisées en risque opérationnel (cf. Tableau 3.1)

Loi de probabilité	Densité
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	$\frac{\alpha}{\theta} \left(1 + \frac{x}{\theta}\right)^{-(\alpha+1)}$
Burr	$\frac{\alpha\gamma}{\theta\gamma} \left[1 + \left(\frac{x}{\theta}\right)^\gamma\right]^{-\alpha}$

Table 3.1 – Exemple de lois utilisées pour la modélisation du risque opérationnel.

L'ajustement de la sévérité par différentes lois de probabilité a fait l'objet de nombreuses projets de recherche depuis quelques années. Suite à ces travaux, la loi adoptée pour la modélisation de la sévérité est la loi log-normale. Elle donne en effet des ajustements de bonne qualité en comparaison avec d'autres lois mais elle présente surtout l'avantage d'être simple à estimer et d'être la loi de référence depuis plusieurs années au Crédit Agricole. Les paramètres  $\mu$  et  $\sigma$  permettent une interprétation directe du modèle estimé.

## 3.3 Méthodes d'estimation standards

Deux méthodes d'estimation classiques des paramètres d'une loi de probabilité sont présentées ici : le maximum de vraisemblance et la méthode des moments généralisée. Elles sont décrites dans le cadre de la loi log-normale<sup>1</sup>  $\mathcal{LN}(\mu, \sigma)$ , de densité

$$f_{\mu, \sigma}(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}.$$

### Méthode du maximum de vraisemblance (ML)

C'est la méthode la plus couramment utilisée et celle qui a les meilleures propriétés sur données simulées. Le principe est de rechercher les paramètres d'une loi de sorte à maximiser la vraisemblance des données. En pratique, la maximisation s'effectue le plus souvent sur la log-vraisemblance, ce qui est équivalent.

La log-vraisemblance de l'échantillon  $x_1, \dots, x_n$  s'écrit

$$\ell(\theta; x_1, \dots, x_n) = \sum_{k=1}^n \ln \tilde{f}_\theta(x_k) = \sum_{k=1}^n \ln f_\theta(x_k) - n \ln(1 - F_\theta(H))$$

1. Il s'agit de la loi la plus utilisée dans l'industrie bancaire.



L'estimateur du maximum de vraisemblance  $\hat{\theta}_{\text{ML}}$  est alors défini par

$$\hat{\theta}_{\text{ML}} = \arg \max_{\theta} \ell(\theta; x_1, \dots, x_n)$$

et on peut montrer que  $\hat{\theta}_{\text{ML}}$  est sans biais, asymptotiquement efficace et normal :

$$\sqrt{n}(\hat{\theta}_{\text{ML}} - \theta_0) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathcal{I}^{-1}(\theta_0)),$$

avec  $\mathcal{I}(\theta_0)$  la matrice d'information de Fisher et  $\theta_0$  la vraie valeur du vecteur des paramètres. Dans le cas log-normal, la log-vraisemblance vaut :

$$\ell(\mu, \sigma; x_1, \dots, x_n) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{k=1}^n (\ln x_k - \mu)^2 + \sum_{k=1}^n \ln \frac{\mathbb{1}_{\{x_k \geq H\}}}{x_k} - n \ln \left( 1 - \Phi\left(\frac{\ln H - \mu}{\sigma}\right) \right)$$

Le dernier terme corrige le seuil de collecte et a un impact important sur l'estimation. Il dépend en effet de  $\mu$  et  $\sigma$  et il n'existe pas de formules fermées pour les estimateurs du maximum de vraisemblance. On doit donc les calculer par optimisation numérique.

Cette méthode peut s'appliquer à un échantillon tronqué dès lors que l'on connaît une expression explicite de la densité conditionnelle au seuil de collecte. L'optimisation de la vraisemblance se fait de façon numérique. En pratique, le seuil de collecte perturbe le programme d'optimisation ce qui rend la méthode instable et parfois non-convergente. En effet, il est fréquent d'obtenir des estimations de  $\mu$  négatives pour la loi log-normale<sup>2</sup>. C'est la raison pour laquelle elle n'est pas mise en production.

### Méthode des moments généralisée (GMM)

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

$$m_p(\theta) = \mathbb{E}[X^p | X \geq H] \text{ et } \hat{m}_p = \frac{1}{n} \sum_{k=1}^n (x_k)^p$$

En posant  $g(\theta) = (m_1(\theta) - \hat{m}_1, \dots, m_p(\theta) - \hat{m}_p)$ , l'estimateur GMM est défini par :

$$\hat{\theta}_{\text{GMM}} = \arg \min_{\theta} g(\theta)' W^{-1} g(\theta)$$

où  $W$  est une matrice symétrique définie positive, optimale en prenant  $W = \mathbb{V}[g(\theta)]$ . L'estimateur  $\hat{\theta}_{\text{GMM}}$  possède également la propriété de normalité asymptotique.

Pour  $X \sim \mathcal{LN}(\mu, \sigma)$  et à l'ordre  $p$ , on a

$$\begin{aligned} \mathbb{E}[X^p | X \geq H] &= \frac{1}{1 - F_{\mu, \sigma}(H)} \int_H^{+\infty} \frac{x^{p-1}}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{1 - \Phi\left(\frac{\ln H - \mu}{\sigma}\right)} \int_{\ln H}^{+\infty} \frac{e^{-\frac{(y - (\mu + p\sigma^2))^2 - 2p\mu\sigma^2 - p^2\sigma^4}{2\sigma^2}}}{\sigma\sqrt{2\pi}} dy \\ &= \frac{1 - \Phi\left(\frac{\ln H - (\mu + p\sigma^2)}{\sigma}\right)}{1 - \Phi\left(\frac{\ln H - \mu}{\sigma}\right)} e^{p\mu + p^2\sigma^2/2} \end{aligned}$$

Alors pour obtenir les deux estimateurs  $\hat{\mu}_{\text{GMM}}$  et  $\hat{\sigma}_{\text{GMM}}$ , il suffit de déterminer les deux premiers moments  $m_1(\mu, \sigma)$  et  $m_2(\mu, \sigma)$  puis de minimiser numériquement la forme quadratique.

Cette méthode a l'inconvénient d'être biaisée. En effet, elle a tendance à sous-estimer la valeur du paramètre  $\sigma$  de la loi log-normale, ce qui entraîne une sous-estimation de la charge en capital résultante.

2. Il n'est pas réaliste d'obtenir des valeurs de  $\mu$  négatives car cela voudrait dire que d'après le modèle, plus de la moitié des pertes sont inférieures à 1€. En effet, la médiane d'une loi log-Normale est  $e^\mu$ .

## 3.4 Tests d'adéquation

Une fois l'estimateur  $\hat{\theta}$  déterminé, on souhaite évaluer l'adéquation de la distribution estimée aux données. On teste pour ce faire l'hypothèse  $H_0$  : "les données de l'échantillon suivent une loi de fonction de répartition  $F_{\hat{\theta}}$ " contre  $H_1$  : "l'échantillon ne suit pas une loi de fonction de répartition  $F_{\hat{\theta}}$ ".

Les données utilisées pour l'estimation étant tronquées, on utilise des statistiques de tests corrigées. On choisit un niveau de test égal à 5%. On accepte alors l'hypothèse  $H_0$  si la p-value<sup>3</sup> du test est supérieure à 5%.

Dans cette étude, l'ajustement est mesuré au moyen de quatre statistiques de test différentes, prenant en compte la troncature des données au seuil  $H$ . L'expression détaillée de ces statistiques est présentée ci-après, et sont issues de l'article de Chernobai *et al.* [2]. On note ici  $z_H = F_{\theta}(H)$  et  $z_j = F_{\theta}(x_{(j)})$  où  $x_{(j)}$  est la  $j^{\text{ème}}$  plus grande perte de l'échantillon.

### Test de Kolmogorov-Smirnov (KS)

Le test de Kolmogorov-Smirnov est un test local, qui mesure la distance maximale entre les fonctions de répartition théorique et empirique :

$$KS = \sqrt{n} \sup_{x \in \mathbb{R}} \left\{ \left| \widehat{F}_n(x) - F_{\theta|H}(x) \right| \right\}$$

Après discrétisation, on obtient :

$$KS = \sqrt{n} \max\{KS_+, KS_-\}$$

$$\text{avec } KS_+ = \max_{j \in [1, n]} \left\{ \left| \frac{j}{n} - \frac{z_j - z_H}{1 - z_H} \right| \right\} \text{ et } KS_- = \max_{j \in [1, n]} \left\{ \left| \frac{z_j - z_H}{1 - z_H} - \frac{j-1}{n} \right| \right\}.$$

On a finalement :

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

### Test de Cramér-Von-Mises (CvM)

Ce test de Cramér-von-Mises est un test global, qui mesure la distance quadratique entre les fonctions de répartition théorique et empirique, en attribuant à chaque observation le même poids :

$$CvM = n \int_{-\infty}^{+\infty} (\widehat{F}_n(x) - F_{\theta|H}(x))^2 \psi(F_{\theta}(x)) dF_{\theta|H}(x), \text{ avec } \psi(F_{\theta}(x)) = 1$$

Après discrétisation, on obtient :

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

### Test quadratique d'Anderson-Darling (AD2)

Ce test est dérivé du test de Cramér-von-Mises, mais attribue un poids plus important aux queues de distribution droite et gauche, avec  $\psi(F_{\theta}(x)) = [F_{\theta}(x)(1 - F_{\theta}(x))]^{-1}$ . Le test est focalisé sur les parties de la distribution où  $F_{\theta}(x)(1 - F_{\theta}(x))$  prend de petites valeurs, c'est-à-dire sur les queues gauche et droite de la distribution :

3. Qui correspond à la probabilité de rejeter à tort l'hypothèse nulle.

$$AD^2 = n \int_{-\infty}^{+\infty} \frac{(\widehat{F}_n(x) - F_{\theta|H}(x))^2}{F_{\theta|H}(x)(1 - F_{\theta|H}(x))} dF_{\theta|H}(x)$$

Après discrétisation, on obtient :

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

### Test quadratique modifié d'Anderson-Darling (AD2up)

Ce dernier test est également dérivé du test de Cramér-von-Mises, mais attribue un poids plus important à la queue de distribution droite qui contient donc les pertes les plus sévères, avec  $\psi(F_{\theta}(x)) = (1 - F_{\theta}(x))^{-2}$  :

$$AD_{up}^2 = n \int_{-\infty}^{+\infty} \frac{(\widehat{F}_n(x) - F_{\theta|H}(x))^2}{(1 - F_{\theta|H}(x))^2} dF_{\theta|H}(x)$$

Après discrétisation, on obtient :

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

### Implémentation de ces tests d'adéquation

En raison de la troncature des données, les statistiques ci-dessus ne sont pas issues d'une distribution standard. On recourt donc à l'algorithme<sup>4</sup> suivant pour déterminer les p-values.

Algorithme : Calcul de p-value par bootstrap paramétrique.

On fixe le seuil de rejet d'hypothèse  $\alpha$ , la taille de l'échantillon initial  $n_0$ , la valeur de la statistique pour l'échantillon  $D$ , l'hypothèse à tester  $H_0$ .

- (1) Simuler un grand nombre  $S$  d'échantillons
- (2) Pour  $s \in \llbracket 1, S \rrbracket$  :
  - (a) estimer les paramètres  $\widehat{\theta}_s$  de l'échantillon  $s$
  - (b) calculer la statistique  $d_s$  correspondant à cet échantillon
- (3) Calculer la p-value comme :  $p = \frac{1}{S} \sum_{s=1}^S \mathbb{1}_{\{d_s > D\}}$ .
- (4) Si  $p < \alpha$ , rejeter  $H_0$ .

En pratique, on prendra  $S = 1000$ , qui est un bon compromis entre temps de calcul et précision de la p-value, et  $\alpha = 5\%$ .

Une analyse comparative de ces différents tests est menée dans l'article [13].

4. Il s'agit de l'algorithme proposé par Chernobai *et al.* [2].

## 3.5 Objectifs du mémoire

Le texte "*Operational Risk - Supervisory Guidelines for the Advanced Measurement Approaches*" [1] publié par le Comité de Bâle en juin 2011 est le premier texte donnant les orientations à suivre dans le cadre des approches avancées. Il préconise un renforcement de la méthodologie employée pour estimer les besoins en fonds propres et incite les banques à justifier et à documenter le choix du modèle. Certaines de ces directives concernent la loi utilisée pour la sévérité ainsi que la méthode d'estimation associée. Il est très important de prendre en compte ces préconisations pour pouvoir anticiper les exigences du régulateur.

D'après le texte, un modèle est dit pertinent s'il est :

- (1) réaliste (il génère une distribution de perte annuelle qui donne une estimation réaliste du capital requis sans avoir à recourir à des ajustements correctifs)
- (2) bien spécifié (les caractéristiques du modèle sont similaires aux données de pertes et logiquement cohérentes)
- (3) flexible (le modèle peut s'adapter à des profils de pertes divers)
- (4) simple (facile à implémenter et à utiliser pour simuler la perte annuelle)

Concernant la distribution de sévérité, le texte précise que le choix d'une distribution de probabilité doit reposer sur une analyse exploratoire des données renforcée ainsi que sur des outils de diagnostic de l'ajustement, en particulier sur la queue de distribution. Le Comité de Bâle recommande l'utilisation de lois dites "subexponentielles" qui représentent mieux le caractère leptokurtique des distributions de pertes. Il incite fortement les banques à adopter une modélisation séparée corps-queue.

Il est mentionné que la méthode d'estimation employée doit respecter des **critères de robustesse** sans pour autant sous-estimer le risque en queue de distribution. Enfin, le texte met en garde les banques qui voudraient utiliser des modèles à moyenne infinie, les charges en capital résultantes n'étant généralement pas applicables<sup>5</sup>.

Au sujet des méthodes d'estimation, il est précisément indiqué dans le texte [1] :

*Robust estimation methods (such as alternatives to classical methods as the Maximum Likelihood and the Probability Weighted Moments), proposed recently in operational risk literature, are reasonably efficient under small deviations from the assumed model. These methods also highlight which observations or deviating substructures have the greatest influence on the statistic to be estimated. A bank may adopt alternatives to classic estimators, provided it can demonstrate that its use does not underestimate risk in the tail.*

C'est dans ce contexte que ce mémoire a pour objet l'étude de la notion de robustesse des estimateurs. Il propose également la définition d'estimateurs conçus de manière à vérifier la propriété de robustesse, avant d'explorer de manière détaillée leur comportement dans le cadre d'un calcul de capital réglementaire au titre du risque opérationnel.

---

5. Dans le cas où la moyenne de la loi est infinie, la CaR étant un quantile d'ordre très élevé, elle est en général trop élevée pour être applicable.

## Partie II

---

# **Modèles robustes**

## 4 Statistique robuste

---

La statistique robuste est une branche de la statistique qui a vu le jour dans les années 1960 avec les premières contributions de Tukey [20]) puis Huber [9] ou encore Hampel [8]. La statistique robuste a pour vocation de fournir des méthodes d'estimation statistique que l'on qualifiera de "résistantes" par rapport aux points marginaux s'écartant du modèle sous-jacent, ainsi qu'à l'ajout ou au retrait de données dans l'échantillon. Cette théorie a été utilisée dans de nombreux domaines, mais son application au risque opérationnel est relativement récente, notamment avec les travaux d'Opdyke et Cavallo [15].

Une des motivations premières à l'application de telles méthodes est le constat suivant : les pertes collectées au titre du risque opérationnel ne s'inscrivent pas dans un cadre statistique idéal, qui serait que ces données soient toutes identiquement distribuées suivant le même modèle paramétrique (iid).<sup>1</sup> (iid). Cependant, cette hypothèse doit être satisfaite afin que les propriétés de l'estimateur du maximum de vraisemblance (ML) soient vérifiées.

En effet, pour être appliqué, l'estimateur ML requiert la vérification de trois hypothèses :

- les pertes sont indépendantes ;
- les données sont homogènes, c'est-à-dire identiquement distribuées ;
- le modèle probabiliste sous-jacent est clairement spécifié.

Si l'une de ces conditions est infirmée, alors l'estimateur ML perd ses propriétés statistiques, il n'est alors plus asymptotiquement normal, sans biais et efficace. Cependant, la statistique robuste a un champ d'application plus large que les méthodes classiques car cette théorie ne nécessite pas de telles contraintes, rarement observées en pratique.

En outre, il est observable que l'ajout d'une seule perte peut perturber fortement l'estimation des paramètres du modèle de risque opérationnel, ce qui a pour conséquence d'entraîner un calcul de charges en capital erroné, pouvant atteindre des montants incohérents. L'enjeu est donc de pouvoir construire un estimateur robuste capable d'être "résistant" à la modification d'une partie des données de pertes, mais également à l'ajout de pertes extrêmes et arbitraires.

Cette partie est consacrée aux fondements de la statistique robuste ainsi qu'aux outils fondamentaux dans l'étude de la robustesse d'estimateurs statistiques.

### 4.1 Fonction d'influence et B-robustesse

La fonction d'influence est un outil analytique puissant dans le domaine de la statistique robuste qui généralise le concept de dérivée à un estimateur, ou plus généralement à une fonctionnelle.

C'est en se basant sur l'étude de cette fonction que l'on peut déterminer si un estimateur est B-robuste ou non. Avant d'aborder le concept de fonction d'influence, nous présentons la définition préliminaire d'une fonctionnelle.

---

1. Cette caractérisation est développée dans les articles [3], [4] et [15].

## Fonctionnelle statistique

Une fonctionnelle statistique  $T$  est une fonction prenant en argument une fonction de répartition  $F_\theta$  paramétrée par  $\theta$ . Ainsi, pour un échantillon  $(x_1, \dots, x_n)$  issu d'une loi de probabilité  $F_\theta$ , le principe de l'estimation consiste à associer à un estimateur  $\hat{\theta}_n$  une fonctionnelle  $T$ . On dit à ce titre qu'un estimateur est consistant au sens de Fisher si  $T(F_\theta) = \theta$ . Pour l'échantillon  $(x_1, \dots, x_n)$  donné dans la pratique, la fonction de répartition empirique associée est définie par

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{x_k \leq x\}} \quad (4.1)$$

Dans ce cas, l'estimateur  $\hat{\theta}_n$  associé à la fonctionnelle  $T$ , défini sur l'échantillon est donné par

$$\hat{\theta}_n = T(F_n)$$

Les concepts d'estimateur et de fonctionnelle statistique sont très proches et seront parfois interchangeables dans la suite. Voici deux exemples de fonctionnelles associées à la moyenne et la médiane :

- moyenne :  $T(F_\theta) = \int x dF_\theta(x)$
- médiane :  $T(F_\theta) = F_\theta^{-1}\left(\frac{1}{2}\right)$

Une autre notion importante avant d'introduire le concept de fonction d'influence est celle de contamination.

## Contamination

Comme mentionné précédemment, les pertes collectées au titre du risque opérationnel sont rarement identiquement distribuées. Afin de s'écarter d'un modèle idéal pour lequel l'ensemble des données de l'échantillon serait *i.i.d.* issues d'une distribution  $F_\theta$ , une idée plutôt intuitive est de contaminer cet échantillon "idéal", ou du moins de se placer dans un voisinage de cette distribution  $F_\theta$ .

Plusieurs techniques de contamination existent (cf. articles [18] et [19]), la plus élémentaire et appréhendable d'entre elles est appelée contamination convexe. Cette méthode consiste à contaminer un pourcentage  $\varepsilon$  de l'échantillon, en supposant alors qu'une proportion  $(1 - \varepsilon)$  suit le modèle  $F_\theta$  et donc que la proportion  $\varepsilon$  restante provient d'un autre modèle  $G$ . Ainsi, pour l'ensemble des observations, on a un modèle de la forme

$$F_{\theta,\varepsilon} = (1 - \varepsilon)F_\theta + \varepsilon G \quad (4.2)$$

où le modèle  $G$  est inconnu, incontrôlable et imprévisible.

Il est important de comprendre qu'on ne s'intéresse ici ni à la distribution de contamination  $G$  ni au pourcentage de contamination  $\varepsilon$  qui représentent une partie de l'échantillon déviant du modèle recherché  $F_\theta$ , mais que le but de la statistique robuste est de pouvoir estimer avec précision le paramètre  $\theta$  malgré la contamination, ou, dans une moindre mesure, de pouvoir contrôler la perturbation induite par la contamination.

### 4.1.1 Fonction d'influence

La fonction d'influence correspond à une dérivée d'une fonctionnelle  $T$ . Or, pour définir une dérivée, il est nécessaire de définir également un taux d'accroissement. Comme l'argument d'une fonctionnelle  $T$  est une distribution  $F$ , il faut spécifier un accroissement élémentaire dans l'espace des distributions  $\mathcal{F}$ . Ainsi, en utilisant la définition du voisinage d'une fonction de répartition donnée au paragraphe précédent (4.2), on peut définir la dérivée de Gâteaux pour un modèle de contamination quelconque  $G$

$$\text{GD}(x; F_\theta, T) = \lim_{\varepsilon \rightarrow 0} \frac{T((1 - \varepsilon)F_\theta + \varepsilon G) - T(F_\theta)}{\varepsilon}, \quad x \in \mathbb{R} \quad (4.3)$$

La fonction d'influence correspond au cas particulier de la dérivée de Gâteaux pour laquelle  $G = \Delta_x$ , la fonction de répartition de la masse de Dirac en  $x$ . On définit alors la fonction d'influence d'un modèle  $F_\theta$  et d'une fonctionnelle  $T$  calculée au point  $x$  par

$$\text{IF}(x; F_\theta, T) = \lim_{\varepsilon \rightarrow 0} \frac{T((1 - \varepsilon)F_\theta + \varepsilon \Delta_x) - T(F_\theta)}{\varepsilon} \quad (4.4)$$

Ce cas de contamination ( $G = \Delta_x$ ) se traduit par une proportion  $\varepsilon$  de l'échantillon prenant la valeur  $x$ . Cela revient donc à comparer une différence infinitésimale entre l'estimateur de l'échantillon et l'estimateur de ce même échantillon auquel une proportion infime de valeur  $x$  a été ajoutée, le tout rapporté au pourcentage de contamination.

Généralement, la fonction d'influence n'est pas utilisée de cette façon, puisqu'il est souvent difficile d'obtenir des expressions simples des fonctionnelles et donc de ce quotient. En pratique, c'est la **fonction d'influence empirique** qui est utilisée sur des données réelles.

Le jeu d'écriture consiste uniquement à remplacer la fonction de répartition par son homologue empirique et  $\varepsilon$  par  $\frac{1}{n+1}$ ,  $n$  étant le nombre de pertes de l'échantillon

$$\text{EIF}(x; F_n, T) = \frac{T\left((1 - \frac{1}{n+1})F_n + \frac{1}{n+1}\Delta_x\right) - T(F_n)}{1/(n+1)}$$

Ceci permet de pouvoir quantifier réellement l'effet d'une seule donnée ajoutée sur les paramètres estimés. De plus, même pour des échantillons de petite taille, la fonction d'influence empirique et sa version théorique sont souvent très proches tracées simultanément. Cela permet de justifier son utilisation afin d'approximer au mieux le comportement théorique de la fonction d'influence et d'en tirer ainsi des conclusions sur la robustesse des méthodes d'estimation utilisées.

Cette fonction d'influence empirique est aussi appelée courbe de sensibilité de Tukey, dont la signification est identique, et dont l'écriture pour un échantillon  $(x_1, \dots, x_{n-1})$  est

$$\text{SC}(x, T) = n [T(x_1, \dots, x_{n-1}, x) - T(x_1, \dots, x_{n-1})]$$

### 4.1.2 B-robustesse

L'observation graphique de la fonction d'influence permet donc de savoir si un estimateur est "résistant" à un changement d'une faible partie de l'échantillon. Pour un montant  $x$  fixé, plus la valeur de la fonction d'influence est élevée en valeur absolue, plus l'estimateur est sensible à l'ajout d'une perte de ce montant. À ce titre, on dit qu'un estimateur est **B-robuste** si les fonctions d'influence de chacun de ses paramètres sont bornées.

Plus généralement, dans le domaine du risque opérationnel, la fonction d'influence permet de donner une réponse quantitative aux questions suivantes :

- Pour une méthode d'estimation, est-elle B-robuste pour différentes lois de sévérité ?
- Pour une distribution de sévérité donnée, quels estimateurs sont robustes ?
- Quel est l'impact lors de l'ajout d'une nouvelle perte sur les paramètres estimés et sur le capital ?



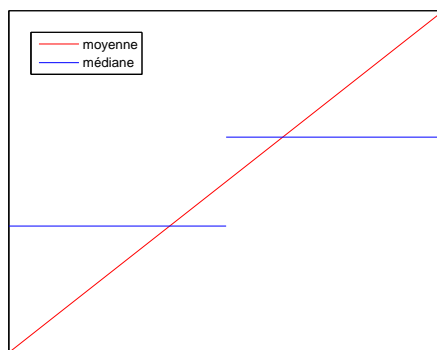


Figure 4.1 – Fonction d'influence de la moyenne et la médiane (ici  $m = \mu$ )

### Exemples de fonctions d'influence

Il peut s'avérer difficile d'obtenir les écritures théoriques des fonctions d'influence pour des fonctionnelles non triviales, ou encore pour des modèles paramétriques complexes. Cependant, certaines d'entre elles sont calculables, comme la moyenne ou la médiane mentionnées plus haut <sup>2</sup>.

Ainsi, pour la moyenne, on a

$$\text{IF}(x; F_\theta, \text{moyenne}) = x - \mu \quad (4.5)$$

Pour l'estimateur du quantile d'ordre  $s$ , on a l'expression

$$\text{IF}(x; F_\theta, T_s) = \begin{cases} \frac{s-1}{f(F_\theta^{-1}(s))} & \text{pour } x < F_\theta^{-1}(s) \\ \frac{1-s}{f(F_\theta^{-1}(s))} & \text{pour } x > F_\theta^{-1}(s) \end{cases} \quad (4.6)$$

Comme la médiane correspond au quantile d'ordre 0.5, sa fonction d'influence est donnée par

$$\text{IF}(x; F_\theta, \text{médiane}) = \begin{cases} \frac{-1}{2f(m)} & \text{pour } x < m \\ \frac{1}{2f(m)} & \text{pour } x > m \end{cases} \quad (4.7)$$

$m$  représentant la médiane de la distribution.

Ces résultats attestent que la moyenne n'est pas robuste contrairement à la médiane. Cela se comprend intuitivement, car en prenant un échantillon de taille  $n$  et de moyenne  $\mu$ , on peut rendre cette moyenne arbitrairement grande en ajoutant une seule donnée de cet échantillon, par exemple en lui donnant une valeur démesurée. Or ce n'est pas le cas de la médiane dans le cas où  $n \geq 3$  <sup>3</sup>.

Afin de rester dans une formulation très mathématique de la notion de robustesse, nous énonçons dans le prochain paragraphe l'idée formelle de robustesse qualitative en termes de métrique et de distance.

2. Les étapes du calcul sont données en Annexe A.

3. Ceci nous conduit à l'introduction d'une définition usuelle en statistique robuste : le point d'effondrement. Le point d'effondrement d'un estimateur est la proportion minimale d'un échantillon que l'on peut modifier arbitrairement sans que l'estimateur ne tende vers l'infini. Ainsi le point d'effondrement de l'estimateur de la moyenne est 0 alors que celui de la médiane est donné par  $\lfloor \frac{n-1}{2} \rfloor$ , où  $\lfloor a \rfloor$  est la partie entière de  $a$ .

### 4.1.3 Robustesse qualitative

La robustesse qualitative est essentiellement équivalente à la continuité de la fonctionnelle  $T$  sur l'ensemble des distributions  $\mathcal{F}$ . Pour définir une continuité, il est nécessaire de définir une distance. La métrique utilisée dans ce cas est celle de Lévy.

**Distance de Lévy.** Soient  $F, G \in \mathcal{F}$  de même support  $\mathcal{X}$ . La distance de Lévy  $\pi : \mathcal{F}^2 \rightarrow \mathbb{R}$  est définie par

$$\pi(F, G) = \inf_{\varepsilon > 0} \{ \forall x \in \mathcal{X}, \quad F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon \}$$

Cette distance sur  $\mathcal{F}$  nous permet alors d'introduire la robustesse qualitative, vue comme la continuité de la fonctionnelle.

**Robustesse qualitative.** On dit que  $T$  est qualitativement robuste en  $F$  si  $T$  est continue en  $F$  pour la distance de Lévy, ce qui s'écrit

$$\forall \varepsilon, \exists \delta \quad \pi(F, G) \leq \delta \Rightarrow |T(F) - T(G)| < \varepsilon$$

Avec cette écriture, on peut montrer que si  $F_n \rightarrow F$  et si  $T$  est qualitativement robuste alors  $T(F_n) \rightarrow T(F)$ .

Bien que l'on comprenne assez intuitivement cette idée, à savoir que deux échantillons relativement proches en termes de fonction de répartition (au sens de la distance de Lévy) fournissent des estimations proches en cas de robustesse qualitative, cette notion est assez difficile à mettre en place pour des estimateurs plus compliqués. Ceci est d'autant moins commode qu'il existe d'autres distances un peu plus complexes comme la distance de Lévy-Prokhorov, ce qui ne facilite pas les choix à effectuer afin d'utiliser ce concept.

C'est pourquoi on préfère l'utilisation des fonctions d'influence, d'une part car la robustesse peut s'observer graphiquement, d'autre part puisqu'on sait en obtenir des calculs explicites pour une certaine catégorie d'estimateurs : les M-estimateurs.

## 4.2 M-estimateurs et maximum de vraisemblance

### 4.2.1 M-estimateurs

Les M-estimateurs sont une catégorie d'estimateurs popularisés par Huber [9] et se caractérisant par

$$T(F) = \arg \min_{\theta} \int_{\mathcal{X}} \rho(x, \theta) d\hat{F}(x)$$

qui se discrétise sur un échantillon donné  $(x_1, \dots, x_n)$  par

$$\hat{\theta} = \arg \min_{\theta} \left( \sum_{i=1}^n \rho(x_i, \theta) \right)$$

Si de plus  $\rho$  est dérivable en  $\theta$ , en notant  $\psi = \frac{\partial \rho}{\partial \theta}$ , on a par équivalence

$$T(F) = \left\{ \theta \mid \int_{\mathcal{X}} \psi(x, \theta) d\hat{F}(x) = 0 \right\} \quad (4.8)$$

ou encore sur un échantillon

$$\hat{\theta} = \left\{ \theta \mid \sum_{i=1}^n \psi(x_i, \theta) = 0 \right\} \quad (4.9)$$

L'estimateur par maximum de vraisemblance est un cas particulier de M-estimateur pour lequel  $\rho(x, \theta) = -\ln(f(x, \theta))$ , avec  $f$  une densité paramétrique choisie<sup>4</sup>.

L'intérêt de tels estimateurs est que le calcul de la fonction d'influence est explicite<sup>5</sup>.

$$\begin{aligned} \text{IF}(x; F_\theta, T) &= \frac{\psi(x, T(F_\theta))}{-\int_{\mathcal{X}} \frac{\partial \psi(y, T(F_\theta))}{\partial \theta} dF_\theta(y)} \\ &= \left( \begin{array}{cc} -\int_{\mathcal{X}} \frac{\partial \psi_{\theta_1}}{\partial \theta_1} dF_\theta(y) & -\int_{\mathcal{X}} \frac{\partial \psi_{\theta_2}}{\partial \theta_1} dF_\theta(y) \\ -\int_{\mathcal{X}} \frac{\partial \psi_{\theta_2}}{\partial \theta_1} dF_\theta(y) & -\int_{\mathcal{X}} \frac{\partial \psi_{\theta_2}}{\partial \theta_2} dF_\theta(y) \end{array} \right)^{-1} \begin{pmatrix} \psi_{\theta_1} \\ \psi_{\theta_2} \end{pmatrix} \\ &:= \mathcal{A}^{-1}(\theta) \psi_\theta \end{aligned} \quad (4.10)$$

où  $\psi_{\theta_i} = \frac{\partial \rho}{\partial \theta_i}$ . La seconde égalité est donnée pour une loi à deux paramètres, cas couramment rencontré, mais on pourrait généraliser pour  $k$  paramètres de la même façon. De cette manière, la fonction d'influence de n'importe quel M-estimateur peut être calculée explicitement, et notamment pour l'estimateur par maximum de vraisemblance, ce qui est présenté dans la section suivante.

En plus de leur fonction d'influence, on connaît la variance asymptotique des M-estimateurs. Ainsi,  $\sqrt{n}(T_n - T(F_\theta))$  est asymptotiquement de moyenne nulle et de variance<sup>6</sup>

$$\mathbb{V}(F_\theta, T) = \int_{\mathcal{X}} \text{IF}(x; F_\theta, T)^2 F_\theta(dx) \quad (4.11)$$

## 4.2.2 Fonction d'influence de l'estimateur par maximum de vraisemblance

L'estimateur par maximum de vraisemblance (MLE, *Maximum Likelihood Estimator*) est communément utilisé pour estimer les paramètres d'une loi sur un échantillon donné. C'est en particulier le cas pour la modélisation de la loi de sévérité des pertes collectées au titre de risque opérationnel. La détermination de la fonction d'influence montre néanmoins que cet estimateur n'est pas B-robuste.

Le MLE est un M-estimateur pour lequel  $\rho(x, \theta) = -\ln(f(x, \theta))$  de fonction de score

$$\psi_\theta(x, \theta) = \frac{\partial \rho(x, \theta)}{\partial \theta} = -\frac{1}{f(x, \theta)} \frac{\partial f(x, \theta)}{\partial \theta},$$

de dérivée

$$\psi'_\theta(x, \theta) = \frac{-\frac{\partial^2 f(x, \theta)}{\partial \theta^2} f(x, \theta) + \left(\frac{\partial f(x, \theta)}{\partial \theta}\right)^2}{[f(x, \theta)]^2}$$

Ainsi pour une distribution  $f$  à  $k$  paramètres, il suffit de calculer

$$\frac{\partial f}{\partial \theta_i} \quad i \in \llbracket 1, k \rrbracket, \quad \frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \quad i, j \in \llbracket 1, k \rrbracket^2$$

4. Les M-estimateurs font partie d'une famille encore plus vaste, les extremum estimateurs qui consistent à maximiser une fonction objectif. L'estimateur par méthode des moments généralisée appartient également à la famille des extremum estimateurs, mais ce n'est pas un M-estimateur.

5. La preuve est donnée en Annexe A.

6. La preuve est également donnée en Annexe A.

afin de pouvoir obtenir la fonction d'influence associée.

Cette définition est explicitée ci-après pour la distribution log-normale, distribution la plus utilisée pour modéliser la sévérité des pertes liées au risque opérationnel.

### Loi log-normale

Pour obtenir la fonction d'influence, il est nécessaire de disposer de la fonction de score et de la matrice  $\mathcal{A}(\theta)$  <sup>7</sup>.

Le vecteur de score est donné par

$$\psi_\theta = \begin{pmatrix} \frac{\mu - \ln x}{\sigma^2} \\ \frac{1}{\sigma} - \frac{(\mu - \ln x)^2}{\sigma^3} \end{pmatrix} \quad (4.12)$$

et la matrice  $\mathcal{A}(\theta)$  par

$$\mathcal{A}(\theta) = \begin{pmatrix} -\frac{1}{\sigma^2} & 0 \\ 0 & -\frac{2}{\sigma^2} \end{pmatrix} \quad (4.13)$$

**Remarque :** Dans le cas du maximum de vraisemblance, la matrice  $\mathcal{A}(\theta)$  est aussi appelée Information de Fisher, et notée  $\mathcal{I}(\theta)$  <sup>8</sup>.

Le calcul de l'inverse est élémentaire. On obtient alors la fonction d'influence pour le MLE avec une densité log-normale suivante

$$\text{IF}(x; \theta, T) = \begin{pmatrix} \ln x - \mu \\ \frac{(\ln x - \mu)^2 - \sigma^2}{2\sigma} \end{pmatrix} \quad (4.14)$$

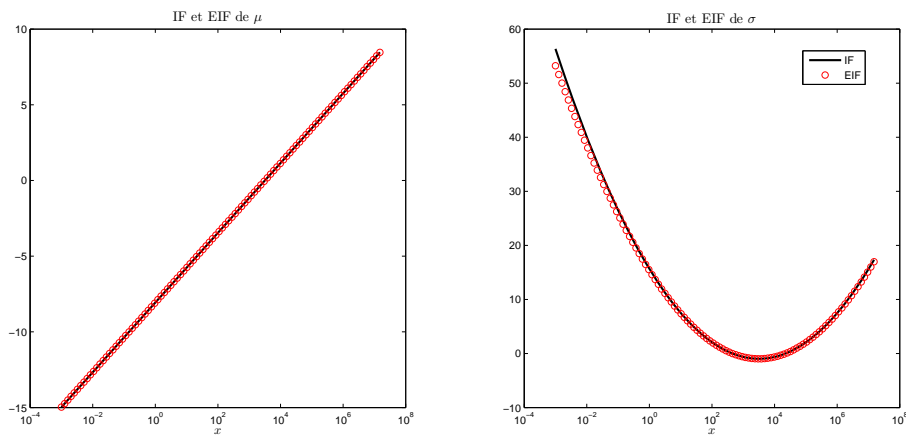


Figure 4.2 – Fonctions d'influence empirique et théorique d'une loi lognormale  $\mathcal{LN}(8, 2)$  non tronquée

7. Ceux-ci se calculent à partir des dérivées première et seconde de la densité, le détail étant donné en Annexe A.

8. Généralement écrite au premier ordre, on a l'égalité  $\mathcal{I}(\theta) = \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta} \ln f(x, \theta) \right)^2 \right] = \mathbb{E} \left[ -\frac{\partial^2}{\partial \theta^2} \ln f(x, \theta) \right]$

On retrouve ici un résultat bien connu sur la robustesse de l'estimateur par maximum de vraisemblance, à savoir que **l'estimateur ML n'est pas B-robuste**. La figure 4.2 montre la non robustesse de l'estimateur, puisque les fonctions d'influence sont clairement non bornées, puisque divergentes lorsque  $x \rightarrow +\infty$  et  $x \rightarrow 0$ . Ceci pose un problème de cohérence puisque l'ajout d'une seule perte peut rendre impossible l'exploitation des résultats obtenus via cette méthode d'estimation.

**Ceci est d'autant plus important que ce phénomène se produit aussi bien pour des pertes extrêmement grandes que pratiquement nulles.**

Une seconde chose intéressante est de noter l'adéquation quasi parfaite entre les fonctions d'influence empiriques et théoriques, alors que pour l'obtention des valeurs empiriques, l'estimation ne s'est faite que pour un échantillon de 250 données. Néanmoins, la fonction d'influence empirique a tendance à être légèrement sous-évaluée pour des valeurs proches de 0. Même si la tendance observée reste la même, le caractère infini de la fonction d'influence quand  $x$  tend vers 0 ne pourra être observé pour la version empirique qui ne donne que des valeurs finies.

**Ceci permet en tout cas de légitimer l'utilisation de fonctions d'influence empirique à des estimateurs pour lesquels il est impossible de déterminer une expression explicite de sa valeur théorique.**

**Cas tronqué.** Il est intéressant d'étendre ces résultats au cas d'une loi log-normale tronquée, étant donnée la situation réelle de collecte des données à partir d'un certain seuil  $H$ . On utilise, dans le cas de données tronquées, la densité suivante :

$$\tilde{f}(x, \theta) = \frac{f(x, \theta)}{1 - F(H, \theta)}$$

qui est bien une densité sur  $[H, +\infty[$ . Sa fonction de répartition se définit donc par

$$\tilde{F}(x, \theta) = \frac{F(x, \theta) - F(H, \theta)}{1 - F(H, \theta)}$$

On a donc  $\rho(x, \theta) = -\ln(f(x, \theta)) + \ln(1 - F(H, \theta))$ , dont la fonction de score est

$$\psi_\theta(x, \theta) = -\frac{1}{f(x, \theta)} \frac{\partial f(x, \theta)}{\partial \theta} - \frac{1}{1 - F(H, \theta)} \frac{\partial F(H, \theta)}{\partial \theta} \quad (4.15)$$

de dérivée

$$\psi'_\theta(x, \theta) = \frac{1}{[f(x, \theta)]^2} \left( \left( \frac{\partial f(x, \theta)}{\partial \theta} \right)^2 - \frac{\partial^2 f(x, \theta)}{\partial \theta^2} f(x, \theta) \right) - \frac{1}{[1 - F(H, \theta)]^2} \left( \left( \frac{\partial F(H, \theta)}{\partial \theta} \right)^2 + \frac{\partial^2 F(H, \theta)}{\partial \theta^2} (1 - F(H, \theta)) \right) \quad (4.16)$$

Ce qui permet d'obtenir la fonction d'influence de la loi log-normale tronquée<sup>9</sup> :

$$\begin{aligned} \text{IF}(x; \theta, T) = & \frac{-\frac{1}{f(x, \theta)} \frac{\partial f(x, \theta)}{\partial \theta} - \frac{1}{1 - F(H, \theta)} \frac{\partial F(H, \theta)}{\partial \theta}}{\left[ -\frac{1}{1 - F(H, \theta)} \int_{\mathcal{Y}} \frac{1}{f(y, \theta)} \left( \left( \frac{\partial f(y, \theta)}{\partial \theta} \right)^2 - \frac{\partial^2 f(y, \theta)}{\partial \theta^2} f(y, \theta) \right) dy \right.} \\ & \left. + \frac{1}{[1 - F(H, \theta)]^2} \left( \left( \frac{\partial F(H, \theta)}{\partial \theta} \right)^2 + \frac{\partial^2 F(H, \theta)}{\partial \theta^2} (1 - F(H, \theta)) \right) \right]} \end{aligned} \quad (4.17)$$

À partir de la fonction d'influence pour une loi log-normale tronquée, on peut obtenir la représentation graphique présentée en figure 4.3 en effectuant les calculs via Matlab<sup>10</sup>.

9. Le détail des calculs permettant d'obtenir les formules exactes des éléments de  $\mathcal{A}(\theta)$  et de ceux de  $\psi_\theta$  est donné en Annexe dans le paragraphe A.A. Ainsi, les coefficients de la matrice  $\mathcal{A}(\theta)$  sont donnés par les formules (A.13), (A.14) et (A.15), ceux de  $\psi_\theta$  par (A.12).

10. Cependant, étant donnée la forme de la matrice  $\mathcal{A}(\theta)$ , il paraît difficile de calculer l'expression analytique de la fonction d'influence car il faudrait être capable de calculer  $\mathcal{A}^{-1}(\theta)$

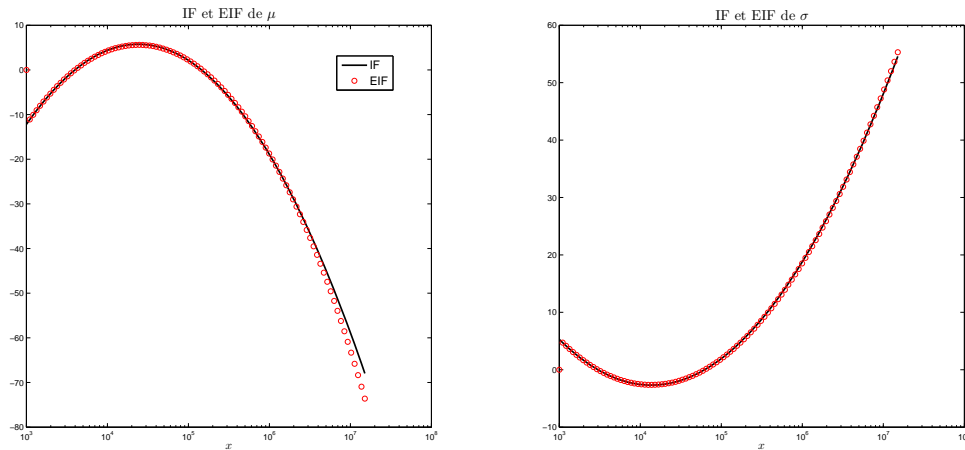


Figure 4.3 – Fonctions d'influence empirique et théorique d'une loi lognormale  $\mathcal{LN}(8, 2)$  tronquée

À la lecture des graphes, on note plusieurs différences majeures entre les résultats sur données tronquées et non tronquées :

- La première est le **changement du sens de variation de la fonction d'influence du paramètre  $\mu$** . Alors que celle-ci est croissante sur données non tronquées, elle est décroissante à partir d'une certaine valeur de  $x$ . En revanche, la fonction d'influence du paramètre  $\sigma$  garde le même sens de variation. Cela signifie que la relation entre les deux paramètres a changé. Pour un modèle non tronqué, l'ajout d'une grande perte de montant  $x$  a pour effet l'augmentation des deux paramètres. Ceci est ici complètement différent puisque pour de grandes valeurs, la fonction d'influence de  $\mu$  devient négative, ce qui veut dire qu'une contamination arbitrairement grande diminue la valeur du paramètre  $\mu$ <sup>11</sup>.
- **La non robustesse de la méthode est de nouveau observable**. Les fonctions d'influence des deux paramètres sont en effet divergentes lorsque  $x \rightarrow \infty$  : celle de  $\mu$  tend vers  $-\infty$  et celle de  $\sigma$  vers  $+\infty$ . Néanmoins, bien qu'un extremum local soit atteint en  $H$ , les valeurs des fonctions d'influence en ce point sont élevées mais non infinies. Ainsi, l'ajout d'une perte arbitrairement faible n'entraîne pas une estimation complètement disproportionnée, mais contribue à la perturber notablement. Ceci est différent du modèle non tronqué, puisqu'on rappelle que le caractère non B-robuste s'observe également lorsque  $x$  tend vers 0.
- On observe également que **la troncature introduit une corrélation entre les paramètres** puisque les éléments extra-diagonaux de la matrice  $\mathcal{A}^{-1}(\theta)$  sont non nuls.

Malgré la bonne adéquation entre les fonctions d'influence empirique et théorique présentée figure 4.3, on peut noter une plus grande instabilité de la version empirique. Bien qu'il ne faille qu'une centaine de points pour assurer une bonne adéquation sur un modèle non tronqué, un millier de points sont ici nécessaires. Ceci est probablement dû à la corrélation entre les paramètres introduite par la troncature au seuil  $H$ .

Afin de mettre en avant ce phénomène, on définit les courbes "critiques" pour une taille d'échantillon donnée, comme les deux courbes s'écartant le plus de la courbe théorique au sens de la norme infinie sur un ensemble de 100 courbes. Pour différentes tailles d'échantillon, la comparaison entre ces courbes "critiques" pour des échantillons non tronqués (figure 4.4) et des échantillons tronqués (figure 4.5) met

11. Ceci est d'autant plus contrintuitif que ce paramètre correspond au paramètre de localisation de la loi log-normale,  $\sigma$  étant le paramètre d'échelle.

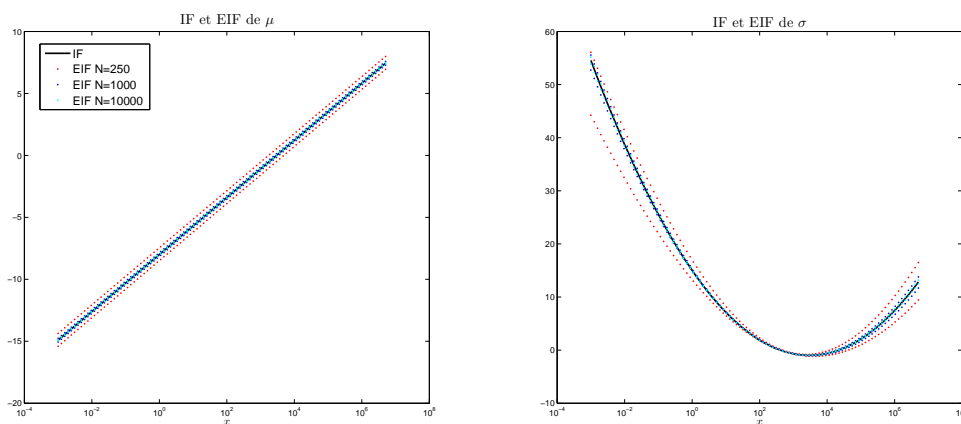


Figure 4.4 – Fonctions d'influence empiriques "critiques" pour différentes tailles d'échantillon d'une loi lognormale  $\mathcal{LN}(8, 2)$  non tronquée

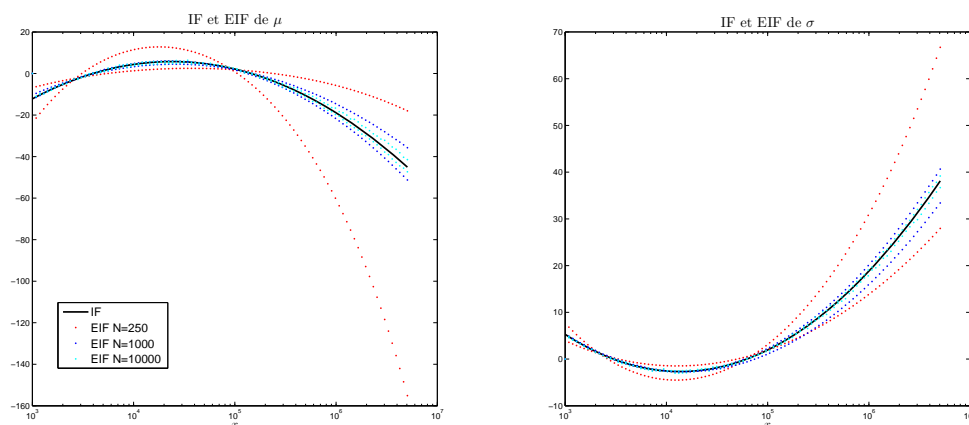
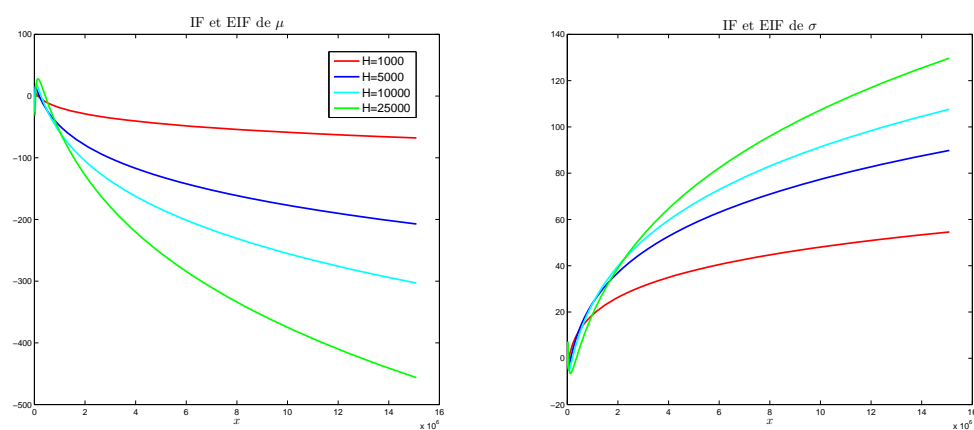


Figure 4.5 – Fonctions d'influence empiriques "critiques" pour différentes tailles d'échantillon d'une loi lognormale  $\mathcal{LN}(8, 2)$  tronquée au seuil  $H = 1000$  euros

en exergue cette variabilité.

À la vue de ces graphiques, **il est préférable par la suite d'utiliser des échantillons de 1000 à 5000 données en cas de troncature** afin de s'assurer que le tracé de la fonction d'influence empirique coïncide bien en moyenne avec le tracé théorique. De plus, il est à noter que **tous ces résultats sont d'autant plus marqués que le seuil de collecte  $H$  est grand**, comme en témoigne le graphique donné figure 4.6.

D'après ces résultats, pour un échantillon de pertes donné, l'ajout d'une seule nouvelle perte peut augmenter le paramètre d'échelle  $\sigma$  et diminuer le paramètre de localisation  $\mu$ , et ce de façon arbitrairement importante. De plus, comme le calcul de la charge en capital dépend des paramètres de la loi, et principalement du paramètre ajustant l'épaisseur de la queue (ici  $\sigma$ ) puisque la méthode de calcul fait intervenir des quantiles extrêmement élevés, une augmentation démesurée de ce paramètre peut avoir des conséquences considérables sur le montant de cette charge en capital. Au vu de ces constats, il semble nécessaire d'implémenter une méthode d'estimation robuste concernant le calcul du capital réglementaire.

Figure 4.6 – Fonctions d'influence théorique pour différents seuils de collecte d'une loi lognormale  $\mathcal{LN}(8, 2)$



## 5 Étude d'estimateurs robustes

---

Sous des conditions idéales d'indépendance et d'homogénéité des données ainsi que de spécification du modèle statistique, l'estimation par maximum de vraisemblance est reconnue comme étant la méthode optimale. Cependant, ce cas idéalisé n'est jamais observé dans la pratique sur des données collectées pour le compte du risque opérationnel. En raison de sa non robustesse mise en évidence dans la partie précédente, disposer d'une méthode robuste qui ne souffre pas des restrictions mentionnées en début de paragraphe apparaît comme nécessaire.

Plusieurs pistes peuvent être envisagées afin d'améliorer la robustesse d'un modèle, en agissant sur le traitement des données en elles-mêmes ou encore sur la méthode d'estimation directement. On peut en effet envisager de :

- (1) Procéder à un "nettoyage" préalable des données en éliminant les points marginaux du modèle, puis utiliser les méthodes classiques d'inférence statistique type MLE ou GMM. On peut par exemple avoir recours à un  $\alpha$ -trimming, qui correspond à la suppression du pourcentage  $\alpha$  des données les plus élevées. D'autres méthodes sont suggérées afin de ne pas supprimer de façon drastique les données, mais plutôt de se référer à un avis d'expert ou bien à des méthodes de détection des points aberrants.
- (2) Implémenter des méthodes d'estimation conçues de manière à vérifier la propriété de robustesse.

Ce mémoire se concentre uniquement sur les méthodes d'estimation robustes, bien que combiner ces deux options est envisageable.

Il existe plusieurs estimateurs robustes dans la littérature, parmi lesquels l'estimateur OBRE, l'estimateur de Cramér-Von-Mises (CvM) ou encore l'estimateur par quantile-distance (QD). Ces trois estimateurs sont définis dans cette partie, et leurs avantages et inconvénients sont discutés en termes d'estimation et d'implémentation. Les propriétés de robustesse de ces estimateurs sont, quant à elles, abordées dans la partie suivante.

### Propriétés d'un estimateur

Avant d'étudier en détail les caractéristiques d'estimateurs robustes, on définit ici les propriétés classiques d'un estimateur, telles que le biais, la convergence et l'efficacité d'un estimateur. Ceci permet alors de caractériser et de comparer entre elle chaque méthode d'estimation présentée dans la suite de cette partie.

**Convergence.** Un estimateur  $\hat{\theta}_n$  du paramètre  $\theta$  d'un échantillon de taille  $n$  est dit convergent s'il converge en probabilité vers  $\theta$ , ce qui s'écrit

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}[\lvert \hat{\theta}_n - \theta \rvert \geq \varepsilon] = 0$$

Si de plus cette convergence est presque sûre, alors l'estimateur est dit fortement convergent

$$\lim_{n \rightarrow \infty} \mathbb{P}[\hat{\theta}_n = \theta] = 1$$

En pratique, une condition suffisante pour assurer la convergence est que

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\theta}_n] = \theta \quad \text{et} \quad \lim_{n \rightarrow \infty} \mathbb{V}[\hat{\theta}_n] = 0$$

où  $\mathbb{E}[\hat{\theta}_n]$  et  $\mathbb{V}[\hat{\theta}_n]$  désignent respectivement l'espérance et la variance de  $\hat{\theta}_n$ .

**Biais.** En fonction de l'échantillon, l'estimateur va plus ou moins fluctuer autour de sa vraie valeur  $\theta$ . Le biais d'un estimateur permet de connaître l'écart moyen entre  $\hat{\theta}_n$  et  $\theta$ . Cette quantité se calcule de la manière suivante

$$\mathbb{B}[\hat{\theta}_n] = \mathbb{E}[\hat{\theta}_n] - \theta$$

Un estimateur est dit sans biais si  $\forall n, \mathbb{E}(\hat{\theta}_n) = \theta$ . On dit également qu'un estimateur est asymptotiquement sans biais lorsque

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\theta}_n] = \theta$$

Dans la suite, on utilise surtout le biais relatif, défini par  $\mathbb{B}_{rel}(\hat{\theta}_n) = \mathbb{B}[\hat{\theta}_n]/\theta$ . On peut ainsi définir la convergence de l'estimateur par

$$\lim_{n \rightarrow \infty} \mathbb{B}_{rel}[\hat{\theta}_n] = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} \sqrt{\mathbb{V}[\hat{\theta}_n]} = 0$$

Ce sont ces deux critères qui sont utilisés dans la suite pour qualifier la convergence des estimateurs.

**Efficacité.** Un estimateur  $\hat{\theta}_n$  est dit efficace s'il est de variance minimale parmi tous les estimateurs de  $\theta$  ayant même moyenne.

On parlera également d'écart-type relatif :

$$\sigma_{rel}[\hat{\theta}_n] = \frac{\sqrt{\mathbb{V}[\hat{\theta}_n]}}{\mathbb{E}[\hat{\theta}_n]}$$

## Mise en place d'un environnement simulé

S'assurer des propriétés "classiques" de l'estimateur revient donc à vérifier ses qualités en termes de convergence. Afin d'avoir une étude la plus complète possible, nous présentons ces résultats à partir de données simulées provenant de quatre jeux de paramètres différents.

En effet, dans la pratique, les données réelles récoltées proviennent des différentes entités et de différentes catégories de risque. Les profils de risque associés sont donc hétérogènes d'une cellule de risque à l'autre. C'est pourquoi on choisit de travailler sur quatre jeux de paramètres différents d'une loi log-normale, représentatifs de l'étendue des panels de risque rencontrés dans la pratique.

Les densités de probabilité associées sont représentées sur la figure 5.1 :

- (8, 2) représentant un profil de haute sévérité à queue de distribution épaisse
- (8, 1) représentant un profil de haute sévérité à queue de distribution plus fine
- (6, 2) représentant un profil de faible sévérité à queue de distribution épaisse
- (6, 1) représentant un profil de faible sévérité à queue de distribution plus fine

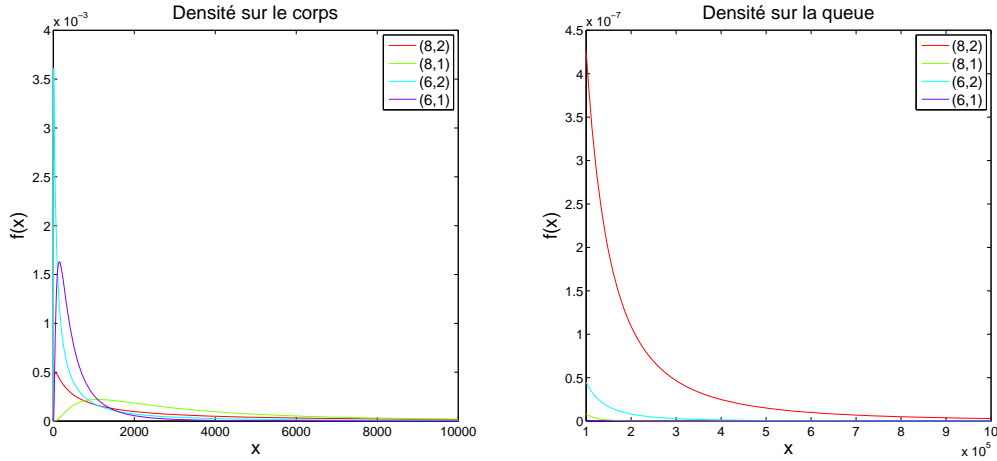


Figure 5.1 – Représentation de la densité de la loi log-normale selon les quatre jeux de paramètres sur le corps  $[0, 10^4]$  et sur la queue  $[10^5, 10^6]$

## 5.1 Estimateur OBRE

L'application de l'estimateur OBRE au risque opérationnel a été l'objet de quelques travaux récents, notamment ceux d'Opdyke et Cavallo ([15] et [16]). L'appellation anglophone "OBRE" signifie *Optimally Bias-Robust Estimator*. Cet estimateur est en fait une version du MLE sous contraintes, transformée de façon à assurer sa B-robustesse. Il est en outre construit de façon à être l'estimateur à fonction d'influence bornée avec une efficacité maximale, d'où le terme *Optimal*.

Sur des données simulées idéales (c'est-à-dire *iid* issues d'un modèle  $F_\theta$  explicite), l'estimateur par maximum de vraisemblance est reconnu pour être optimal au sens de l'efficacité (sa variance asymptotique est minimale). Cependant, sur données réelles, il peut même arriver que l'estimateur robuste OBRE soit plus efficace<sup>1</sup>.

### 5.1.1 Définition

Pour un échantillon donné  $(x_1, \dots, x_n)$ , l'estimateur OBRE  $\hat{\theta}$  du paramètre  $\theta$  est donné par

$$\hat{\theta} = \left\{ \theta \mid \sum_{i=1}^n \psi(x_i, \theta) = 0 \right\}$$

où  $\psi(x_i, \theta)$  est défini par  $\psi(x_i, \theta) = A(\theta)(s(x_i, \theta) - a(\theta))W_c(x_i, \theta)$  avec  $s(x_i, \theta) = \frac{\partial \ln f(x_i, \theta)}{\partial \theta}$  la fonction de score classique et  $W_c(x_i, \theta)$  une fonction de pondération donnée par

$$W_c(x_i, \theta) = \min \left( 1, \frac{c}{\|A(\theta)(s(x_i, \theta) - a(\theta))\|} \right)$$

Les quantités  $A(\theta)$  et  $a(\theta)$  de dimensions respectives  $\dim(\theta) \times \dim(\theta)$  et  $\dim(\theta) \times 1$  sont tels que les deux équations suivantes sont vérifiées

$$\mathbb{E} [\psi(x, \theta) \times \psi(x, \theta)'] = I \quad (5.1)$$

$$\mathbb{E} [\psi(x, \theta)] = 0 \quad (5.2)$$

1. Voir pour plus de détails, les articles [15], [16], [4], [21].

Comme l'estimateur OBRE est un M-estimateur, sa fonction d'influence est définie par l'équation (4.10). Elle est donc linéaire en  $\psi_\theta$ . Ainsi, si  $\psi_\theta$  est non bornée alors la fonction d'influence est non bornée également, ce qui est le cas de l'estimateur ML. Dans le cas de l'OBRE, on s'arrange pour que la quantité  $\psi(x, \theta)$  soit bornée. C'est le rôle assuré par la fonction de pondération  $W_c$ , comprise entre 0 et 1. Elle permet d'affecter un poids à chaque donnée, ce poids spécifiant la concordance entre ces données et le modèle paramétrique estimé. Ainsi, si une perte est pondérée par un poids quasi nul, c'est qu'elle est considérée comme totalement divergente du modèle sous-jacent, et inversement si ce poids est proche de 1.

$A(\theta)$  et  $a(\theta)$  peuvent être vus comme des multiplicateurs de Lagrange pour lesquels les équations (5.1) et (5.2) permettent de satisfaire les contraintes de B-robustesse et de consistance de Fisher<sup>2</sup> que doit vérifier l'estimateur OBRE.

La fonction de pondération  $W_c$  est dépendante du paramètre  $c$ . C'est un paramètre de modulation permettant un arbitrage entre robustesse et efficacité. Ce paramètre, choisi par l'utilisateur, doit obligatoirement vérifier  $\sqrt{\dim(\theta)} \leq c \leq \infty$ . Plus  $c$  est proche de son minimum, plus l'estimation gagne en robustesse mais perd en efficacité, et inversement lorsque  $c$  devient grand, l'estimateur OBRE coïncide avec l'estimateur ML.

Il existe plusieurs variantes de l'estimateur OBRE, mais celle décrite plus haut, appelée estimateur OBRE "normalisé", est reconnue pour être la plus stable numériquement.

### 5.1.2 Algorithme

Un algorithme permet de trouver l'estimation  $\hat{\theta}$  satisfaisant les équations (5.1) et (5.2) de manière itérative<sup>3</sup>. Cet algorithme<sup>4</sup> en quatre étapes est établi comme suit :

#### 1. Etape 1

- Choix du seuil de précision  $\eta$ , inhérent à la convergence de l'algorithme.
- Choix d'une valeur initiale  $\theta_{\text{init}}$ , généralement la valeur obtenue par maximum de vraisemblance en intégrant ou non le seuil  $H$ .
- Initialisation des valeurs de  $A(\theta)$  et  $a(\theta)$  respectivement à  $\sqrt{[J(\theta)^{-1}]'}$  et 0, avec

$$J(\theta) = \int s(x, \theta) s(x, \theta)' dF_\theta(x)$$

la matrice d'information de Fisher, que l'on retrouve plus communément écrite sous la forme  $\mathbb{E} \left[ \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 \right]$ .

#### 2. Etape 2<sup>5</sup>

On obtient les nouvelles valeurs de  $A$  et  $a$  en résolvant le système suivant :

$$\begin{cases} A(\theta)' A(\theta) = M_2^{-1}(\theta) \\ a(\theta) = \frac{\int s(x, \theta) W_c(x, \theta) dF_\theta(x)}{\int W_c(x, \theta) dF_\theta(x)} \end{cases} \quad (5.3)$$

$$(5.4)$$

avec  $M_k(\theta) = \int (s(x, \theta) - a(\theta))(s(x, \theta) - a(\theta))' W_c(x, \theta)^k dF_\theta(x)$  pour  $k = 1, 2$ .

2. C'est-à-dire que l'estimateur vérifie  $T(F_\theta) = \theta$

3. En utilisant une méthode de convergence de type Newton-Raphson.

4. Il est présenté dans les articles [15], [4] et [21].

5. À cette étape, l'écriture de  $W_c(x, \theta)$  ne dépend des valeurs  $A(\theta)$  et  $a(\theta)$  calculées à l'étape précédente (Etape1 si initialisation, Etape2 précédente sinon).

### 3. *Etape 3*

On calcule  $M_1$  avec les nouvelles valeurs de  $a(\theta)$  et  $A(\theta)$  calculées à l'*Etape 2*, puis on obtient la variation de paramètre  $\theta$  suivante

$$\Delta\theta = M_1^{-1}(\theta) \left[ \frac{1}{n} \sum_{i=0}^n [s(x_i, \theta) - a(\theta)] W_c(x_i, \theta) \right] \quad (5.5)$$

*Condition d'arrêt* : l'algorithme s'arrête lorsque  $\left| \frac{\Delta\theta_j}{\theta_j} \right| < \eta$ , pour  $j = 1, 2$

Sinon, on pose :

$$\theta \rightarrow \theta - \frac{\Delta\theta}{\gamma} \quad (5.6)$$

et l'algorithme reprend à l'*Etape 2*.

*Remarque :*

- Les équations écrites sous forme d'intégrale matricielle correspondent simplement à l'intégrale de chacun des éléments de la matrice.
- Le paramètre  $\gamma$  permet de diminuer si besoin le pas  $\Delta\theta$  à chaque itération. Son utilité est présentée dans la suite.

Dans la suite, nous explicitons certains des calculs mentionnés dans l'algorithme dans le cadre d'une loi lognormale tronquée, comme nous l'avons fait pour le calcul des fonctions d'influence. En effet, nous rappelons que cette loi est la plus communément utilisée pour la modélisation de la sévérité. Ensuite, nous exposons les difficultés rencontrées lors de l'implémentation de cet algorithme, notamment en raison de la complexité des calculs.

### 5.1.3 Calculs pour une loi log-normale tronquée

Reprenons  $f$  la densité d'une loi log-normale tronquée dont les propriétés ont déjà été définies dans le paragraphe 4.2.2. La fonction de score, appelée ici  $s(x, \theta)$ , est celle d'une loi log-normale tronquée, et son calcul est donné en Annexe.

Concernant les éléments de la matrice d'information de Fisher utilisée afin d'initialiser la valeur de  $A(\theta)$ , ceux-ci sont calculables. Néanmoins, leur écriture étant très lourde, le résultat est donné en Annexe A.

Malgré la complexité des éléments de la matrice  $J(\theta)$ , ceux-ci ont l'avantage d'être explicites, ce qui n'est pas le cas des autres calculs que l'on trouve dans l'algorithme. En effet les calculs des matrices  $M_k$  (2) ainsi que celui de  $a(\theta)$  (5.4) font intervenir des intégrales dont la quantité  $W_c$  intervient dans l'intégrande. Cependant, étant donnée la forme de cette fonction, il paraît impossible d'en donner une expression explicite. C'est pourquoi on doit avoir recours à des approximations numériques sous Matlab. De même, bien qu'il existe des algorithmes pour obtenir la décomposition de Cholesky d'une matrice, obtenir celle de  $A(\theta)$  (5.3) de façon formelle semble difficile.

### Problématiques numériques

Parmi ces deux problématiques d'ordre numérique, obtenir la décomposition de Cholesky d'une matrice se fait très facilement sous MATLAB. C'est donc l'approximation numérique des intégrales qui doit faire

l'objet d'une attention particulière. Dans la littérature, on peut trouver des approches permettant de contourner en partie ce problème.

Comme cela est présenté en détail en Annexe B, on est amené à travailler sur deux versions différentes d'implémentation de l'algorithme de l'estimateur OBRE. La première faisant intervenir la fonction de répartition empirique associée aux données, est appelée *méthode empirique*, la seconde utilisant des méthodes d'approximations numériques d'intégrales, est appelée *méthode reparamétrée*.

Avant de s'atteler aux propriétés de robustesse ainsi qu'à l'étude de la sensibilité des paramètres d'entrée, on doit s'assurer que l'estimateur vérifie les propriétés de convergence énoncés en début de partie dans le paragraphe 5.5. Il est intéressant de comparer ces propriétés pour les deux méthodes proposées afin de sélectionner la meilleure.

La comparaison entre ces deux approches est également proposée en Annexe B. Et à la vue de ces résultats, il est préférable de n'utiliser par la suite que la version empirique de l'algorithme d'estimation OBRE. Ses propriétés techniques et la sensibilité de ses paramètres sont présentés dans la prochaine partie.

### Sensibilité des paramètres

Il est également intéressant de tester la sensibilité des estimations au niveau de précision  $\eta$  ainsi qu'aux paramètres initiaux. Les résultats détaillés de cette analyse sont présentés en Annexe B et aboutit aux choix exposés ci-après.

**Sensibilité au paramètre  $\eta$ .** Nous choisissons  $\eta = 0,001$  afin de garder en plus une marge de sécurité, la convergence étant assurée pour les quatre modèles pour  $\eta$  de l'ordre de 0,05.

**Sensibilité aux paramètres d'initialisation.** Dans l'algorithme, l'initialisation des paramètres par la valeur obtenue par maximum de vraisemblance sur les données non tronquées. Cela signifie que, même si le seuil est de  $H = 1000$ , cette initialisation se fait pour un seuil  $H = 0$ . En effet, cela permet d'obtenir des paramètres d'ordre de grandeur convenable, et de passer outre les estimations parfois aberrantes que peut fournir le maximum de vraisemblance sur données tronquées (un  $\mu$  négatif et un  $\sigma$  très élevé).

**En somme, rien n'invalide cette façon d'initialiser. Dans la suite, l'initialisation se fait donc sur données non modifiées et non tronquées.**

Les propriétés de convergence de l'estimateur OBRE sont vérifiées et le calibrage des paramètres est établi. Les propriétés de robustesse sont abordées dans le chapitre suivant.

## 5.2 Estimateur de Cramér-von-Mises

De façon générale, le but de la méthode d'estimation de Cramér-von-Mises est de minimiser une distance entre une fonction de répartition  $G$  donnée et un ensemble de fonctions de répartition paramétriques  $\mathcal{F}(\theta)$ , tel que  $F_\theta \in \mathcal{F}(\theta)$ , pour une mesure  $\nu$ .

La distance utilisée pour trouver ce minimum est

$$d_{CvM}(F, G) = \int_{\mathcal{X}} (F(x) - G(x))^2 \nu(dx) \quad (5.7)$$

Mathématiquement, l'estimateur CvM se traduit par

$$T(F_\theta) = \arg \min_{\theta} d_{CvM}(F_\theta, G) \quad (5.8)$$

Il s'agit d'un extremum estimateur, puisqu'on cherche à minimiser une fonction objectif.

On utilise cet estimateur afin de trouver le paramètre  $\theta$  qui minimise la distance de Cramér-von-Mises entre la fonction de répartition empirique de l'échantillon et la fonction de répartition de la famille paramétrique  $F_\theta$ . Ainsi, dans cette configuration,  $G = F_n$ ,  $\nu(dx) = F_\theta(dx)$ . Si de plus on se place dans un cadre où les données sont tronquées au seuil  $H$ , on fait intervenir la densité tronquée  $\widetilde{F}_\theta$  et cet estimateur devient

$$T(\widetilde{F}_\theta) = \arg \min_{\theta} \int_H^{+\infty} \left( \widetilde{F}_\theta(x) - F_n(x) \right)^2 d\widetilde{F}_\theta(x)$$

ou encore

$$T(\widetilde{F}_\theta) = \arg \min_{\theta} d_{CvM}(\widetilde{F}_\theta, F_n)$$

Puisque l'on travaille avec les fonctions de répartition, deux options sont envisageables pour prendre en compte le seuil  $H$  dans l'estimation.

### 5.2.1 Prise en compte normale du seuil

On entend par prise en compte "normale" du seuil le fait que la minimisation de la distance par rapport au paramètre  $\theta$  se fasse de manière usuelle, en faisant intervenir une distribution paramétrique tronquée à gauche. Ainsi, si on note  $(x_{(1)}, \dots, x_{(n)})$  l'échantillon ordonné par ordre croissant, on considère que  $x_{(1)}$  est la plus petite de nos pertes, en faisant abstraction des pertes inférieures au seuil  $H$ . La fonction de répartition empirique a alors la même expression qu'énoncé dans l'équation (4.1), et que  $F_n(x_{(1)}) = \frac{1}{n}$ . Il n'y a donc aucune prise en compte des hypothétiques valeurs en-dessous du seuil dans l'écriture de la fonction de répartition empirique.

En pratique, sur un échantillon  $(x_1, \dots, x_n)$ , on discrétise l'équation (??) sur cet échantillon pour obtenir l'estimateur CvM

$$T(\widetilde{F}_\theta) = \arg \min_{\theta} \sum_{i=1}^n \left( \widetilde{F}_\theta(x_i) - F_n(x_i) \right)^2$$

En développant l'écriture de  $\widetilde{F}_\theta$

$$T(\widetilde{F}_\theta) = \arg \min_{\theta} \sum_{i=1}^n \left( \frac{F_\theta(x_i) - F_\theta(H)}{1 - F_\theta(H)} - \frac{i}{n} \right)^2 \quad (5.9)$$

### Étude de convergence

On se place ici dans les mêmes conditions que l'étude déjà effectuée pour l'estimateur OBRE. On effectue donc des tests de convergence pour les quatre jeux de paramètres de loi log-normale, en étudiant à la fois le biais relatif et l'écart-type relatif pour les deux paramètres  $\mu$  et  $\sigma$ . On procède toujours pour 1000 simulations.

D'après les graphes fournis figure 5.2, on peut conclure que l'estimateur est convergent puisque le biais relatif et l'écart-type relatif tendent vers 0 lorsque  $n$  tend vers l'infini, quelque soit le profil de risque.

Néanmoins, pour de petites tailles d'échantillon, les biais observés sur les profils  $\mathcal{LN}(6,1)$  et  $\mathcal{LN}(6,2)$  sont très importants. De plus, quelque soit le profil de risque, le biais relatif est toujours plus important que celui observé avec l'emploi de la méthode OBRE.

En outre, l'écart-type relatif du paramètre  $\mu$  est explosif pour les profils  $\mathcal{LN}(6,1)$  et  $\mathcal{LN}(6,2)$ , et reste plus important que celui observé pour l'OBRE pour les autres profils (même si cela se voit difficilement

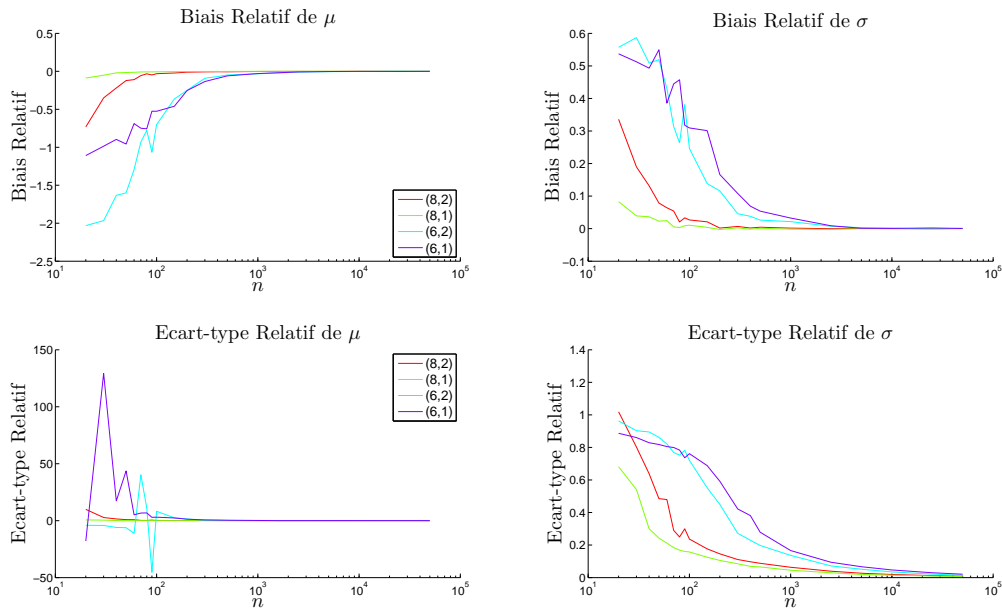


Figure 5.2 – Convergence de l'estimateur CvM pour différents profils de risque

graphiquement ; par exemple, celui du couple  $\mathcal{LN}(8,2)$  atteint une valeur de 10 pour  $n = 20$ ).

D'après ces graphiques, on peut conclure que cette méthode d'estimation est plus biaisée et plus volatile que la méthode OBRE. Ceci semble d'autant plus vrai lorsque la quantité  $F_\theta(H)$  est importante. En effet, plus cette quantité est élevée, plus la proportion de données censurées par rapport au modèle sans troncature est grande. Ainsi, pour un seuil de  $H = 1000$  on observe numériquement

$$F_{(8,1)}(H) < F_{(8,2)}(H) < F_{(6,2)}(H) < F_{(6,1)}(H)$$

C'est pourquoi, dans le prochain paragraphe, nous testons une nouvelle approche abordant la prise en compte du seuil différemment.

### 5.2.2 Prise en compte des données sous le seuil

Contrairement à la méthode utilisée pour une prise en compte "normale" du seuil de collecte énoncée précédemment, l'idée est ici de prendre en compte les pertes sous le seuil dans l'écriture de la fonction de répartition empirique, celle-ci n'étant donc plus nulle en  $H$ .

On dispose d'un échantillon de  $n$  pertes  $(x_1, \dots, x_n)$  toutes supérieures ou égales au seuil  $H$ . On note  $m$  le nombre de pertes sous le seuil, sur lesquelles on n'a aucune information en règle générale. On note également

$$n_{tot} = n + m$$

le nombre total de pertes. En supposant que la distribution totale provienne d'un modèle paramétrique  $F_\theta$ , on a l'équivalence entre les proportions suivantes

$$\frac{m}{n} = \frac{F_\theta(H)}{1 - F_\theta(H)} \quad (5.10)$$

il en résulte que :



$$n_{tot} = \frac{n}{1 - F_\theta(H)} \quad (5.11)$$

En considérant alors ces  $m$  pertes inobservables, on peut exprimer la fonction de répartition empirique

$$F_n^*(x) = \frac{m + \sum_{k=1}^n \mathbb{1}_{\{x_i \leq x\}}}{n_{tot}} \quad (5.12)$$

ce qui peut se réécrire à partir des équations (5.10) et (5.11)

$$F_n^*(x) = F_\theta(H) + \frac{1 - F_\theta(H)}{n} \sum_{k=1}^n \mathbb{1}_{\{x_i \leq x\}}$$

Le but de l'estimation est toujours de minimiser la distance entre les fonctions de répartition empirique et théorique. Comme on se place ici sur un modèle non censuré prenant en compte les données sous le seuil, l'estimateur CvM devient

$$T(F_\theta) = \arg \min_{\theta} \int_0^\infty (F_\theta(x) - F_n^*(x))^2 dF_\theta(x)$$

qui se discrétise sur un échantillon par

$$T(F_\theta) = \arg \min_{\theta} \sum_{i=1}^n \left( F_\theta(x_i) - \frac{m+i}{n_{tot}} \right)^2$$

ou encore

$$T(F_\theta) = \arg \min_{\theta} \sum_{i=1}^n \left( F_\theta(x_i) - F_\theta(H) - \frac{1 - F_\theta(H)}{n} i \right)^2$$

La recherche de minimum est faite sous Matlab, la fonction étant rentrée telle quelle, l'optimisation se faisant en une seule étape. Les résultats de convergence sont donnés figure 5.3.

Mis à part le profil de risque  $\mathcal{LN}(6,1)$ , cet estimateur est convergent. En s'intéressant aux grandes valeurs de  $n$ , sur les graphes, on montrerait que le biais relatif et l'écart-type des deux paramètres tendent vers 0.

Cependant, l'estimation est beaucoup trop instable pour des échantillons moins volumineux. En effet, les biais relatifs se comptent en millier de pourcent, les écarts-types sont colossaux, il faut des échantillons d'au moins 5000 pertes pour obtenir une convergence convenable.

La méthodes avec prise en compte des données sous le seuil apparait donc comme moins inadaptée dans la pratique. Comme l'OBRE, les propriétés de robustesse de cet estimateur pour ses deux versions sont étudiées dans la prochaine partie.

## 5.3 Estimateur Quantile-Distance

La méthode quantile-distance (QD) est une méthode d'estimation qui, comme l'estimation CvM, est basée sur la minimisation d'une distance. Cette méthode d'estimation ainsi que ses propriétés théoriques ont été originellement présentées par LaRiccia [11], et LaRiccia et Wehrly [12]. Son application au risque

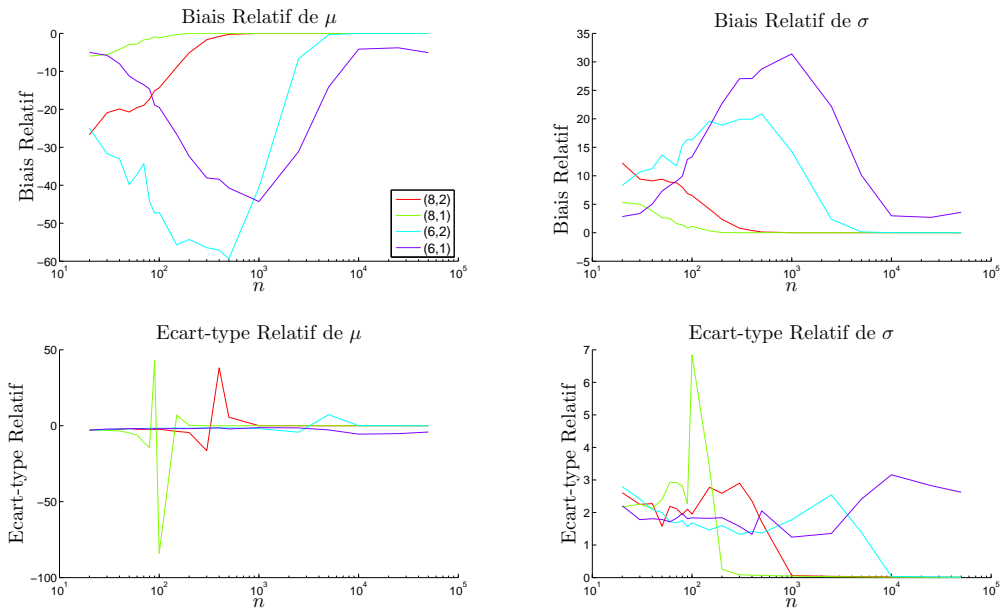


Figure 5.3 – Convergence de l'estimateur CvM avec prise en compte des données sous le seuil pour différents profils de risque

opérationnel a été proposée par Ergashev [5].

Alors que la méthode CvM se fonde sur l'écart entre fonction de répartition empirique et théorique, la méthode QD réside dans la minimisation d'une distance quadratique pondérée entre quantiles empiriques et théoriques. Cette distance est définie par

$$\mathcal{Q}^2(\theta, \{\alpha_1, \dots, \alpha_K\}) = \sum_{k=1}^K \omega_k (\hat{q}(\alpha_k) - F_\theta^{-1}(\alpha_k))^2 \quad (5.13)$$

où  $0 < \alpha_1 < \dots < \alpha_K < 1$  sont des niveaux de quantile,  $\omega_1, \dots, \omega_K$  sont des poids,  $F_\theta^{-1}$  correspond à la fonction quantile de la distribution  $F_\theta$ , c'est-à-dire la fonction inverse de  $F_\theta$ , et  $\hat{q}$  est une interpolation linéaire des quantiles empiriques, définie comme suit

$$\hat{q}(\alpha_k) = x_{(\lfloor n\alpha_k \rfloor)} + (n\alpha_k - \lfloor n\alpha_k \rfloor) (x_{(\lfloor n\alpha_k \rfloor + 1)} - x_{(\lfloor n\alpha_k \rfloor)}), k \in \llbracket 1, K \rrbracket \quad (5.14)$$

L'intérêt d'utiliser une interpolation des quantiles est de pouvoir avoir une estimation plus flexible, dans la mesure où la distance à minimiser est "lissée".

Afin de prendre en compte le seuil de troncature  $H$ , il est nécessaire de travailler avec des quantiles tradés. On peut en effet faire le constat que, pour une valeur de  $x$  donnée, le quantile associé pour la distribution tronquée  $\tilde{\alpha}$  est inférieur à celui associé pour la distribution totale  $\alpha$ , puisque  $\tilde{F}_\theta(x) < F_\theta(x)$ . En reprenant l'expression de la fonction de répartition pour une distribution tronquée :

$$\tilde{F}_\theta(x) = \frac{F_\theta(x) - F_\theta(H)}{1 - F_\theta(H)}$$

on en déduit

$$\alpha_k = \tilde{\alpha}_k + (1 - \tilde{\alpha}_k)F_\theta(H) \quad (5.15)$$

La distance quadratique à minimiser devient par conséquent

$$\tilde{\mathcal{Q}}^2(\theta, \{\alpha_1, \dots, \alpha_K\}) = \sum_{k=1}^K \omega_k \left( \hat{q}(\alpha_k) - F_{\theta}^{-1}(\hat{\alpha}_k + (1 - \hat{\alpha}_k)F_{\theta}(H)) \right)^2 \quad (5.16)$$

La méthode d'estimation QD peut donc se traduire par

$$T(\tilde{F}_{\theta} \mid \{\alpha_1, \dots, \alpha_K\}) = \arg \min_{\theta} \tilde{\mathcal{Q}}^2(\theta, \{\alpha_1, \dots, \alpha_K\}) \quad (5.17)$$

### 5.3.1 Spécification des paramètres

Dans ce paragraphe sont abordées les questions de calibration des différents paramètres entrant en jeu dans la fonctionnelle d'estimation : la pondération  $\omega_k$ , la répartition des niveaux de quantile  $\alpha_k$  et leur nombre,  $K$ .

#### Pondération

Selon Ergashev [5], les pondérations ont une influence négligeable sur l'estimation. Cependant, afin de limiter l'instabilité numérique liée à l'estimation sur les valeurs les plus sévères, il paraît opportun de limiter le poids qui leur est accordé en utilisant la pondération suivante :

$$\omega_k = \frac{1}{\hat{q}(\alpha_k)^2}, k \in \llbracket 1, K \rrbracket \quad (5.18)$$

Cela permet notamment d'éviter que l'estimation ne soit fondée que sur les quantiles empiriques les plus élevés.

#### Définition des niveaux de quantiles à ajuster, $\alpha_k$

La répartition des quantiles  $\alpha_k$  peut aussi bien se faire de façon uniforme que de façon exponentielle. En utilisant une répartition uniforme des quantiles sur  $[0,1]$ , aucune partie de la distribution n'est privilégiée dans l'estimation :

$$\alpha_k^{(\text{unif})} = \left(1 - \frac{1}{n+1}\right) \frac{k}{K}, \quad k \in \llbracket 1, K \rrbracket \quad (5.19)$$

On ne peut pas utiliser simplement l'écriture  $k/K$ , puisqu'alors on aurait  $\alpha_K = 1$  et par conséquent  $F_{\theta}^{-1}(\alpha_K)$  serait infini. Le coefficient  $(1 - \frac{1}{n+1})$  permet ainsi de prendre en compte (au moins partiellement) la perte la plus sévère<sup>6</sup>.

La figure 5.4 présente un exemple de répartition uniforme des niveaux de quantiles (représentés en points rouges).

La répartition exponentielle permet quant à elle de densifier le nombre de quantiles sur la queue de distribution. Cette option paraît opportune dans le contexte du risque opérationnel, puisqu'elle permet de se focaliser davantage sur la queue de distribution, partie de la distribution qui importe le plus lors du calcul de charge en capital. Les niveaux de quantiles peuvent être distribués exponentiellement selon

$$\alpha_k^{(\text{exp})} = \left(1 - \frac{1}{n+1}\right) \frac{1 - \exp(-km/K)}{1 - \exp(-m)}, \quad k \in \llbracket 1, K \rrbracket \quad (5.20)$$

$m$  est un paramètre permettant de moduler la concentration de quantiles sur la queue de distribution. Ainsi, plus  $m$  est grand, plus on se focalise sur la queue de distribution. À l'inverse, lorsque  $m$  tend vers 0, on retrouve une distribution uniforme des quantiles<sup>7</sup>.

6. Ergashev [5] choisit la valeur de 0.999 par défaut, mais il paraît plus intéressant d'utiliser une valeur dépendant de la taille de l'échantillon.

7. En effet, par développement limité on retrouve l'équation (5.19).

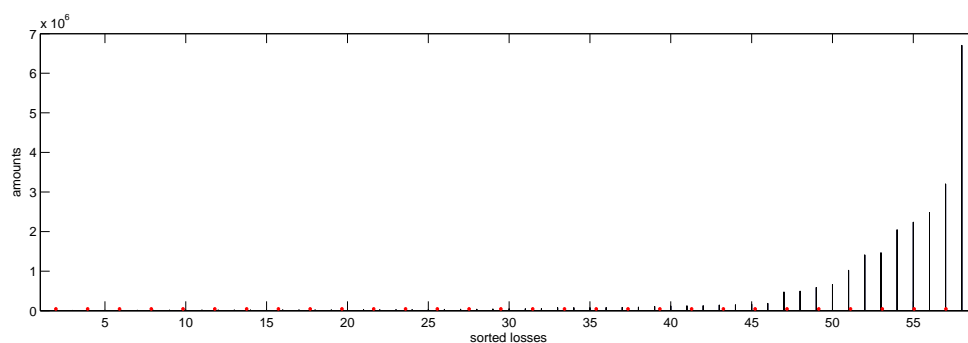


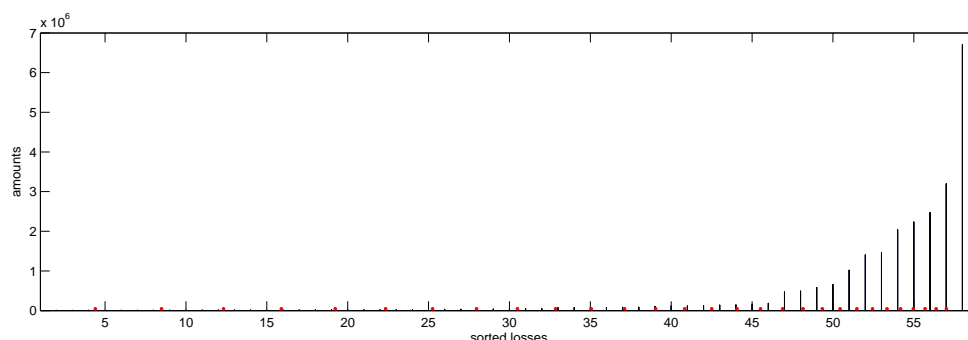
Figure 5.4 – Pertes ordonnées et niveaux de quantiles uniformes à ajuster.

Dans la suite, on utilise la valeur  $m = 2$  qui permet, tout en se focalisant un peu plus sur la queue, de garder également un nombre de niveaux de quantiles suffisant sur le corps de la distribution<sup>8</sup>. Le tableau 5.5 représente en pourcentage la répartition des quantiles sur différents intervalles de  $[0,1]$ .

Intervalle	Répartition	
	Uniforme	Exponentielle
$[0.5,1]$	50%	71.7%
$[0.9,1]$	10%	24.7%
$[0.99,1]$	1%	3.1%
$[0.999,1]$	0.1%	0.3%

Figure 5.5 – Répartition des quantiles distribués uniformément et exponentiellement ( $m = 2$ )

La figure 5.6 illustre la répartition exponentielle des niveaux de quantiles, avec  $m = 2$ .

Figure 5.6 – Pertes ordonnées et niveaux de quantiles exponentiels ( $m = 2$ ) à ajuster

Finalement, la dernière grandeur à fixer est le nombre total de quantiles. On voit sur la figure 5.7 que l'estimation des paramètres de la loi est peu affectée par la valeur de  $K$ .

Afin d'arbitrer entre flexibilité et précision de l'estimation, et ainsi avoir suffisamment de quantiles pour capter le comportement du corps et de la queue de la distribution et optimiser le temps de calcul, on choisit de poser :

$$K = \left\lfloor \frac{n}{2} \right\rfloor \quad (5.21)$$

---

8. Voir l'article [14].

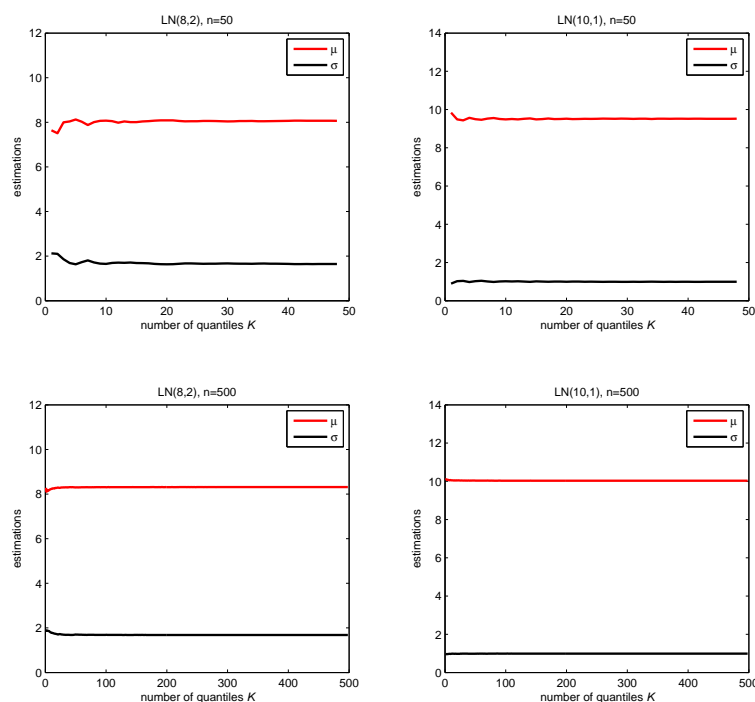


Figure 5.7 – Étude de différents nombres de quantiles à ajuster.

Deux versions de cette méthode sont proposées dans les prochains paragraphes, la différence réside dans la méthode d'optimisation numérique de l'équation (5.17).

### 5.3.2 Version par optimisation directe

L'équation d'optimisation (5.17) comporte deux valeurs de  $\theta$ . Par optimisation "directe", on entend une optimisation simultanée de ces deux valeurs. Cette optimisation se fait sous Matlab.

Comme pour les autres estimateurs, une étude de convergence est menée dans les mêmes conditions, qui sont, on le rappelle, un modèle général log-normal, 4 profils de risques, un seuil de  $H = 1000$ , différentes valeurs de  $n$  et les paramètres de l'estimateur QD donnés par les équations (5.18), (5.20) et (5.21). Les graphiques de cette étude sont présentés figure 5.8.

L'examen de ces graphes nous apprend que **l'estimateur est convergent**, avec une convergence appréciable quelque soit le profil de risque à partir d'une taille d'échantillon de 1000 données. **Le comportement de l'estimateur QD est d'ailleurs comparable à celui de l'estimateur CvM**, à savoir des biais de même grandeur selon les profils de risque, un paramètre  $\sigma$  biaisé à la hausse, un paramètre  $\mu$  biaisé un peu plus fortement à la baisse et un écart-type important pour de faibles tailles d'échantillon. La convergence s'avère cependant un peu plus rapide avec l'estimation par méthode CvM. Le biais à la hausse sur  $\sigma$  est cependant problématique, puisque il entrainera automatiquement une estimation surévaluée de la CaR. C'est pourquoi testons une seconde méthode afin de tenter de pallier à ce problème.

### 5.3.3 Version par optimisation scindée

Le paramètre  $\theta$  apparaissant deux fois dans l'équation (5.17), il est possible que cela entraîne une instabilité lors de la phase d'optimisation. L'idée est donc de procéder en deux étapes en estimant d'abord

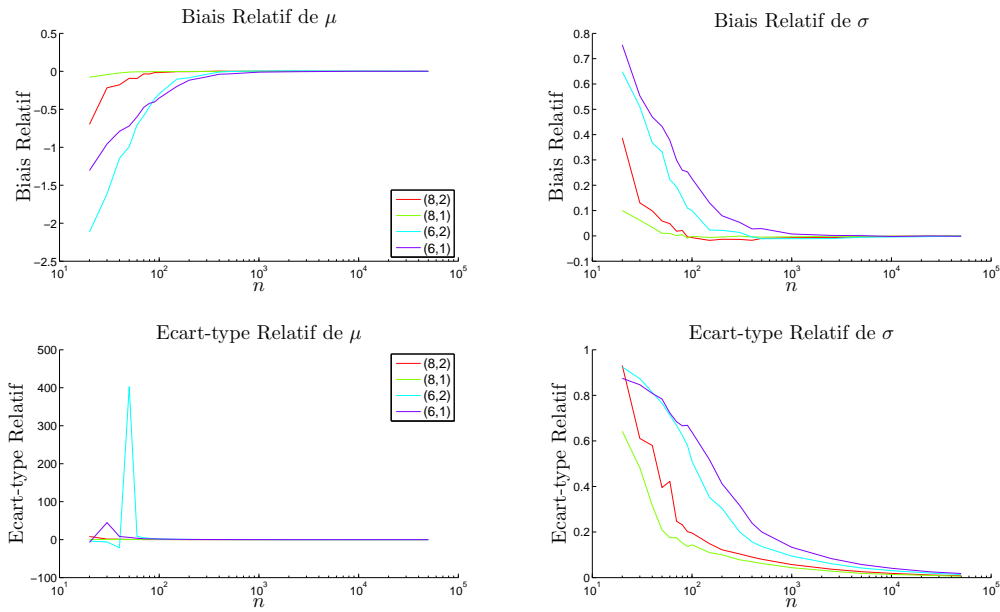


Figure 5.8 – Convergence de l'estimateur QD par optimisation "directe" pour différents profils de risque

une première valeur de  $\theta$ , que l'on notera  $\bar{\theta}$ , pour obtenir  $F_{\bar{\theta}}(H)$ , puis de fixer cette valeur afin d'obtenir la valeur définitive de l'estimation en minimisant la distance quadratique. Une autre raison à l'origine de l'élaboration d'une autre nouvelle version est que les résultats obtenus sur données réelles avec la première version ne sont pas satisfaisants.

La procédure est donc la suivante. On introduit l'estimation des quantiles empiriques translatés

$$q(\alpha_k) = q(\widetilde{\alpha}_k) - (1 - \widetilde{\alpha}_k)H \quad (5.22)$$

utilisés afin de minimiser la distance translatée correspondante

$$\bar{\theta} = \arg \min_{\theta} \sum_{k=1}^K \omega_k ((q(\widetilde{\alpha}_k) - (1 - \widetilde{\alpha}_k)H) - F_{\theta}^{-1}(\widetilde{\alpha}_k))^2 \quad (5.23)$$

Puis, l'estimation finale est obtenue, en fixant  $\bar{\theta}$  pour  $F_{\bar{\theta}}(H)$ , en minimisant la distance

$$\tilde{\mathcal{Q}}^2(\theta, \{\alpha_1, \dots, \alpha_K\}) = \sum_{k=1}^K \omega_k (\tilde{q}(\alpha_k) - F_{\theta}^{-1}(\widetilde{\alpha}_k + (1 - \widetilde{\alpha}_k)F_{\bar{\theta}}(H)))^2 \quad (5.24)$$

Les résultats obtenus avec cette optimisation en deux étapes sont résumés figure 5.9.

La première chose à noter à la vue de ces graphiques est que **cet estimateur n'est pas convergent**, puisque le biais ne tend pas vers 0, même si c'est le cas pour l'écart-type. Cependant, on peut observer une propriété de l'estimateur encore inobservée sur les méthodes d'estimation précédentes (OBRE, CvM, QD version directe) : pour un jeu de paramètres donné, **le biais de l'estimation ne dépend pas de la taille de l'échantillon**. Graphiquement, cela se constate très bien par la forme plane des courbes. De plus, si on compare l'écart-type relatif des paramètres estimés pour des petites tailles d'échantillon, on remarque qu'il est inférieur à toutes les autres méthodes d'estimation en ce qui concerne  $\mu$ , et équivalent à celui de l'OBRE (qui est actuellement le plus faible) en ce qui concerne  $\sigma$ .

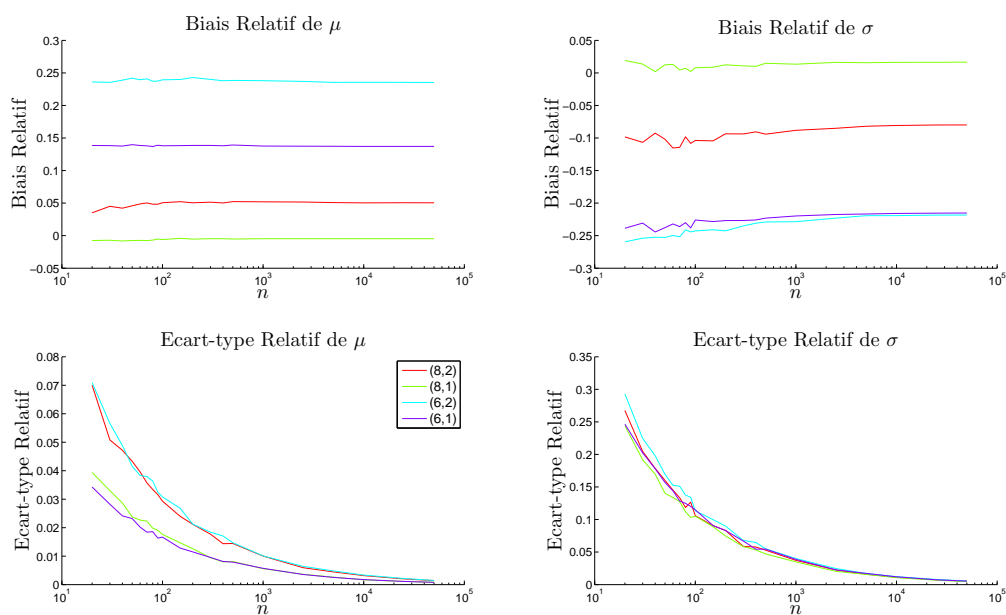


Figure 5.9 – Convergence de l'estimateur QD par optimisation "scindée" pour différents profils de risque

Comme dans la pratique, la taille des échantillons est très variable (de 20-30 pertes à 10000), cette méthode d'estimation apparaît très adaptée puisque la qualité et la précision de l'estimation ne dépend pas (ou très peu) du nombre d'observations.

Malgré ces propriétés remarquables, il n'en demeure pas moins que l'estimation reste biaisée. De plus, ce biais est lié au seuil de collecte  $H$  comme le montrent les courbes figure 5.10 obtenues pour le profil de risque  $\mathcal{LN}(6,2)$  pour différentes valeurs du seuil.

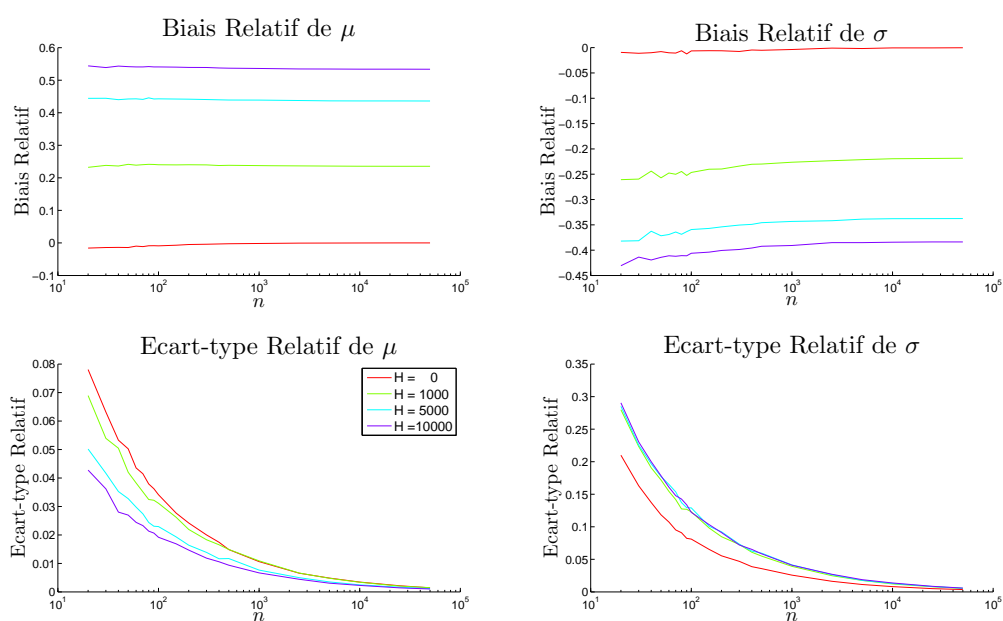


Figure 5.10 – Convergence de l'estimateur QD par optimisation "scindée" pour le profil (6,2) selon différentes valeurs de  $H$



## Partie III

---

# **Application au calcul du capital réglementaire**

## 6 Tests de robustesse

---

Cette partie aborde les propriétés de robustesse des estimateurs, le but étant de pouvoir les comparer entre eux sur ce critère.

Les tests de robustesse proposés dans cette partie sont de deux natures. Les premiers tests concernent la robustesse locale, qui se résument à des études de fonctions d'influence ; les seconds sont des tests de robustesse dite globale, dans lesquels on étudie le comportement de l'estimateur face à des contaminations de plus grande envergure.

L'étude présentée ici est réalisée pour tous les estimateurs robustes présentés ci-dessus, ainsi que les estimateurs par maximum de vraisemblance et par méthode des moments généralisée. Elle est également effectuée pour différents profils de risque, afin d'avoir une vision plus globale du comportement de ces estimateurs.

### 6.1 Robustesse locale

On étudie la fonction d'influence de chacun des estimateurs afin d'évaluer s'ils sont bien B-robustes. On rappelle que la fonction d'influence correspond à l'équivalent de la dérivée pour une méthode d'estimation, et qu'un estimateur est dit B-robuste si sa fonction d'influence est bornée.

On utilise ici les fonctions d'influence empiriques afin d'approximer les fonctions d'influence théoriques. On présente les fonctions d'influence des estimateurs pour un échantillon de 500 pertes simulées à partir d'une loi-lognormale de paramètres (8,2) tronquée au seuil  $H = 1000$ <sup>1</sup>.

Les fonctions d'influence empiriques sont tracées sur l'intervalle  $[H, Q_\theta(0.99999)]$ , où  $Q_\theta(0.99999)$  est le quantile à 99.999% de la distribution log-normale de paramètre  $\theta$ .  $x$  balaye cet intervalle en 250 valeurs réparties de façon exponentielle, afin d'avoir une répartition uniforme à l'échelle logarithmique. On représente également la variation relative de la CaR suite à l'ajout de la nouvelle perte  $x$ , qui est donc égale à la variation relative entre la CaR calculée à partir des nouveaux paramètres estimés et ceux estimés sans ajout de la perte  $x$ . Outre fournir l'impact d'une nouvelle perte  $x$  sur les paramètres estimés, cela nous permet également de quantifier l'impact final de cet ajout sur le calcul de capital.

Les graphes sont présentés séparément suivant les estimateurs. Ainsi, l'estimateur OBRE est présenté pour trois valeurs de  $c$ , les deux versions de la méthode QD sont proposées (dans les légendes, "QD" correspond à la méthode par optimisation directe et "QD scindé" par optimisation scindée), les deux versions de l'estimateur CvM (dans les légendes, "CvM" correspond à la méthode avec prise en compte normale du seuil "CvM v2" correspond à la version modifiée avec prise en compte des données sous le seuil), ainsi que l'estimateur ML et GMM. Enfin, les fonctions d'influence sur les paramètres  $\mu$  sont toujours représentées en trait plein, celles sur  $\sigma$  en trait pointillé.

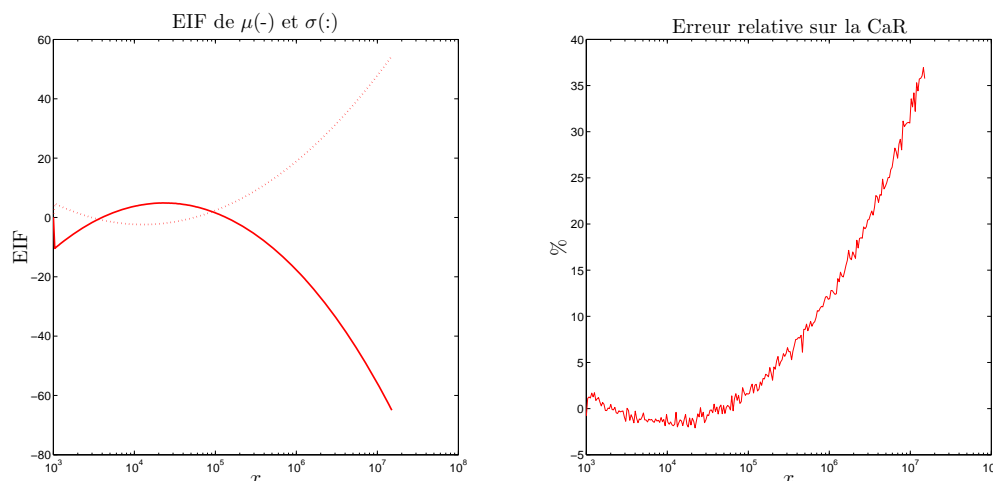


Figure 6.1 – Fonction d’influence empirique et variation relative de la CaR associée pour un échantillon de 500 données issues d’une loi  $\mathcal{LN}(8, 2)$  et tronquées au seuil  $H = 1000$  estimé par ML

### B-robustesse de l’estimateur ML

Les résultats pour l’estimateur ML sont donnés figure 6.1. On retrouve des fonctions d’influence pour  $\mu$  et  $\sigma$  évoluant en sens opposé et non bornées. Les fonctions d’influence restent proches de 0 sur  $[H, 10^5]$ , ce qui signifie que l’impact d’une nouvelle perte  $x \in [H, 10^5]$  ne fait que peu varier les paramètres. Ce qui explique la variation de la CaR quasi nulle sur cet intervalle. La valeur de  $10^5$  correspond environ au quantile à 95% de la distribution. Cependant, pour des valeurs supérieures, les fonctions d’influence deviennent non bornées, celle de  $\sigma$  tendant vers  $+\infty$  et  $\mu$  vers  $-\infty$ . Comme  $\sigma$  est le paramètre influant sur l’épaisseur de la queue, si sa fonction d’influence tend vers  $+\infty$ , cela signifie que sa valeur augmente indéfiniment, on observe alors une queue de plus en plus épaisse. Ceci se retrouve en regardant la variation relative de la CaR, la forme de cette courbe étant analogue à sur la fonction d’influence de  $\sigma$ . On observe ainsi une variation relative de la CaR non bornée, ce qui signifie que la nouvelle CaR estimée est non bornée également et donc tend vers  $+\infty$ .

En somme, en plus de retrouver que l’estimateur ML n’est pas B-robuste, on constate que l’ajout d’une seule perte de montant  $x$  peut aboutir à une estimation très élevée de la CaR.

### B-robustesse de l’estimateur GMM

Pour des soucis de lisibilité dûs à l’échelle, les résultats pour la méthode GMM sont également présentés avec un focus sur l’intervalle  $[H, 10^6]$  figure 6.2. On constate que cet estimateur n’est pas B-robuste, car les fonctions d’influence sont clairement non bornées, atteignant respectivement -700 pour  $\mu$  et 500 pour  $\sigma$ . Ceci s’explique par le procédé de la méthode d’estimation, puisque il consiste à se baser sur les moments empiriques d’ordre 1 et 2 des pertes. Or, la moyenne (moment d’ordre 1) n’étant pas robuste, donc rajouter une perte extrêmement élevée modifie nettement cette moyenne. Ceci est d’autant plus extrême pour le moment d’ordre 2. Ainsi, puisque l’estimation résulte de la mise en égalité entre moments empiriques et moments théoriques, on observe des fonctions d’influence empiriques très importantes pour de grandes valeurs de  $x$ .

Parallèlement, pour de plus petites valeurs de  $x$  et des valeurs autour de la moyenne, la fonction d’influence empirique observée est plane et proche de 0. C’est d’ailleurs, parmi les estimateurs présentés, celle

1. Les autres résultats pour les paramètres de simulation (8,1), (6,2) et (6,1) sont présentés en Annexe C, et les conclusions que nous pouvons en tirer sont similaires.

qui est la plus proche de l'axe des abscisses sur l'intervalle  $[H, 10^5]$ . Néanmoins, pour de grandes valeurs de  $x$ , l'estimateur devient fortement biaisé, ce qui conduit à des estimations de CaR aberrantes. C'est ainsi qu'on observe des variations relatives de CaR dépassant les 1000%, c'est-à-dire que l'ajout d'une seule perte dans un échantillon de 500 données peut aboutir à une multiplication par 10 du capital.

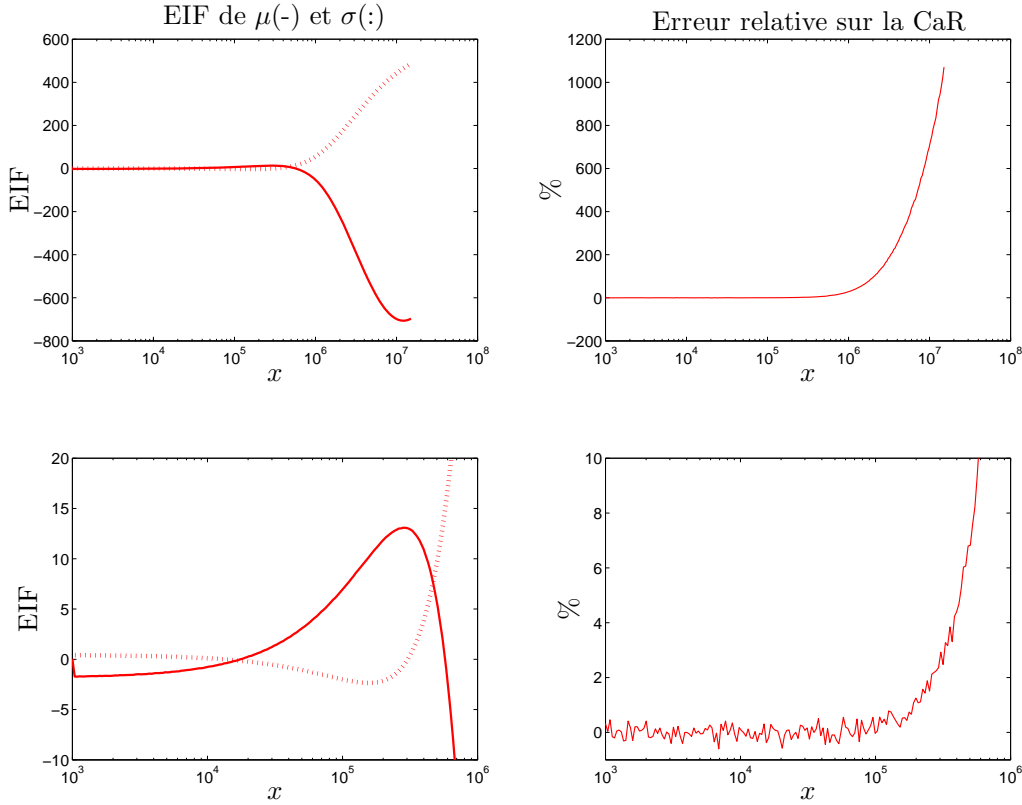


Figure 6.2 – Fonction d'influence empirique et variation relative de la CaR associée pour un échantillon de 500 données issues d'une loi  $\mathcal{LN}(8, 2)$  et tronquées au seuil  $H = 1000$  estimé par GMM

### B-robustesse de l'estimateur OBRE

Les résultats présentés ici sont donnés pour trois valeurs de  $c$ , à savoir  $\sqrt{2}$ , 2 et 5. Ces valeurs sont un moyen de jauger l'impact sur la robustesse de cet estimateur, puisqu'elles correspondent à trois niveaux de robustesse, en ordre décroissant. Les graphiques pour ces trois valeurs sont représentés figure 6.3. Comme l'estimateur OBRE est un estimateur dérivé du ML que l'on a rendu B-robuste, sont également superposés les graphes déjà obtenus pour l'estimateur ML afin de se rendre compte de l'apport de l'OBRE.

On peut constater que la valeur de  $c$  a bien un impact sur la robustesse de l'estimation puisque plus cette valeur est proche de  $\sqrt{2}$ , plus l'influence de l'ajout d'une nouvelle perte  $x$  sur les paramètres estimés est asymptotiquement faible. Mise à part la valeur  $c = 5$  pour laquelle les fonctions d'influence des paramètres semblent non bornées, **les valeurs  $\sqrt{2}$  et 2 donnent des estimateurs B-robustes.**

En comparant ces résultats avec ceux de l'estimateur ML, on constate plusieurs choses. La première est que, pour la valeur  $c = 5$ , les fonctions d'influence de l'OBRE et du ML sont égales, puis le caractère robuste opère dès lors que la valeur de perte  $5 \times 10^5$  est dépassée, valeur correspondant au quantile à

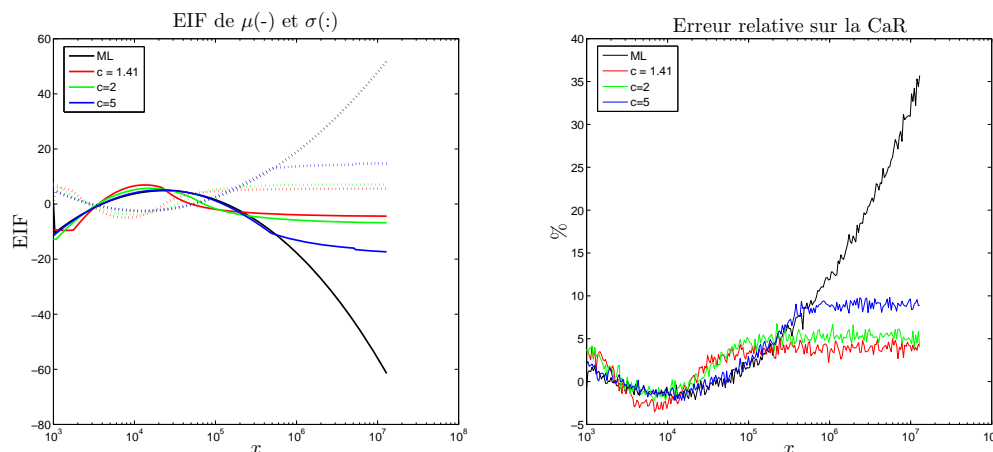


Figure 6.3 – Fonction d'influence empirique et variation relative de la CaR associée pour un échantillon de 500 données issues d'une loi  $\mathcal{LN}(8, 2)$  et tronquées au seuil  $H = 1000$  estimé par OBRE pour trois valeurs de  $c$

99.5% de la distribution. En revanche, pour les deux autres valeurs de  $c$ , bien qu'ayant la même forme, les courbes des EIF sont un peu plus élevées en valeur absolue au niveau du seuil et de "la bosse" (autour de  $x = 10^4$ ), ce qui signifie que l'impact de l'ajout d'une nouvelle perte  $x$  dans ces zones a plus d'incidence sur l'estimation des paramètres que la méthode ML. Cependant, c'est pour permettre d'être B-robuste que cette méthode est impactée sur ces certaines zones. Ainsi, **asymptotiquement, la variation relative de la CaR reste modérée**, elle est au maximum de 4% pour  $c = \sqrt{2}$  et de 5% pour  $c = 2$ , bien que les pertes les plus importantes représentent des quantiles de niveau 99.999%.

### B-robustesse de l'estimateur QD

D'après les graphes obtenus figure 6.4, on constate que **les deux versions sont B-robustes**. On peut par ailleurs remarquer que la version scindée de l'estimateur QD est plus robuste que la version par optimisation directe. Là encore, l'influence des points entre  $H$  et  $10^5$  est faible comparée à celles correspondant à des quantiles plus élevés. Le maximum des fonctions d'influence est atteint pour un  $x$  avoisinant  $10^6$  qui correspond quasiment au niveau de quantile pour lequel est calculée la CaR, à savoir 99.9%. À partir de cette valeur, les courbes se stabilisent, voire, chose intéressante, décroissent lentement. Ainsi, la variation relative de l'estimation par optimisation directe est au maximum de 27% et atteint 25% pour la valeur de  $x$  la plus élevée, alors que celle par optimisation scindée est de 17% au maximum puis retombe à 16%.

### B-robustesse de l'estimateur CvM

Si on observe qu'une seule couleur de courbe sur la figure 6.5 (graphe de gauche), c'est que les fonctions d'influence empiriques de l'estimateur CvM avec prise en compte normale et modifiée du seuil sont confondues. On constate que **cet estimateur est bien B-robuste pour ses deux versions**, et que l'allure de ses fonctions d'influence est comparable à celle de l'OBRE pour  $c = \sqrt{2}$ , à savoir une amplitude importante au niveau du seuil (-11 pour l'EIF de  $\mu$  et 6 pour  $\sigma$ ), un extremum local tout juste avant  $x = 10^4$  ainsi qu'une stabilité asymptotique aux alentours de  $5 \times 10^5$  et une valeur limite comparable (6 pour  $\sigma$  et -3 pour  $\mu$ ). De même que l'OBRE, les variations de CaR associées sont semblables, celles avec prise en compte des données sous le seuil étant légèrement supérieures. La valeur maximale atteinte est simplement de 4%.

Cette étude de robustesse locale à partir des fonctions d'influence empiriques prouve que

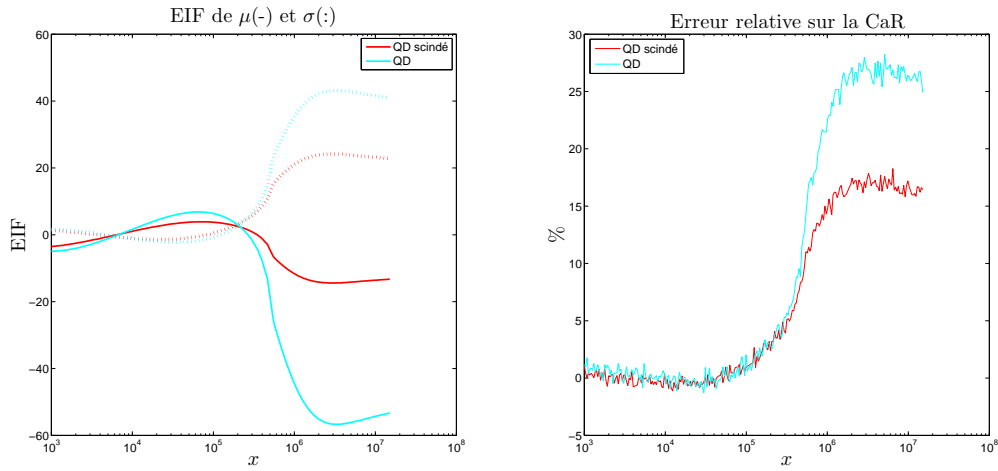


Figure 6.4 – Fonction d'influence empirique et variation relative de la CaR associée pour un échantillon de 500 données issues d'une loi  $\mathcal{LN}(8, 2)$  et tronquées au seuil  $H = 1000$  estimé par les deux versions de la méthode QD

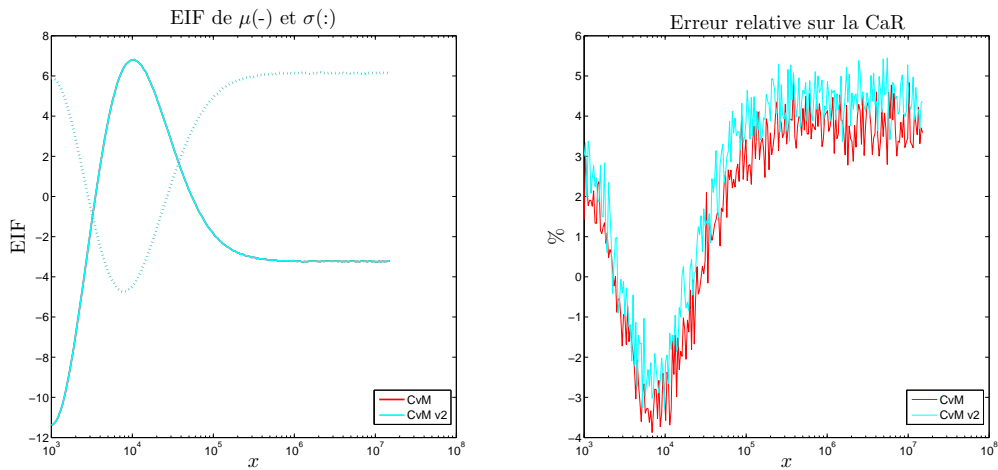


Figure 6.5 – Fonction d'influence empirique et variation relative de la CaR associée pour un échantillon de 500 données issues d'une loi  $\mathcal{LN}(8, 2)$  et tronquées au seuil  $H = 1000$  estimé par les deux versions de la méthode QD

les estimateurs présentés dans la partie 5 sont **B-robustes**. Afin d'affiner la comparaison entre ces estimateurs, un examen plus approfondi, portant sur la robustesse globale, est présenté dans la section suivante.

## 6.2 Robustesse globale

Alors que dans le cas de la robustesse locale, on ne contamine que par une seule perte, la robustesse globale consiste en une contamination d'un pourcentage  $\alpha$  de la distribution, ce pourcentage pouvant être non négligeable.

On procède à une contamination comme proposée équation (4.2), à savoir :

$$F_{\theta,\alpha} = (1 - \alpha)F_{\theta} + \alpha G$$

La distribution  $F_{\theta}$  reste toujours une loi log-normale de paramètre  $\theta$ , et  $G$  est le modèle contaminant sur lequel nous pouvons travailler. Ainsi nous présentons dans la suite différents modèles de contamination qui se divisent en deux catégories : les modèles de contamination à queue fine, qui sont modélisés par des distributions log-normales, et les modèles à queue épaisse, modélisés par des distributions de Pareto.

L'objectif de l'estimation robuste est, dans la mesure du possible, d'estimer correctement le paramètre  $\theta$  de la distribution contaminée, et cela quelque soit le pourcentage de contamination  $\alpha$  et le modèle contaminant  $G$ . Dans cette étude, on observe l'évolution de l'estimation en fonction de  $\alpha$  et de  $G$  pour deux modèles  $F_{\theta}$ , un modèle  $\mathcal{LN}(8, 2)$  et un  $\mathcal{LN}(8, 1)$ .

On se place dans des conditions pour lesquelles les estimateurs sont non biaisés en moyenne, c'est-à-dire des tailles d'échantillon de 1000 données. La procédure de l'étude est la suivante :

- (1) On se donne un modèle de base  $F_{\theta}$ , un modèle contaminant  $G$  et un vecteur regroupant différentes valeurs de contamination  $\alpha$ .
- (2) 1000 échantillons de 1000 données sont simulés à partir de la distribution non contaminée  $F_{\theta}$ .
- (3) Pour chaque estimateur, on effectue alors 1000 estimations sur ces 1000 échantillons. Elles servent de référence pour les estimations sur données contaminées.
- (4) Ensuite, pour chaque valeur de  $\alpha$  et chaque échantillon simulé à l'étape 2., on retire aléatoirement un pourcentage  $\alpha$  de données non contaminées que l'on remplace par des données issues du modèle contaminant  $G$ .
- (5) Pour chaque estimateur, on réeffectue 1000 estimations des paramètres du modèle  $F_{\theta,\alpha}$ .
- (6) Pour chaque estimateur, on compare les estimations sur les modèles contaminés  $F_{\theta,\alpha}$  et le modèle de référence  $F_{\theta}$ .
- (7) Les estimateurs peuvent être comparés entre eux.

La comparaison faite à l'étape (6) se base sur trois critères :

- les paramètres estimés ;
- les biais relatifs<sup>2</sup> ;
- les CaRs.

Pour comparer les paramètres estimés, on s'intéresse à la moyenne des estimations, à leur écart-type ainsi qu'à l'intervalle de confiance empirique à 95%. Pour obtenir cet intervalle de confiance, on ordonne les 1000 estimations des paramètres  $(\hat{\mu}_i, \hat{\sigma}_i)$  par rapport à  $\hat{\sigma}$ , les quantiles empiriques à 2.5% et 97.5% sont calculés pour le paramètre  $\hat{\sigma}$ , en supposant qu'ils correspondent respectivement aux indices  $k$  et  $l$  des 1000 estimations, on obtient alors l'intervalle de confiance  $[(\hat{\mu}_k, \hat{\sigma}_k), (\hat{\mu}_l, \hat{\sigma}_l)]$ .

Il est important de comprendre l'intérêt de ne pas prendre simplement les quantiles empiriques de  $\sigma$  et de  $\mu$ . Si on regarde, pour 1000 échantillons, la répartition des paramètres estimés entre  $\hat{\mu}$  et  $\hat{\sigma}$  n'est pas uniforme sur un pavé, mais a la forme d'une "diagonale", ceci étant dû aux paramètres extra diagonaux non nuls de la matrice de covariance de l'estimateur. Un exemple est donné figure 6.6. Ainsi, on peut voir en rouge les deux points correspondant à l'intervalle de confiance à 95%, et en vert ceux correspondant à une prise en compte des quantiles sur  $\hat{\mu}$  et sur  $\hat{\sigma}$ . On comprend ainsi que les quantiles calculés de cette façon ne correspondent pas à une estimation pouvant être observée.

Pour chacun des 1000 échantillons contaminés, on obtient un biais relatif en comparant l'erreur relative entre l'estimation de cet échantillon contaminé et l'estimation de ce même échantillon non contaminé. On obtient ainsi 1000 biais relatifs, dont on calcule la moyenne et l'intervalle de confiance empirique à 95%.

---

2. On entend par biais, la différence relative entre les paramètres estimés sur données contaminées et sur données non contaminées, et non pas celle entre les paramètres estimés et le vrai paramètre  $\theta$  du modèle  $F_{\theta}$ .

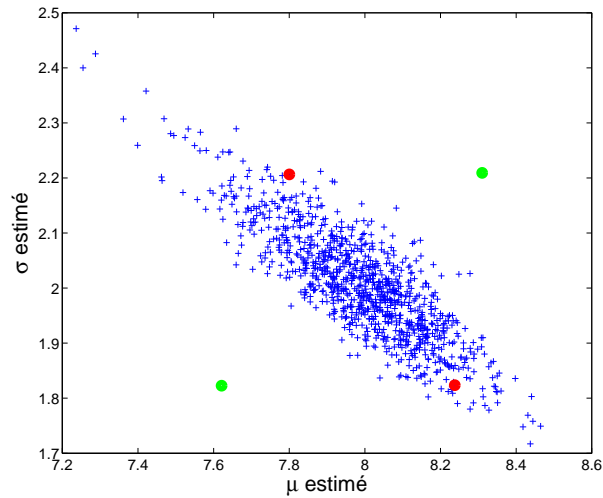


Figure 6.6 – Fonction d’influence empirique et variation relative de la CaR associée pour un échantillon de 500 données issues d’une loi  $\mathcal{LN}(8, 2)$  et tronquées au seuil  $H = 1000$  estimé par GMM

Les CaR moyennes sont obtenues à partir de l’estimation moyenne des paramètres, et l’intervalle de confiance à 95% est obtenu à partir de l’intervalle de confiance à 95% des estimations. Là encore, on voit l’intérêt de la prise en compte des quantiles choisie qui aboutit à des estimations de capital cohérentes avec l’échantillon, alors que la prise en compte "basique" (points verts sur la figure 6.6) amènerait quant à elle à une borne haute de la CaR surévaluée et une borne basse sous-évaluée.

Les résultats de robustesse globale sont réalisés sur les modèles  $\mathcal{LN}(8, 2)$  et  $\mathcal{LN}(8, 1)$  que l’on contamine aussi bien par une loi log-normale de paramètres différents que par une loi de Pareto. La loi de Pareto  $\mathcal{P}(\alpha, \beta)$  est une loi puissance à queue épaisse à deux paramètres. Sa fonction de densité est définie par

$$f(x) = \frac{\alpha}{\beta} \left(1 + \frac{x}{\beta}\right)^{-(\alpha+1)}, \quad (\alpha, \beta) \in \mathbb{R}_+^{*2}$$

On ne présente ici que les résultats obtenus pour un modèle  $\mathcal{LN}(8, 2)$  contaminé par un modèle  $\mathcal{LN}(8, 4)$ , car ce modèle de contamination est celui qui affecte le plus les données et donc les estimations sur un modèle  $\mathcal{LN}(8, 2)$ .

En Annexe C sont également donnés les résultats de contamination pour  $\mathcal{LN}(8, 2)/\mathcal{LN}(8, 3)$ ,  $\mathcal{LN}(8, 1)/\mathcal{LN}(8, 2)$  et  $\mathcal{LN}(8, 1)/\mathcal{LN}(8, 3)$  ainsi que pour les modèles  $\mathcal{LN}(8, 2)/\mathcal{P}(0.7, 4000)$  et  $\mathcal{LN}(8, 2)/\mathcal{P}(0.6, 4000)$ . Les conclusions et les constats généraux énoncés dans ce paragraphe pour une contamination  $\mathcal{LN}(8, 2)/\mathcal{LN}(8, 4)$  sont également valables aux autres modèles de contamination présentés dans cette annexe.

Les différents pourcentages de contamination  $\alpha$  utilisés sont les suivants : 1‰, 5‰, 1%, 5%, 10%, 15%, 20%, 25% et 50%. On choisit des valeurs de  $\alpha$  très élevées afin de pouvoir dégager une tendance plus globale, mais l’hypothèse d’un pourcentage de contamination très important est discuté par la suite.

### Contamination $\mathcal{LN}(8, 2)$ par $\mathcal{LN}(8, 4)$

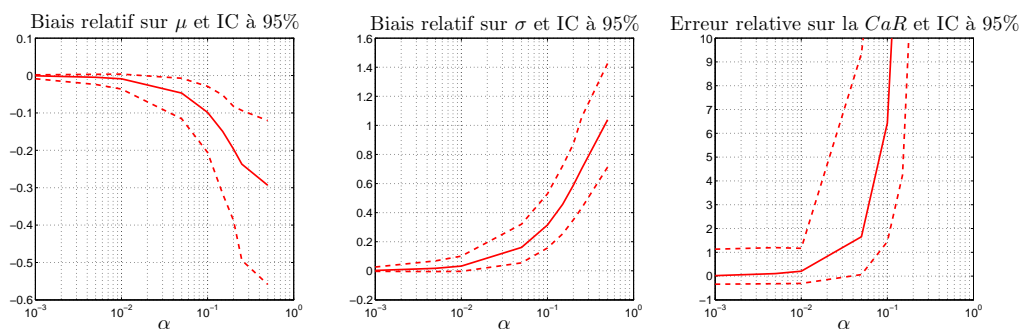
Le modèle  $\mathcal{LN}(8, 4)$  est extrêmement sévère comparativement au modèle  $\mathcal{LN}(8, 2)$ , comme en témoigne le tableau 6.1.



Distribution	Moyenne	Moment d'ordre 2	Quantile à 99.9%
$\mathcal{LN}(8,2)$	$2.2^4$	$2.6^{10}$	$1.4^6$
$\mathcal{LN}(8,4)$	$8.9^6$	$7.0^{20}$	$7.0^8$

Table 6.1 – Comparaison des distributions  $\mathcal{LN}(8,2)$  et  $\mathcal{LN}(8,4)$ 

## Résultats obtenus pour l'estimateur ML

Figure 6.7 – Étude sur l'estimateur ML de la contamination d'une loi  $\mathcal{LN}(8,2)$  par une loi  $\mathcal{LN}(8,4)$  pour différents pourcentages de contamination

On constate sur la figure 6.7 que le biais relatif sur  $\sigma$  augmente et que celui sur  $\mu$  est négatif et augmente en valeur absolue en fonction de  $\alpha$ . Les graphes sont ici représentés en échelle logarithmique sur l'axe des abscisses afin de mieux visualiser les courbes pour les faibles valeurs de  $\alpha$ . Néanmoins, sur une échelle linéaire, on pourrait constater que les biais relatifs sont linéaires en  $\alpha$  sur  $[1\%, 25\%]$ , et que la pente devient plus faible sur  $[25\%, 50\%]$ .

Pour une contamination très faible,  $\alpha = 1\%$ ,  $5\%$  et  $1\%$ , l'impact sur l'estimation des paramètres est faible, même si on peut observer tout de même un biais relatif de 2% et 3% en moyenne sur  $\sigma$  pour  $\alpha = 5\%$  et  $1\%$  respectivement. On observe également que la CaR augmente de 11% et 20% pour ces deux valeurs de  $\alpha$ .

Pour 5% et 10% de contamination, le biais sur  $\mu$  reste modéré (5% et 10% respectivement) mais celui sur  $\sigma$  devient non négligeable avec des valeurs de 16% pour  $\alpha = 5\%$  et 32% pour  $\alpha = 10\%$ . En conséquence, la CaR obtenue est multipliée par 1.7 pour  $\alpha = 5\%$  et par 6.4 pour  $\alpha = 10\%$ . Bien que le biais soit linéaire en  $\alpha$ , ce n'est pas le cas de la CaR : on le constate ici car, en doublant  $\alpha$  pour passer de 5 à 10%, le biais relatif sur les paramètres est bien doublé, mais l'erreur relative sur la CaR est quadruplée.

Pour des valeurs de  $\alpha$  plus élevées, on observe toujours une linéarité entre le biais relatif et  $\alpha$ , ainsi qu'une augmentation exponentielle de la CaR par rapport au pourcentage de contamination.

## Résultats obtenus pour l'estimateur GMM

Le comportement de l'estimateur GMM face à la contamination est différent des précédents. En effet, la première chose à remarquer est que le biais relatif sur  $\mu$  devient positif, alors qu'il est négatif pour les autres estimateurs, et nul dans le cas de la méthode QD scindée. La seconde est que les intervalles de confiance à 95% sont très étendus, et ce même pour de faibles contaminations.

Sur l'intervalle  $[1\%, 10\%]$ , le biais sur  $\mu$  reste d'abord très bas pour atteindre ensuite 13% pour  $\alpha = 10\%$ . Le biais sur ce paramètre reste acceptable pour des contaminations ne dépassant pas 10%.

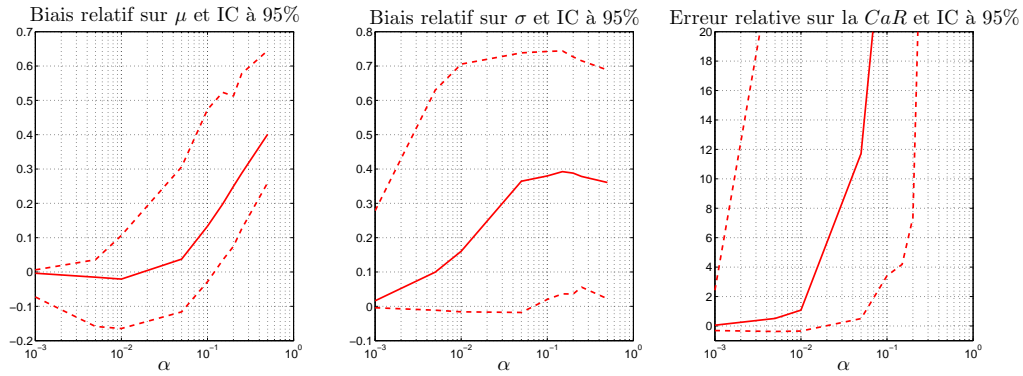


Figure 6.8 – Étude sur l’estimateur GMM de la contamination d’une loi  $\mathcal{LN}(8, 2)$  par une loi  $\mathcal{LN}(8, 4)$  pour différents pourcentages de contamination

Le biais sur  $\sigma$  est en revanche plus problématique. En effet, même pour une contamination de 1‰, c’est-à-dire en ne remplaçant qu’une seule perte de l’échantillon de taille 1000 par une observation issue d’une loi  $\mathcal{LN}(8, 4)$ , le biais moyen obtenu est de 2%. Il passe ensuite à 10% pour  $\alpha = 5‰$ , puis 16% pour 1% et se stabilise ensuite à hauteur de 36% à partir de 5% de contamination.

Ainsi, on observe parallèlement une erreur relative sur la CaR de 5%, 51%, 108%, 1172% puis 2995% pour les valeurs respectives de  $\alpha$  : 1‰, 5‰, 1%, 5% et 10%.

### Résultats obtenus pour l’estimateur OBRE

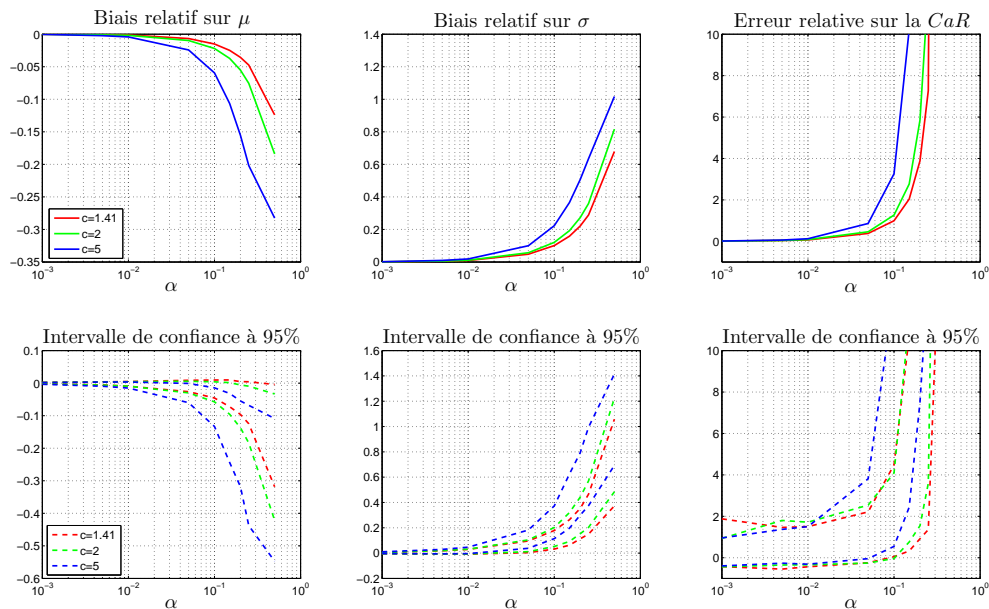


Figure 6.9 – Étude sur l’estimateur OBRE pour 3 valeurs de  $c$  de la contamination d’une loi  $\mathcal{LN}(8, 2)$  par une loi  $\mathcal{LN}(8, 4)$  pour différents pourcentages de contamination

On constate premièrement figure 6.9 que, pour chacune des trois valeurs de  $c$ , les biais relatifs et l’erreur relative sur la CaR pour un  $\alpha$  donné sont bien inférieurs à ceux de l’estimateur ML, cela étant d’autant

plus marqué lorsque  $c$  diminue. En effet, alors que pour  $\alpha = 5\%$  le biais relatif moyen est de  $-5\%$  pour  $\mu$  et  $16\%$  pour  $\sigma$  pour le ML, il est de  $-1\%$ ,  $-1\%$  et  $-2\%$  sur  $\mu$  et  $5\%$ ,  $6\%$  et  $10\%$  sur  $\sigma$ , respectivement pour les valeurs de  $c : \sqrt{2}, 2$  et  $5$ . Pour  $\alpha = 10\%$ , on peut également comparer les valeurs moyennes de biais  $-10\%$  sur  $\mu$  et  $31\%$  sur  $\sigma$  aux valeurs  $-2\%$ ,  $-2\%$  et  $-6\%$  sur  $\mu$  et  $10\%$ ,  $12\%$  et  $22\%$  sur  $\sigma$  obtenues pour l'OBRE. Ainsi, l'erreur relative sur la CaR est aussi minimiser puisque les biais sont réduits. On observe pour  $\alpha = 10\%$ , une erreur relative de  $99\%$ ,  $127\%$  et  $326\%$  pour les valeurs respectives  $c = \sqrt{2}, 2$  et  $5$ , alors que cette erreur était de  $643\%$  pour l'estimateur ML.

Même pour des valeurs de  $\alpha$  très élevées, l'estimateur reste très résistant à la contamination, principalement pour  $c = \sqrt{2}$  et  $c = 2$ .

Les intervalles de confiance empiriques à  $95\%$  restent équivalents à ceux obtenus pour la méthode ML.

Les résultats obtenus pour l'estimateur OBRE sont donc bien meilleurs que ceux pour l'estimateur ML. On peut également se rendre compte de l'impact du paramètre  $c$  sur l'estimation des paramètres et donc sur la CaR estimée.

## Résultats obtenus pour l'estimateur QD

On observe figure 6.10 que les comportements des deux versions de l'estimateur sont très différents.

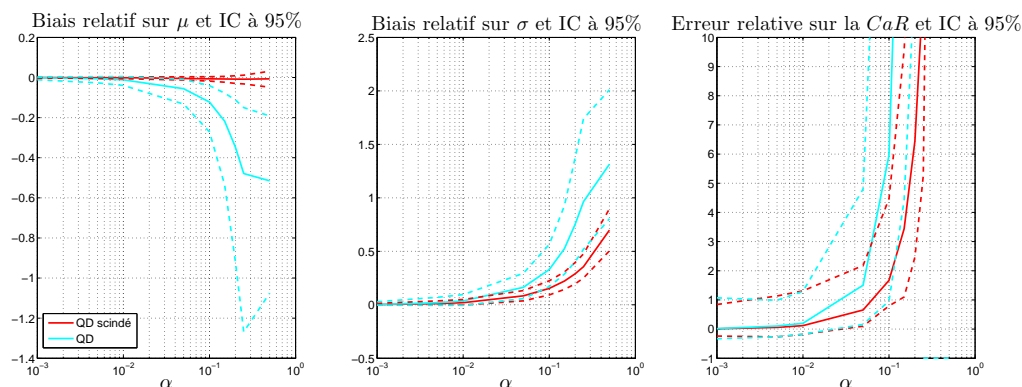


Figure 6.10 – Étude sur les deux versions de l'estimateur QD de la contamination d'une loi  $\mathcal{LN}(8, 2)$  par une loi  $\mathcal{LN}(8, 4)$  pour différents pourcentages de contamination

En effet, l'estimation de  $\mu$  par optimisation scindée est très robuste, puisque, même pour des contaminations extrêmes, le biais moyen reste proche de  $0\%$  avec un intervalle de confiance très concentré autour de la moyenne, alors que le biais sur  $\mu$  par optimisation directe est comparable à celui de l'estimateur ML, voire plus important ( $-13\%$  de biais pour  $\alpha = 10\%$  alors qu'on observe  $-10\%$  pour le ML).

De même, le biais relatif moyen observé sur  $\sigma$  par optimisation directe est plus important que par optimisation scindée. Il est même légèrement plus important que celui obtenu avec l'estimateur ML. En revanche, pour l'autre version, les biais obtenus sont très ressemblants à ceux obtenus avec l'estimateur OBRE pour la valeur  $c = 2$ , en restant supérieurs de quelques pourcents.

D'après ces observations sur les paramètres, il est logique de constater que l'erreur relative sur la CaR par optimisation directe est équivalente à celle obtenue avec la méthode ML, et que cette erreur obtenue par optimisation scindée est du même ordre que celle obtenue avec l'OBRE pour  $c = 2$ , en étant légèrement supérieure. Pour  $\alpha = 5\%$ , on peut vérifier que cette erreur relative est de  $150\%$  par version directe alors qu'elle est de  $166\%$  par ML, et qu'elle est de  $66\%$  par version scindée alors qu'elle est de  $46\%$  pour l'OBRE.

avec  $c = 2$ .

D'après ces résultats, **l'estimateur par optimisation directe n'apparaît pas adapté au problème de contamination globale**, puisque ses performances sont comparables à celles de l'estimateur ML. L'estimateur par optimisation scindée semble quant à lui plus apte à répondre à ce problème.

## Résultats obtenus pour l'estimateur CvM

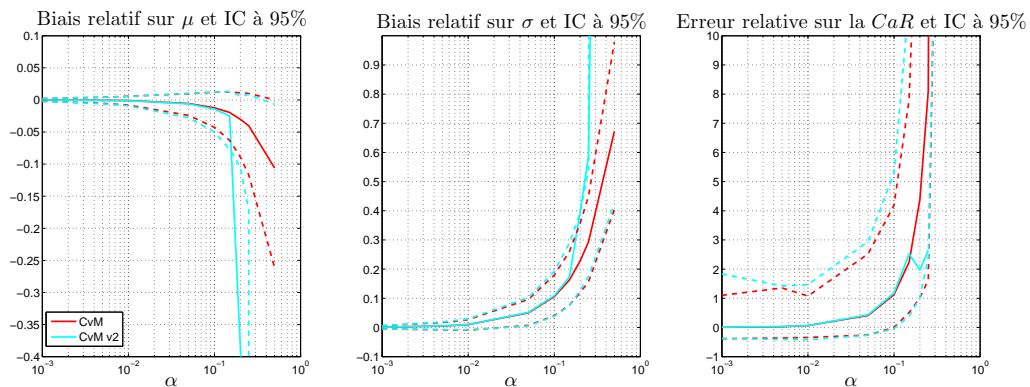


Figure 6.11 – Étude sur les deux versions de l'estimateur CvM de la contamination d'une loi  $\mathcal{LN}(8, 2)$  par une loi  $\mathcal{LN}(8, 4)$  pour différents pourcentages de contamination

Alors que leurs courbes étaient indiscernables dans l'étude de robustesse locale, ce n'est plus le cas ici pour de hauts pourcentages  $\alpha$  comme on peut le voir figure 6.11. D'ailleurs, l'estimation moyenne par méthode de prise en compte des données sous le seuil devient très instable à partir de 20% de contamination, mais sur  $[1\%, 15\%]$  les deux estimateurs apportent les mêmes résultats.

Les résultats avec prise en compte "normale" du seuil sont presque superposables à ceux obtenus pour l'OBRE avec  $c = \sqrt{2}$ . C'est également le cas de la version avec prise en compte des données sous le seuil mais sur l'intervalle  $[1\%, 15\%]$ . Ainsi, on peut observer pour une valeur  $\alpha = 10\%$  des biais de -1% sur  $\mu$  et 11% sur  $\sigma$ , que l'on peut comparer à celles de l'OBRE pour  $c = \sqrt{2}$ , à savoir -2% et 10%, ou encore -2% et 16% pour  $\alpha = 15\%$ , valeurs identiques à celles de l'OBRE.

Comme les biais sont comparables, les erreurs relatives obtenues sur les CaR sont également similaires.

L'intervalle de confiance à 95% sur les paramètres est également analogue à celui de l'OBRE pour  $c = \sqrt{2}$ , à savoir concentré autour de la moyenne.

D'après ces résultats de robustesse globale, l'estimateur CvM apparaît comme une méthode d'estimation concurrençant la méthode OBRE.

En conclusion de cette étude de robustesse, on peut affirmer que, malgré la B-robustesse des estimateurs OBRE, QD et CvM, leur caractère robuste est très variable d'un estimateur à l'autre sur données contaminées de façon globale. **Il ressort que les estimateurs OBRE, QD par optimisation scindée et CvM avec prise en compte normale du seuil sont les mieux adaptés pour traiter ce problème.** Les estimateurs OBRE et CvM sont également ceux produisant les résultats les plus satisfaisants, l'OBRE ayant de plus l'avantage d'être modulable via son paramètre  $c$ .

D'autre part, cette étude montre que **les estimateurs ML et GMM ne sont pas capables d'aborder de tels problèmes de contamination.** Ceci est d'autant plus vrai pour l'estimateur GMM, qui ne parvient pas à pallier cette difficulté, même pour des contaminations très faibles.

## 7 Calcul de capital sur données réelles

---

Après avoir constaté les propriétés théoriques de ces estimateurs sur données simulées, il est intéressant de les tester sur données réelles.

### 7.1 Résultats généraux

Afin de tester l'ensemble de ces estimateurs, nous disposons d'un panel de 14 échantillons, notés E1 à E14, provenant de différentes entités du groupe et de différentes cellules de risque. Pour l'ensemble de ces échantillons, le seuil de collecte  $H$  est de 1000€, sauf pour l'échantillon E3 dont celui-ci est de 1€. Le nombre d'observations varie de 55 à 7474 selon les échantillons.

Pour chaque échantillon, nous estimons les paramètres de la loi de sévérité par 9 méthodes d'estimations : ML, OBRE pour les valeurs  $c = \{\sqrt{2}, 2, 5\}$ , QD par optimisation directe et scindée, CvM avec et sans prise en compte des données sous le seuil, et GMM. Nous calculons également la valeur des fonctions de répartition au seuil  $H$  et les CaR associées à ces estimations, ainsi que les  $p$ -values des 4 tests d'adéquation.

Les tableaux de résultats sont fournis en Annexe D. Afin de rendre ces tableaux plus lisibles, nous adoptons le code couleur suivant :

- Une cellule coloriée en vert correspond à une  $p$ -value supérieure à 5%, ce qui signifie que le test d'adéquation n'est pas rejeté au risque de 5%. Dans le cas contraire, le test est rejeté.
- Une cellule coloriée en rouge correspond soit à une  $p$ -value que l'on ne peut calculer à cause des paramètres estimés tels que  $F_{\hat{\theta}}(H) \approx 1$ , ce qui ne permet pas son calcul par méthode de Monte Carlo, soit à une valeur de CaR qu'on ne peut calculer pour les mêmes raisons.
- L'estimation du paramètre  $\hat{\mu}$  est écrite en rouge lorsque celui-ci est inférieur à 0.

La première conclusion que nous pouvons tirer à la vue de ces tableaux, est que, pour un échantillon donné,  $(\hat{\mu}, \hat{\sigma})$  **diffère nettement d'une méthode d'estimation à l'autre**. Alors que dans un cadre *iid* ou même dans des tests de contamination, ces estimations sont du même ordre. Cela souligne donc l'hétérogénéité des échantillons rencontrés dans la pratique.

On peut remarquer beaucoup d'estimations  $\hat{\mu}$  inférieures à 0, ce qui n'a pas réellement de sens d'un point de vue métier. En effet, les fonctions de répartition en  $H$  associées à ces estimations sont toutes égales à 100%, ce qui signifie que toutes les pertes se situent sous ce seuil. De plus, comme la moyenne d'une loi log-normale est égale à  $\exp(\mu + \sigma^2/2)$ , une estimation  $\hat{\mu}$  négative revient à supposer que la perte moyenne est proche de 0, ou du moins très inférieure au seuil  $H$ .

À ce titre, on peut voir que les résultats obtenus avec l'estimateur CvM avec prise en compte des données sous le seuil sont ceux pour lequel ce phénomène est le plus observé (12 cas sur 14). On constate que, pour l'autre version de l'estimateur CvM, ce problème est rencontré pour la moitié des échantillons. Parmi ces estimateurs, le CvM est celui qui engendre le plus d'estimations  $\hat{\mu}$  négatives. Bien que ces deux versions de l'estimateur CvM donnent des résultats analogues sur données simulées et données contaminées, ce n'est plus le cas sur données réelles.

Estimateur	B-robustesse	Nb $\hat{\mu} < 0$	Nb $F_{\hat{\theta}}(H) > 50\%$	Nb Tests "OK"
GMM	Non	1	6	5
ML	Non	5	10	21
OBRE $c = \sqrt{2}$	Oui	6	10	15
OBRE $c = 2$	Oui	5	10	19
OBRE $c = 5$	Oui	5	10	21
QD "scindé"	Oui	0	2	7
QD "direct"	Oui	5	10	20
CvM "normal"	Oui	7	10	19
CvM "seuil"	Oui	12	12	4

Table 7.1 – Tableau comparatif des résultats sur données réelles

On note que l'estimateur ML conduit également à des estimations négatives de  $\mu$  (5 sur 14). L'OBRE étant dérivé de l'estimateur ML, lorsque l'estimation de  $\mu$  est négative, elle l'est aussi pour l'OBRE, même si sa valeur est atténuée en valeur absolue. De plus, l'estimation n'est pas linéaire en  $c$ . En effet, alors que cette linéarité est observable sur données simulées et contaminées, on peut voir ici, pour l'échantillon E4, des estimations du couple  $(\hat{\mu}, \hat{\sigma})$  valant (7.47, 3.12), (8.86, 0.62) et (8.41, 1.74) pour les valeurs respectives  $c = \sqrt{2}$ , 2 et 5.

L'estimateur QD par optimisation directe souffre du même problème, 5 des estimations de  $\mu$  étant négatives. Les estimateurs GMM et QD par optimisation scindée semblent le pallier, puisque une seule estimation est inférieure à 0.

Cependant, c'est pour ces deux estimateurs (CvM avec prise en compte des données sous le seuil mis à part) que l'on observe le moins de tests d'adéquation concluant : 7 pour le QD et seulement 5 pour la méthode GMM, la majorité étant le test AD2up qui, au vu des résultats de ces tests, est le plus simple à satisfaire. A l'inverse, le nombre de tests concluants pour les autres estimateurs est plus satisfaisant : 21 pour la méthode ML, 20 pour l'estimateur QD par optimisation directe, 19 pour le CvM avec prise en compte normale du seuil, et ce nombre décroît pour l'estimateur OBRE en fonction de sa robustesse, 21, 19 et 15 pour les valeurs respectives  $c = 5$ , 2 et  $\sqrt{2}$ .

Ces résultats se résument dans le tableau 7.1.

Ces résultats mettent en relief la complexité du choix de l'estimateur. D'un côté, nous disposons d'estimateurs pour lesquels les tests d'adéquations sont satisfaisants, mais dont les modèles estimés n'ont pas d'interprétation métier. De l'autre, c'est l'inverse, les méthodes estiment des paramètres pour lesquels le modèle est appréhendable, mais qui ne satisfont pas ces mêmes tests d'adéquation.

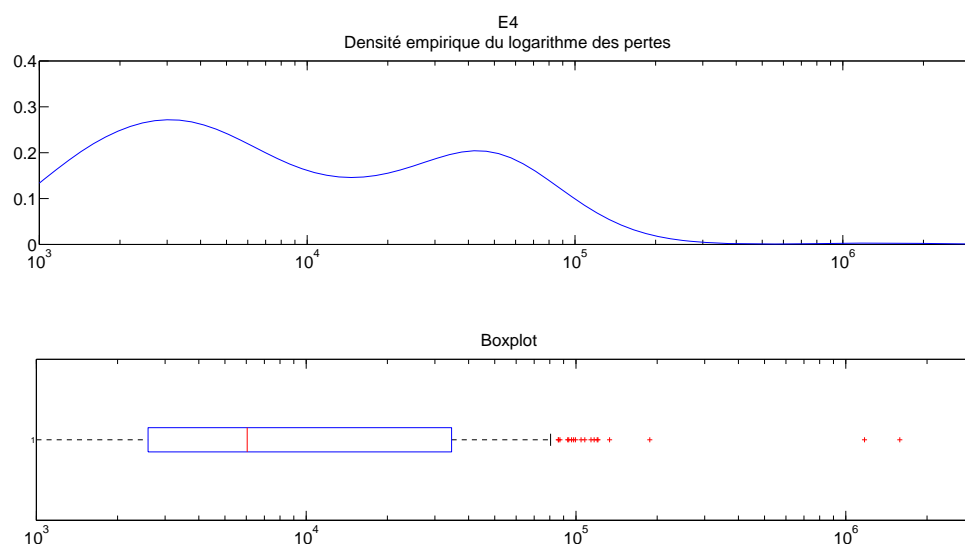
Dans le prochain paragraphe, nous présentons plus précisément trois des différents échantillons afin de comprendre pourquoi l'estimation est si complexe.

## 7.2 Focus sur trois échantillons

Les échantillons que nous présentons ici sont le reflet de la variété des données observables dans la pratique. Pour chacun d'eux, nous présentons des caractéristiques générales de l'échantillon, sa fonction de densité empirique et le diagramme en boîte associé, représentant la médiane, les quantiles à 25 et 75%, les points extrêmes (en pointillés), ainsi que les points considérés aberrants par Matlab (tracé individuellement en rouge). Nous présentons également les estimations obtenues pour chacun des estimateurs.

## Echantillon E4

	Nombre	Maximum	moyenne	médiane	écart-type	skewness	kurtosis
E4	529	1 585 394	25 258	6 046	88 355	15	239



Estimateur	$\hat{\mu}$	$\hat{\sigma}$	$F_{\hat{\sigma}}(H)$	KS	CVM	AD2	AD2up	CaR
GMM	8,60	1,66	15%	0	0	0	0,012	9,84E+06
ML	8,22	1,94	25%	0	0	0	0,021	2,09E+07
OBRE $c = \sqrt{2}$	7,47	3,12	43%	0	0	0	0,006	1,70E+09
OBRE $c = 2$	8,86	0,62	0%	0	0	0	0	1,26E+06
OBRE $c = 5$	8,41	1,74	19%	0	0	0	0,026	1,09E+07
QD "scindé"	8,65	1,49	12%	0	0	0	0,007	5,81E+06
QD "direct"	8,66	1,47	12%	0	0	0	0,008	5,53E+06
CvM "normal"	7,30	2,60	44%	0	0	0	0,041	1,46E+08
CvM "seuil"	-677,87	79,78	100%	-	-	-	-	-

Cet échantillon est un cas atypique puisqu'on peut observer 2 modes distincts sur le tracé de la densité empirique. Or, la loi log-normale n'ayant qu'un seul mode, on comprend que l'estimation délicate, ce qui est constaté en regardant les couples  $(\hat{\mu}, \hat{\sigma})$  qui diffèrent significativement d'une méthode d'estimation à l'autre, notamment pour l'OBRE suivant les trois valeurs de  $c$ .

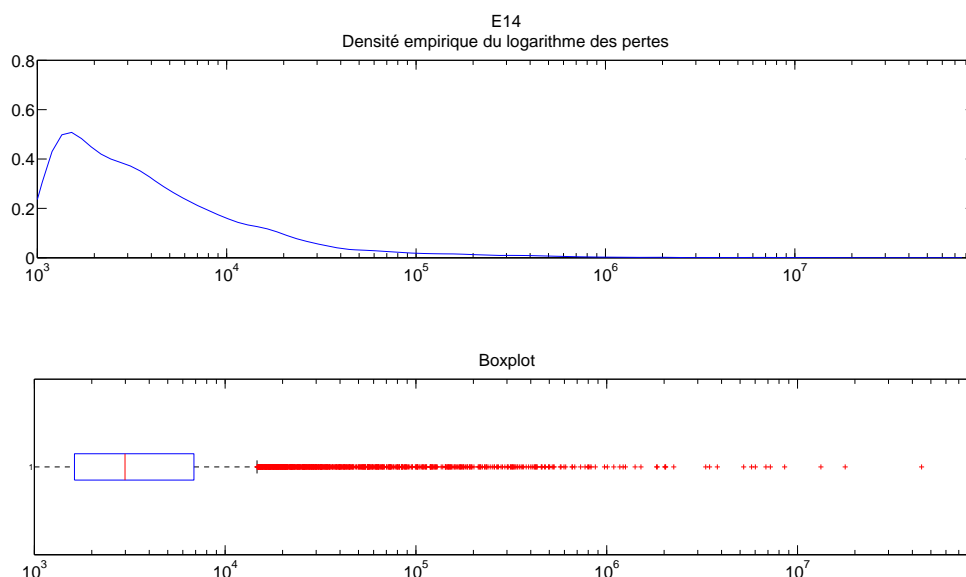
Etant données les deux modalités observées, le modèle log-normale apparaît inadapté. Ce qui est d'ailleurs confirmé par les tests d'adéquation, car toutes les  $p$ -value sont nulles quelque soit l'estimateur. Une alternative permettant de modéliser ce genre de sévérité serait un modèle mélange qui se définit comme suit

$$\begin{cases} X \sim \mathcal{LN}(\mu_1, \sigma_1) & \text{avec probabilité } p \\ X \sim \mathcal{LN}(\mu_2, \sigma_2) & \text{avec probabilité } (1 - p) \end{cases}$$



## Echantillon E14

	Nombre	Maximum	moyenne	médiane	écart-type	skewness	kurtosis
E14	7 474	44 703 038	32 393	2 986	616 862	56	3 813



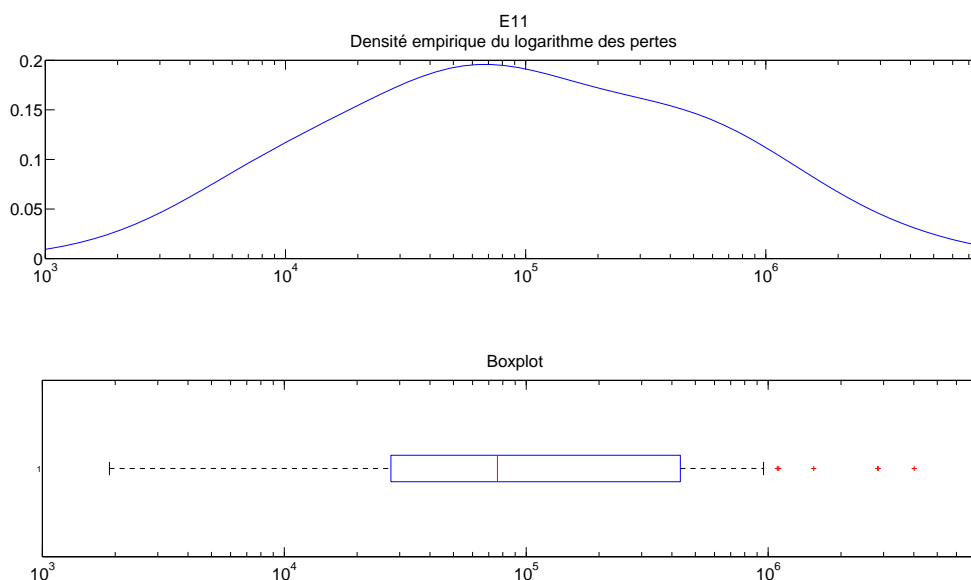
Estimateur	$\hat{\mu}$	$\hat{\sigma}$	$F_{\hat{\theta}}(H)$	KS	CVM	AD2	AD2up	CaR
GMM	5,76	2,65	67%	0	0	0	0,003	2,50E+08
ML	2,64	3,06	92%	0	0	0	0,004	1,84E+08
$c = \sqrt{2}$	5,61	2,17	72%	0	0	0	0	3,32E+07
$c = 2$	4,62	2,51	82%	0	0	0	0	6,32E+07
$c = 5$	2,72	3,04	92%	0	0	0	0,01	1,78E+08
QD "scindé"	7,25	1,77	42%	0	0	0	0	2,72E+07
QD "direct"	-7,88	4,79	100%	0	0	0	0,025	-
CvM "normal"	5,87	2,09	69%	0	0	0	0	3,06E+07
CvM "seuil"	-815,74	98,05	100%	-	-	-	-	-

Cet échantillon est en revanche un cas plus couramment rencontré dans la pratique. Ici, la distribution est extrêmement sévère, la queue de distribution étant très épaisse. Sur cet exemple, on peut se rendre compte de l'impact de la valeur  $c$  de l'OBRE régissant la robustesse, puisque les estimations des paramètres obtenus est de (5.61,2.17) et (4.62,2.51) pour  $c = \sqrt{2}$  et 2 respectivement, et de (2.64,3.06) pour l'estimateur ML.

Cependant, pour ce type d'échantillon, le modèle log-normale semble impropre car celui-ci est à queue fine. Comme précédemment, ceci est confirmé par les tests d'adéquation qui rejette unanimement le modèle estimé. Etant donné ce profil très sévère, une distribution à queue épaisse serait plus adéquate, comme la loi de Pareto. On peut également penser à des modèles composites qui permettent de modéliser une loi pour le corps et une loi différente pour la queue, la coupure se faisant à partir d'un certain seuil  $u$ . On pourrait alors obtenir des modèles de la forme  $\mathcal{C}(\mathcal{LN}(\mu, \sigma), \mathcal{P}(\alpha, \beta), u)$ .

## Echantillon E11

	Nombre	Maximum	moyenne	médiane	écart-type	skewness	kurtosis
E11	89	4 013 108	333 290	76 030	636 494	4	19



Estimateur	$\hat{\mu}$	$\hat{\sigma}$	$F_{\hat{\theta}}(H)$	KS	CVM	AD2	AD2up	CaR
GMM	11,95	1,24	0%	0	0	0	0,136	2,56E+07
ML	11,42	1,72	0%	0,155	0,205	0,207	0,49	7,97E+07
c = 1,41	11,38	1,85	1%	0,331	0,321	0,26	0,343	1,24E+08
c = 2	11,45	1,91	1%	0,412	0,253	0,162	0,201	1,63E+08
c = 5	11,43	1,74	0%	0,189	0,214	0,216	0,471	8,34E+07
QD scindé	11,39	1,71	0%	0,07	0,157	0,141	0,424	7,35E+07
QD direct	11,39	1,71	0%	0,132	0,243	0,253	0,401	7,27E+07
CvM normal	11,37	1,87	1%	0,158	0,16	0,266	0,464	1,28E+08
CvM seuil	11,37	1,87	1%	0,386	0,279	0,317	0,364	1,34E+08

Nous présentons également cet échantillon car il est le seul de l'ensemble de notre panel pour lequel les estimations des paramètres par chacun des 10 estimateurs sont comparables bien qu'il n'y ait que 89 observations. Ceci s'explique par le fait que sa fonction de densité empirique à l'échelle logarithmique est approchante d'une loi normale, et que le mode de cette densité n'est pas tronqué par le seuil de collecte.

Ce focus permet de rendre compte de la diversité des échantillons sur lesquels nous travaillons. Il suggère également que la loi log-normale n'est pas la plus adaptée à certains profils et que d'autres alternatives doivent être examinées.

# Conclusions

---

L'application de méthodes d'estimation robustes à la sévérité des risques opérationnels a fait l'objet de nombreuses études récemment. Ce type d'estimation permet en effet de prendre en compte l'hétérogénéité des données, chose qui n'est pas possible avec les méthodes d'estimation standards.

Dans ce mémoire, nous mettons en avant le fait que l'estimation des paramètres de la distribution de sévérité des risques opérationnels est un problème très délicat. L'hétérogénéité des données complexifie grandement l'estimation, et les méthodes d'estimation standards, que sont le maximum de vraisemblance et la méthode des moments généralisée, apparaissent inadaptées.

En effet, après avoir présenté la fonction d'influence, outil analytique fondamental de la théorie de la statistique robuste, qui permet à partir de déductions graphiques de juger de la robustesse des estimateurs, nous constatons effectivement la non robustesse de ces deux estimateurs standards. En adéquation avec les directives du Comité de Bâle, nous présentons ensuite trois estimateurs robustes, les estimateurs OBRE, CvM et QD.

Le premier, l'estimateur OBRE, qui est une version robuste de l'estimateur ML, est celui qui a demandé le plus de travail, aussi bien sur le plan de la théorie, que sur celui de l'implémentation et du calibrage des différents paramètres du modèle. Concernant les deux autres estimateurs, robustes par construction, la prise en compte de la troncature à gauche des données a donné lieu à deux versions distinctes par estimateurs. On constate, pour l'ensemble de ces paramétrages, la robustesse locale de ces estimateurs en se basant sur l'étude de la fonction d'influence.

Par ailleurs, il est apparu opportun de vérifier la robustesse globale de ces estimateurs, en mettant en place des tests de contamination à plus grande échelle afin de se rapprocher des conditions observables dans un échantillon de pertes réelles particulièrement hétérogènes. Enfin, afin d'être le plus exhaustif possible, tous ces résultats sont présentés pour différents profils de risque théoriques et pour différents tests de contamination.

Si les tests de robustesse globale sont concluants pour les estimateurs robustes implémentés, leur application à des données réelles apporte des résultats mitigés. Ainsi, nous identifions deux caractères bien distincts des estimateurs : les estimateurs validant les tests d'adéquation mais donnant des estimations de paramètres aberrantes d'un point de vue métier, et les estimateurs pour lesquels les paramètres estimés sont appréhendables mais avec un ajustement aux données médiocre.

Cependant, une étude plus poussée de ces données de production suggère que la loi log-normale, bien qu'étant la plus appliquée par les institutions bancaires, n'est pas appropriée à certains types de données. Une piste à poursuivre serait alors de s'intéresser à des modèles plus complexes mais plus flexibles comme des modèles mélanges ou des modèles composites.

Mais alors, la mise en place de méthodes d'estimation spécifiquement robustes s'en trouverait considérablement complexifiée. De même, pour certains échantillons, le recours aux distributions issues de la théorie des valeurs extrêmes. L'estimation de leurs paramètres et en particulier de l'indice de valeurs extrêmes est alors très sensible et peut conduire à des niveaux de capital aberrants.

De manière générale, ces travaux viennent contribuer aux nombreuses études dédiées à la modélisation de la sévérité des risques opérationnels, qu'il s'agisse des nombreuses distributions de probabilité envisageables ou des méthodes à mettre en place pour estimer leurs paramètres. Cela demeure un sujet délicat, pour lequel aucun modèle générique ne parvient à faire consensus au sein des établissements bancaires.

Ces constats soulèvent la légitimité d'un autre élément fondamental : le niveau de quantile (99.9%) imposé pour déterminer le capital. Il semble en effet particulièrement difficile de fonder la quantification du capital réglementaire à ce niveau de précision quand on constate la complexité de l'estimation de la sévérité.

C'est en raison de ces différents challenges que se sont ouvertes depuis quelques mois des discussions entre le Comité de Bâle et les fédérations bancaires internationales afin de modifier en profondeur l'approche avancée. Le niveau de quantile utilisé pourrait ainsi être abaissé (et compensé par des coefficients multiplicatifs pour maintenir un niveau de capital équivalent), et surtout des exigences plus prescriptives pourraient être formulées pour s'appliquer au modèle interne.

# Références bibliographiques

---

- [1] Bank of International Settlements (2011), *Operational Risk - Supervisory Guidelines for the Advanced Measurement Approaches*.
- [2] Chernobai, A., Rachev, S.T., and Fabozzi, F.J. (2012), *Composite goodness-of-fit tests for left-truncated loss samples* in Handbook of Financial Econometrics and Statistics, C.-F. Lee (ed), Springer.
- [3] Chernobai, A., and Rachev, S.T. (2006), *Applying robust methods to operational risk modeling*, Technical report.
- [4] Dupuis, D.J. (1998) *Exceedances over high thresholds : a guide to threshold selection*, Technical report.
- [5] Ergashev, B. (2008), *Should Risk Managers Rely on Maximum Likelihood Estimation Method While Quantifying Operational Risk ?*, The Journal of Operational Risk, 3(2), 63-86.
- [6] Frachot, A., Georges, P. and Roncalli, T. (2001), *Loss distribution approach for operational risk*, Crédit Lyonnais, Groupe de Recherche Opérationnelle, Working paper.
- [7] Frachot, A., Moudoulaud O. and Roncalli, T. (2003), *Loss distribution approach in practice*, in M. Ong (ed.), The Basel Handbook : A Guide for Financial Practitioners, Risk Books.
- [8] Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A. (1986), *Robust statistics : the approach based on influence functions*, Wiley.
- [9] Huber, P. (1981), *Robust statistics*, Wiley.
- [10] Klugman, S.A., H.H. Panjer and G.E. Willmot (1998), *Loss Models : From Data to Decisions*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons.
- [11] LaRiccia, V. N. (1982), *Asymptotic properties of weighted  $L^2$  quantile distance estimators*, The Annals of Statistics, 10, 621-62.
- [12] LaRiccia, V. N., and Wehrly, T. E. (1982), *Asymptotic properties of a family of minimum quantile distance estimators*, Journal of the American Statistical Association, 80, 742-747.
- [13] Lavaud, S., and Leherissé, V. (2014) *Goodness-of-fit tests and selection methods for operational risk*, The Journal of Operational Risk, 9(3), 1-30.
- [14] Leherissé, V. and Renaudin, A. (2013), *Quantile distance estimation for operational risk : a practical application*, The Journal of Operational Risk, 8(2), 73-102.
- [15] Opdyke J.D., Cavallo A. (2012), *Estimating operational risk capital : the challenges of truncation, the hazards of maximum likelihood estimation, and the promise of robust statistics*, The Journal of Operational Risk.
- [16] Opdyke J.D., Cavallo A. (2012), *Operational risk capital estimation and planning : exact sensitivity analysis and business decision making using the influence function*, Chapter submission for Operational risk : new frontiers explored.
- [17] Roncalli, T. (2009), *La Gestion des risques Financiers*, Economica.
- [18] Ruckdeschel, P., Horbenko, N., Bae, T. (2011), *Robust estimation of operational risk*, Journal of operational risk.
- [19] Ruckdeschel, P., Horbenko, N. (2011), *Optimally robust estimators in generalized pareto models*, Research article.
- [20] Tukey, J. W. (1960), A survey of sampling from contaminated distributions. *Contribution to probability and statistics*, Stanford University press.
- [21] Victoria-Feser, M.P. (2000), *Robust income estimation with missing data*, Discussion paper.

# **Annexes**

# A Calculs de fonctions d'influence

---

## IF de la moyenne

La fonctionnelle associée à l'estimateur de la moyenne est donnée par  $T(F) = \int x \, dF(x) = \mu$ . Comme celle-ci est linéaire, le calcul de la fonction d'influence se fait de la façon suivante

$$\begin{aligned} \text{IF}(x; F_\theta, T) &= \lim_{\varepsilon \rightarrow 0} \frac{T((1-\varepsilon)F_\theta + \varepsilon\Delta_x) - T(F_\theta)}{\varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{(1-\varepsilon)T(F_\theta) + \varepsilon T(\Delta_x) - T(F_\theta)}{\varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{(1-\varepsilon)\mu + \varepsilon x - \mu}{\varepsilon} \\ &= x - \mu \end{aligned}$$

## IF de la médiane

Pour la médiane, prenons le cas plus général de l'estimateur du quantile d'ordre  $s$ ,  $T_s(F_\theta) = F_\theta^{-1}(s)$  ou encore  $T_s(F_{\theta,\varepsilon}) = F_{\theta,\varepsilon}^{-1}(s)$  avec une distribution contaminée.

Partons de l'identité

$$F_{\theta,\varepsilon}(F_{\theta,\varepsilon}^{-1}(s)) = s$$

L'idée est maintenant de dériver cette expression par rapport à  $\varepsilon$ . On rappelle que, si  $f$  et  $g$  sont deux fonctions à deux paramètres  $(x, t)$ , alors on obtient par la règle de la chaîne

$$\frac{\partial}{\partial t} [g(t, f(t, x))] = \frac{\partial f(t, x)}{\partial t} \times \frac{\partial g(t, x)}{\partial x} + \frac{\partial g(t, f(t, x))}{\partial t}$$

En appliquant ce résultat à notre identité de départ, on obtient

$$\begin{aligned} \frac{\partial}{\partial \varepsilon} [F_{\theta,\varepsilon}(F_{\theta,\varepsilon}^{-1}(s))] &= \frac{\partial T_s(F_{\theta,\varepsilon})}{\partial \varepsilon} \times \left[ (1-\varepsilon) \frac{\partial F_\theta}{\partial x} + \varepsilon \frac{\partial G}{\partial x} \right] \circ (F_{\theta,\varepsilon}^{-1}(s)) + [G - F_\theta] \circ (F_{\theta,\varepsilon}^{-1}(s)) \\ &=_{\varepsilon=0} \frac{\partial T_s(F_\theta)}{\partial \varepsilon} \times f(F_\theta^{-1}(s)) + G(F_\theta^{-1}(s)) - F_\theta(F_\theta^{-1}(s)) \end{aligned}$$

et d'autre part  $\frac{\partial}{\partial \varepsilon} [F_{\theta,\varepsilon}(F_{\theta,\varepsilon}^{-1}(s))] = \frac{\partial s}{\partial \varepsilon} = 0$ , ce qui nous permet d'obtenir

$$\frac{\partial T_s(F_\theta)}{\partial \varepsilon} = \frac{s - G(F_\theta^{-1}(s))}{f(F_\theta^{-1}(s))}$$

En prenant  $G = \Delta_x$ , cela donne la fonction d'influence de  $T_s$  au point  $x$

$$\text{IF}(x; F_\theta, T_s) = \begin{cases} \frac{s-1}{f(F_\theta^{-1}(s))} & \text{pour } x < F_\theta^{-1}(s) \\ \frac{s}{f(F_\theta^{-1}(s))} & \text{pour } x > F_\theta^{-1}(s) \end{cases}$$

Ainsi, comme la médiane correspond au quantile d'ordre 0.5, sa fonction d'influence est donnée par

$$\text{IF}(x; F_\theta, \text{médiane}) = \begin{cases} \frac{-1}{2f(m)} & \text{pour } x < m \\ \frac{1}{2f(m)} & \text{pour } x > m \end{cases}$$

## Propriétés des M-Estimateurs - Démonstrations

### Calcul de la fonction d'influence

On rappelle que l'on veut montrer que le résultat de l'équation (4.10).

On commence par remplacer  $F_\theta$  dans (4.8) par sa distribution contaminée puis on différencie par rapport à  $\varepsilon$ .

$$\begin{aligned} \frac{\partial}{\partial \varepsilon} \left[ \int_{\mathcal{X}} \psi(y, T(F_{\theta, \varepsilon})) F_{\theta, \varepsilon}(dy) \right] &= \int_{\mathcal{X}} \frac{\partial \psi(y, T(F_{\theta, \varepsilon}))}{\partial \varepsilon} F_{\theta, \varepsilon}(dy) + \int_{\mathcal{X}} \psi(y, T(F_{\theta, \varepsilon})) \frac{\partial F_{\theta, \varepsilon}(dy)}{\partial \varepsilon} \\ &= \int_{\mathcal{X}} \frac{\partial T(F_{\theta, \varepsilon})}{\partial \varepsilon} \frac{\partial \psi(y, T(F_{\theta, \varepsilon}))}{\partial \theta} F_{\theta, \varepsilon}(dy) \\ &\quad + \int_{\mathcal{X}} \psi(y, T(F_{\theta, \varepsilon})) [G - F_\theta](dy) \\ &= 0 \quad \text{d'après (4.8)} \end{aligned}$$

En prenant maintenant  $G = \Delta_x$  et  $\varepsilon = 0$ , on obtient

$$\text{IF}(x; F_\theta, T) \int_{\mathcal{X}} \frac{\partial \psi(y, T(F_\theta))}{\partial \theta} F_\theta(dy) - \underbrace{\int_{\mathcal{X}} \psi(y, T(F_\theta)) F_\theta(dy)}_{=0} + \psi(x, T(F_\theta)) = 0$$

On en déduit

$$\text{IF}(x; F_\theta, T) = \frac{\psi(x, T(F_\theta))}{-\int_{\mathcal{X}} \frac{\partial \psi(y, T(F_\theta))}{\partial \theta} dF_\theta(y)}$$

### Variance asymptotique

On veut démontrer ici le résultat de l'équation (4.11).

On utilisera pour cette démonstration les trois théorèmes énoncés plus haut dans cette même annexe.

D'après (4.9),  $\sum_{i=1}^n \psi(x_i, \hat{\theta}) = 0$ . En utilisant un développement en séries de Taylor au voisinage de  $\theta$ , on obtient

$$\begin{aligned} \sum_{i=1}^n \psi(x_i, \hat{\theta}) &= \sum_{i=1}^n \psi(x_i, \theta) + (\hat{\theta} - \theta) \sum_{i=1}^n \frac{\partial \psi(x_i, \theta)}{\partial \theta} + O(1) \\ \sqrt{n} \times \frac{1}{n} \sum_{i=1}^n \psi(x_i, \hat{\theta}) &= \sqrt{n} \times \frac{1}{n} \sum_{i=1}^n \psi(x_i, \theta) + \sqrt{n} (\hat{\theta} - \theta) \times \frac{1}{n} \sum_{i=1}^n \frac{\partial \psi(x_i, \theta)}{\partial \theta} + O(1/\sqrt{n}) \end{aligned}$$

Comme  $\mathbb{E}[\psi(X, \theta)] = 0$ , par le théorème central limite, on a

$$\sqrt{n} \times \left( \frac{1}{n} \sum_{i=1}^n \psi(x_i, \theta) \right) \xrightarrow[n \rightarrow \infty]{d} Z \sim \mathcal{N}(0, \xi^2) \quad (\text{A.1})$$

avec

$$\xi^2 = \mathbb{V}[\psi(X, \theta)] = \mathbb{E}[\psi(X, \theta)^2] \quad (\text{A.2})$$



De plus, d'après la loi faible des grands nombres, on a

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial \psi(x_i, \theta)}{\partial \theta} \xrightarrow[n \rightarrow \infty]{p} \mathbb{E} \left[ \frac{\partial \psi(X, \theta)}{\partial \theta} \right] \quad (\text{A.3})$$

Puis, en utilisant le théorème de Slutsky, on obtient

$$\sqrt{n} (\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{d} \frac{-Z}{\mathbb{E} \left[ \frac{\partial \psi(X, \theta)}{\partial \theta} \right]} \sim \mathcal{N}(0, \sigma^2) \quad (\text{A.4})$$

où

$$\sigma^2 = \frac{\xi^2}{\mathbb{E} \left[ \frac{\partial \psi(X, \theta)}{\partial \theta} \right]} \quad (\text{A.5})$$

or, d'après (4.10)

$$\begin{aligned} \mathbb{V} [\text{IF}(X; F_\theta, T)] &= \mathbb{V} \left[ -\psi(x, \theta) \middle/ \int_{\mathcal{X}} \frac{\partial \psi(y, \theta)}{\partial \theta} dF_\theta(y) \right] \\ &= \mathbb{E} \left[ \left( -\psi(X, \theta) \middle/ \mathbb{E} \left[ \frac{\partial \psi(X, \theta)}{\partial \theta} \right] \right)^2 \right] - \mathbb{E} \left[ -\psi(X, \theta) \middle/ \mathbb{E} \left[ \frac{\partial \psi(X, \theta)}{\partial \theta} \right] \right]^2 \\ &= \mathbb{E} [\psi(X, \theta)^2] \middle/ \mathbb{E} \left[ \frac{\partial \psi(X, \theta)}{\partial \theta} \right]^2 - \left( \mathbb{E} [-\psi(X, \theta)] \middle/ \mathbb{E} \left[ \frac{\partial \psi(X, \theta)}{\partial \theta} \right] \right)^2 \\ &= \mathbb{E} [\psi(X, \theta)^2] \middle/ \mathbb{E} \left[ \frac{\partial \psi(X, \theta)}{\partial \theta} \right]^2 \\ &= \sigma^2 \\ &= \mathbb{E} [\text{IF}(X; F_\theta, T)^2] \\ &= \int_{\mathcal{X}} \text{IF}(x; F_\theta, T)^2 F_\theta(dx) \end{aligned}$$

Ce qui nous donne bien le résultat (4.11).

## Calculs complémentaires pour la loi log-normale

### Loi log-normale non tronquée

Nous aurons besoin dans la suite des différentes dérivées premières et secondes de la densité log-normale  $f(x; \mu, \sigma) =$

$$\frac{1}{\sqrt{2\pi}\sigma x} \exp \left\{ -\frac{1}{2} \left( \frac{\ln x - \mu}{\sigma} \right)^2 \right\} :$$

$$\left. \begin{aligned} \frac{\partial f}{\partial \mu} &= \frac{\ln x - \mu}{\sigma^2} f(x; \mu, \sigma) \\ \frac{\partial f}{\partial \sigma} &= \left( \frac{(\ln x - \mu)^2}{\sigma^3} - \frac{1}{\sigma} \right) f(x; \mu, \sigma) \\ \frac{\partial^2 f}{\partial \mu^2} &= \left( \frac{(\ln x - \mu)^2}{\sigma^4} - \frac{1}{\sigma^2} \right) f(x; \mu, \sigma) \\ \frac{\partial^2 f}{\partial \sigma^2} &= \left( \frac{2}{\sigma^2} - 5 \frac{(\ln x - \mu)^2}{\sigma^4} + \frac{(\ln x - \mu)^4}{\sigma^6} \right) f(x; \mu, \sigma) \\ \frac{\partial^2 f}{\partial \mu \partial \sigma} &= \left( \frac{(\ln x - \mu)^3}{\sigma^5} - 3 \frac{\ln x - \mu}{\sigma^3} \right) f(x; \mu, \sigma) = \frac{\partial^2 f}{\partial \sigma \partial \mu} \end{aligned} \right\} \quad (\text{A.6})$$

Le vecteur de score s'obtient naturellement comme étant

$$\psi_\theta = \begin{pmatrix} -\frac{1}{f} \frac{\partial f}{\partial \mu} \\ -\frac{1}{f} \frac{\partial f}{\partial \sigma} \end{pmatrix} = \begin{pmatrix} \frac{\mu - \ln x}{\sigma^2} \\ \frac{1}{\sigma} - \frac{(\mu - \ln x)^2}{\sigma^3} \end{pmatrix}$$

Comme  $\psi'_\theta(x, \theta) = \left( -\frac{\partial^2 f(x, \theta)}{\partial \theta^2} f(x, \theta) + \left( \frac{\partial f(x, \theta)}{\partial \theta} \right)^2 \right) / [f(x, \theta)]^2$ , le calcul des éléments de la matrice  $\mathcal{A}(\theta)$  se fait assez facilement en utilisant des intégrations par partie ainsi que des changements de variables. Ainsi, en se servant des dérivées de  $f$  (A.6).

$$\begin{aligned} - \int_0^\infty \frac{\partial \psi_\mu}{\partial \mu} dF(x) &= - \int_0^\infty \frac{-\frac{\partial^2 f(x)}{\partial \mu^2} f(x) + \left( \frac{\partial f(x)}{\partial \mu} \right)^2}{f(x)^2} dF(x) \\ &= - \int_0^\infty \left( -\frac{(\ln x - \mu)^2}{\sigma^4} + \frac{1}{\sigma^2} + \frac{(\ln x - \mu)^2}{\sigma^4} \right) dF(x) \\ &= -\frac{1}{\sigma^2} \\ - \int_0^\infty \frac{\partial \psi_\sigma}{\partial \sigma} dF(x) &= - \int_0^\infty \frac{-\frac{\partial^2 f(x)}{\partial \sigma^2} f(x) + \left( \frac{\partial f(x)}{\partial \sigma} \right)^2}{[f(x)]^2} dF(x) \\ &= - \int_0^\infty \left[ -\left( \frac{2}{\sigma^2} - 5 \frac{(\ln x - \mu)^2}{\sigma^4} + \frac{(\ln x - \mu)^4}{\sigma^6} \right) + \left( \frac{(\ln x - \mu)^2}{\sigma^3} - \frac{1}{\sigma} \right)^2 \right] f(x) dx \\ &= \frac{1}{\sigma^2} - \frac{3}{\sigma^2} \int_0^\infty \left( \frac{\ln x - \mu}{\sigma} \right)^2 \frac{1}{\sqrt{2\pi}\sigma x} \exp \left\{ -\frac{1}{2} \left( \frac{\ln x - \mu}{\sigma} \right)^2 \right\} dx \\ &= \frac{1}{\sigma^2} - \frac{3}{\sigma^2} \int_{-\infty}^\infty u^2 e^{-u^2/2} \quad \text{en posant} \quad u = \frac{\ln x - \mu}{\sigma} \\ &= -\frac{2}{\sigma^2} \end{aligned}$$

car la dernière intégrale, qui correspond au moment d'ordre 2 d'une loi normale centrée réduite, vaut 1.

$$\begin{aligned} - \int_0^\infty \frac{\partial \psi_\mu}{\partial \sigma} dF(x) &= - \int_0^\infty \frac{-\frac{\partial^2 f(x)}{\partial \mu \partial \sigma} f(x) + \frac{\partial f(x)}{\partial \mu} \frac{\partial f(x)}{\partial \sigma}}{[f(x)]^2} dF(x) \\ &= - \int_0^\infty \left[ -\left( \frac{(\ln x - \mu)^3}{\sigma^5} - 3 \frac{\ln x - \mu}{\sigma^3} \right) + \frac{\ln x - \mu}{\sigma^2} \left( \frac{(\ln x - \mu)^2}{\sigma^3} - \frac{1}{\sigma} \right) \right] f(x) dx \\ &= \frac{2}{\sigma^2} \int_0^\infty \left( \frac{\ln x - \mu}{\sigma} \right) \frac{1}{\sqrt{2\pi}\sigma x} \exp \left\{ -\frac{1}{2} \left( \frac{\ln x - \mu}{\sigma} \right)^2 \right\} dx \\ &= \frac{2}{\sigma^2} \int_{-\infty}^\infty u e^{-u^2/2} \quad \text{en posant} \quad u = \frac{\ln x - \mu}{\sigma} \\ &= 0 \end{aligned}$$

car la dernière intégrale, qui correspond au moment d'ordre 1 d'une loi normale centrée réduite, vaut 0.

Ainsi :

$$\mathcal{A}(\theta) = \begin{pmatrix} -\frac{1}{\sigma^2} & 0 \\ 0 & -\frac{2}{\sigma^2} \end{pmatrix}$$

## Loi log-normale tronquée

L'expression de la fonction de score s'obtient à partir de (4.15) et du vecteur de score pour une loi log-normale non tronquée (4.12).

$$\psi_\theta = \begin{pmatrix} -\frac{\ln x - \mu}{\sigma^2} - \frac{\int_0^H \frac{\ln y - \mu}{\sigma^2} f(y, \theta) dy}{1 - F(H, \theta)} \\ -\left(\frac{(\ln x - \mu)^2}{\sigma^3} - \frac{1}{\sigma}\right) - \frac{\int_0^H \left(\frac{(\ln y - \mu)^2}{\sigma^3} - \frac{1}{\sigma}\right) f(y, \theta) dy}{1 - F(H, \theta)} \end{pmatrix} \quad (\text{A.7})$$

Pour obtenir l'expression de la matrice  $\mathcal{A}(\theta)$ , nous avons besoin de la formule de la dérivée de la fonction de score donnée par l'équation (4.17) et des formules de dérivées premières et secondes de la dérivée de la densité données plus haut.

Reprenons d'abord la formule de la dérivée du score (4.16) appliquée à la dérivée seconde en  $\mu$  :

$$\begin{aligned} \frac{\partial \psi_\mu}{\partial \mu} &= \frac{1}{[f(y, \theta)]^2} \left( \left( \frac{\partial f(y, \theta)}{\partial \mu} \right)^2 - \frac{\partial^2 f(y, \theta)}{\partial \mu^2} f(y, \theta) \right) \\ &\quad - \frac{1}{[1 - F(H, \theta)]^2} \left( \left( \frac{\partial F(H, \theta)}{\partial \mu} \right)^2 + \frac{\partial^2 F(H, \theta)}{\partial \mu^2} (1 - F(H, \theta)) \right) \end{aligned}$$

en intégrant sur  $[H, \infty)$

$$\begin{aligned} - \int_H^\infty \frac{\partial \psi_\mu}{\partial \mu} d\widetilde{F}_\theta(y) &= - \int_H^\infty \frac{1}{[f(y, \theta)]^2} \left( \left( \frac{\partial f(y, \theta)}{\partial \mu} \right)^2 - \frac{\partial^2 f(y, \theta)}{\partial \mu^2} f(y, \theta) \right) d\widetilde{F}_\theta(y) \\ &\quad + \frac{1}{[1 - F(H, \theta)]^2} \left( \left( \frac{\partial F(H, \theta)}{\partial \mu} \right)^2 + \frac{\partial^2 F(H, \theta)}{\partial \mu^2} (1 - F(H, \theta)) \right) \end{aligned}$$

en utilisant les expressions des dérivées données en (A.6) et en explicitant les dérivées de  $F$

$$\begin{aligned} - \int_H^\infty \frac{\partial \psi_\mu}{\partial \mu} d\widetilde{F}_\theta(y) &= - \int_H^\infty \frac{1}{[f(y, \theta)]^2} \left[ \left( \frac{\ln y - \mu}{\sigma^2} f(y; \mu, \sigma) \right)^2 \right. \\ &\quad \left. - \left( \frac{(\ln y - \mu)^2}{\sigma^4} - \frac{1}{\sigma^2} \right) f(y; \mu, \sigma)^2 \right] d\widetilde{F}_\theta(y) \\ &\quad + \frac{1}{[1 - F(H, \theta)]^2} \left[ \left( \int_0^H \frac{\partial f(y, \theta)}{\partial \mu} dF_\theta(y) \right)^2 \right. \\ &\quad \left. + (1 - F(H, \theta)) \int_0^H \frac{\partial^2 f(y, \theta)}{\partial \mu^2} dF_\theta(y) \right] \end{aligned}$$

Finalement, il reste

$$\begin{aligned}
-\int_H^\infty \frac{\partial \psi_\mu}{\partial \mu} d\tilde{F}(y) &= -\frac{1}{\sigma^2} + \frac{1}{[1 - F(H, \theta)]^2} \left[ \int_0^H \frac{\ln y - \mu}{\sigma^2} f(y, \theta) dy \right]^2 \\
&\quad + \frac{1}{1 - F(H, \theta)} \int_0^H \left( \frac{(\ln y - \mu)^2}{\sigma^4} - \frac{1}{\sigma^2} \right) f(y, \theta) dy
\end{aligned} \tag{A.8}$$

De la même manière, on trouve les expressions des autres éléments de la matrice, à savoir

$$\begin{aligned}
-\int_H^\infty \frac{\partial \psi_\sigma}{\partial \sigma} d\tilde{F}(y) &= \frac{1}{\sigma^2} - \frac{1}{1 - F(H, \theta)} \int_H^\infty \frac{3(\ln y - \mu)^2}{\sigma^4} f(y, \theta) dy \\
&\quad + \frac{1}{[1 - F(H, \theta)]^2} \left[ \int_0^H \left( \frac{(\ln y - \mu)^2}{\sigma^3} - \frac{1}{\sigma} \right) f(y, \theta) dy \right]^2 \\
&\quad + \frac{1}{1 - F(H, \theta)} \int_0^H \left( \frac{2}{\sigma^2} - \frac{5(\ln y - \mu)^2}{\sigma^4} + \frac{(\ln y - \mu)^4}{\sigma^6} \right) f(y, \theta) dy
\end{aligned} \tag{A.9}$$

$$\begin{aligned}
-\int_H^\infty \frac{\partial \psi_\mu}{\partial \sigma} d\tilde{F}(y) &= \frac{1}{1 - F(H, \theta)} \int_H^\infty \frac{-2(\ln y - \mu)}{\sigma^3} f(y, \theta) dy \\
&\quad + \frac{1}{[1 - F(H, \theta)]^2} \left[ \int_0^H \frac{\ln y - \mu}{\sigma^2} f(y, \theta) dy \right] \times \\
&\quad \quad \quad \left[ \int_0^H \left( \frac{(\ln y - \mu)^2}{\sigma^3} - \frac{1}{\sigma} \right) f(y, \theta) dy \right] \\
&\quad + \frac{1}{1 - F(H, \theta)} \int_0^H \left( \frac{(\ln y - \mu)^3}{\sigma^5} - \frac{3(\ln y - \mu)}{\sigma^3} \right) f(y, \theta) dy \\
&= -\int_H^\infty \frac{\partial \psi_\sigma}{\partial \mu} d\tilde{F}(y)
\end{aligned} \tag{A.10}$$

Afin de calculer complètement les éléments de  $\mathcal{A}(\theta)$ , il nous faut les expressions exactes des différentes intégrales intervenant dans les équations (A.8), (A.9) et (A.10).

Prenons l'intégrale suivante

$$\int_0^H \left( \frac{\ln y - \mu}{\sigma} \right)^2 f(y, \theta) dy = \int_0^H \left( \frac{\ln y - \mu}{\sigma} \right)^2 \frac{1}{\sqrt{2\pi}\sigma y} \exp \left\{ -\frac{1}{2} \left( \frac{\ln y - \mu}{\sigma} \right)^2 \right\} dy$$

on pose le changement de variables suivant  $u = (\ln y - \mu)/\sigma$

$$\int_0^H \left( \frac{\ln y - \mu}{\sigma} \right)^2 f(y, \theta) dy = \int_{-\infty}^{(\ln H - \mu)/\sigma} \frac{1}{\sqrt{2\pi}} u^2 \exp \left\{ -\frac{u^2}{2} \right\} du$$

puis par IPP en posant  $v' = u \exp \{-u^2/2\}$  et  $w = u$

$$\int_0^H \left( \frac{\ln y - \mu}{\sigma} \right)^2 f(y, \theta) dy = \frac{1}{\sqrt{2\pi}} \left[ -u \exp \left\{ -\frac{u^2}{2} \right\} \right]_{-\infty}^{(\ln H - \mu)/\sigma} + \int_{-\infty}^{(\ln H - \mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{u^2}{2} \right\} du$$

on reconnait une fonction de répartition d'une loi normale centrée réduite

$$\int_0^H \left( \frac{\ln y - \mu}{\sigma} \right)^2 f(y, \theta) dy = -\frac{1}{\sqrt{2\pi}} \left( \frac{\ln H - \mu}{\sigma} \right) \exp \left\{ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right\} + \Phi \left( \frac{\ln H - \mu}{\sigma} \right)$$

et enfin :

$$\int_0^H \left( \frac{\ln y - \mu}{\sigma} \right)^2 f(y, \theta) dy = -\frac{1}{\sqrt{2\pi}} \left( \frac{\ln H - \mu}{\sigma} \right) \exp \left\{ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right\} + F(H, \theta)$$

De manière identique, en utilisant le même changement de variables et une IPP, on peut prouver les résultats suivant

$$\begin{aligned} \int_0^H \left( \frac{\ln y - \mu}{\sigma} \right) f(y, \theta) dy &= -\frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \\ \int_0^H \left( \frac{\ln y - \mu}{\sigma} \right)^2 f(y, \theta) dy &= F(H, \theta) - \frac{\ln H - \mu}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \\ \int_0^H \left( \frac{\ln y - \mu}{\sigma} \right)^3 f(y, \theta) dy &= -\frac{1}{\sqrt{2\pi}} \left( 2 + \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right) \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \\ \int_0^H \left( \frac{\ln y - \mu}{\sigma} \right)^4 f(y, \theta) dy &= 3F(H, \theta) - \frac{1}{\sqrt{2\pi}} \left( \left( \frac{\ln H - \mu}{\sigma} \right)^3 + 3 \frac{\ln H - \mu}{\sigma} \right) \times \\ &\quad \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \\ \int_H^\infty \left( \frac{\ln y - \mu}{\sigma} \right) f(y, \theta) dy &= \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \\ \int_H^\infty \left( \frac{\ln y - \mu}{\sigma} \right)^2 f(y, \theta) dy &= \frac{\ln H - \mu}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] + [1 - F(H, \theta)] \end{aligned} \tag{A.11}$$

En réinjectant ces formules dans les expressions (A.8), (A.9) et (A.10) ainsi que dans celle de la fonction de score (A.7) on obtient les équations suivantes :

$$\psi_\theta = \begin{pmatrix} -\frac{\ln x - \mu}{\sigma^2} + \frac{1}{1 - F(H, \theta)} \left( \frac{1}{\sqrt{2\pi}\sigma} \right) \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \\ - \left( \frac{(\ln x - \mu)^2}{\sigma^3} - \frac{1}{\sigma} \right) + \frac{1}{1 - F(H, \theta)} \left( \frac{\ln H - \mu}{\sqrt{2\pi}\sigma^2} \right) \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \end{pmatrix} \quad (\text{A.12})$$

$$\begin{aligned} - \int_H^\infty \frac{\partial \psi_\mu}{\partial \mu} d\tilde{F}(y) &= -\frac{1}{\sigma^2} + \frac{1}{[1 - F(H, \theta)]^2} \left( \frac{1}{2\pi\sigma^2} \right) \exp \left[ -\left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \\ &\quad - \frac{1}{1 - F(H, \theta)} \left( \frac{\ln H - \mu}{\sqrt{2\pi}\sigma^3} \right) \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \end{aligned} \quad (\text{A.13})$$

$$\begin{aligned} - \int_H^\infty \frac{\partial \psi_\sigma}{\partial \sigma} d\tilde{F}(y) &= -\frac{2}{\sigma^2} - \frac{1}{1 - F(H, \theta)} \left( \frac{\ln H - \mu}{\sqrt{2\pi}\sigma^3} \right) \left[ 1 + \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \\ &\quad \times \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \\ &\quad + \frac{1}{[1 - F(H, \theta)]^2} \frac{(\ln H - \mu)^2}{2\pi\sigma^4} \exp \left[ -\left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \end{aligned} \quad (\text{A.14})$$

$$\begin{aligned} - \int_H^\infty \frac{\partial \psi_\mu}{\partial \sigma} d\tilde{F}(y) &= -\frac{1}{1 - F(H, \theta)} \left( \frac{1}{\sqrt{2\pi}\sigma^2} \right) \left( 1 + \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right) \\ &\quad \times \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \\ &\quad + \frac{1}{[1 - F(H, \theta)]^2} \left( \frac{\ln H - \mu}{2\pi\sigma^3} \right) \exp \left[ -\left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] \end{aligned} \quad (\text{A.15})$$

## Matrice d'initialisation $J(\theta)$ (OBRE)

L'information de Fisher  $J(\theta)$  servant à initialiser la valeur de  $A(\theta)$  dans l'algorithme de calcul de l'OBRE est définie par

$$J(\theta) = \int s(x, \theta) s(x, \theta)' dF_\theta(x)$$

où  $s(x, \theta)$  est la fonction de score pour la loi log-normale tronquée, qui a déjà été calculée, donnée formule (A.12). Ainsi en écrivant,  $J(\theta)$

$$J(\theta) = \begin{pmatrix} J_{1,1}(\theta) & J_{1,2}(\theta) \\ J_{2,1}(\theta) & J_{2,2}(\theta) \end{pmatrix}$$

on peut obtenir  $J_{i,j}(\theta)$ ,  $i, j \in \{1, 2\}^2$  par

$$\int_H^\infty \psi_i(x, \theta) \psi_j(x, \theta) dF_\theta(x)$$

Après des calculs fastidieux faisant encore intervenir changements de variables et IPP, on peut obtenir

$$\begin{aligned}
J_{1,1}(\theta) &= \left( \frac{\ln x}{\sigma^2} \right)^2 + \frac{1}{1 - F(H, \theta)} \left( \frac{1}{\sqrt{2\pi}\sigma^5} \right) [-2 \ln x Q_1(\theta) + Q_2(\theta)] \\
J_{2,2}(\theta) &= \left( \frac{(\ln x - \mu)^2}{\sigma^3} \right)^2 + \frac{1}{1 - F(H, \theta)} \left( \frac{1}{\sqrt{2\pi}\sigma^2} \right) \left[ -2 \left( \frac{\ln x - \mu}{\sigma} \right)^2 R_1(\theta) + R_2(\theta) \right] \\
J_{1,2}(\theta) &= \frac{\ln x}{\sigma^2} \times \frac{(\ln x - \mu)^2}{\sigma^3} + \frac{1}{1 - F(H, \theta)} \left( \frac{1}{\sqrt{2\pi}\sigma^2} \right) \left[ - \left( \frac{\ln x - \mu}{\sigma^2} \right)^2 Q_1(\theta) - \ln x R_1(\theta) + S(\theta) \right] \\
&= J_{2,1}(\theta)
\end{aligned}$$

avec

$$\begin{aligned}
Q_1(\theta) &= \sigma^2 \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] + \sqrt{2\pi}\sigma\mu(1 - F(H, \theta)) \\
Q_2(\theta) &= \sigma^2 (\ln H + \mu) \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] + \sqrt{2\pi}\sigma(\sigma^2 + \mu^2)(1 - F(H, \theta)) \\
R_1(\theta) &= \frac{\ln H - \mu}{\sigma} \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] + \sqrt{2\pi}(1 - F(H, \theta)) \\
R_2(\theta) &= \left( \left( \frac{\ln H - \mu}{\sigma} \right)^3 + 3 \left( \frac{\ln H - \mu}{\sigma} \right) \right) \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] + 3\sqrt{2\pi}(1 - F(H, \theta)) \\
S(\theta) &= \left( \left( \frac{\ln H - \mu}{\sigma} \right)^2 + \frac{\mu}{\sigma} \left( \frac{\ln H - \mu}{\sigma} \right) + 2 \right) \exp \left[ -\frac{1}{2} \left( \frac{\ln H - \mu}{\sigma} \right)^2 \right] + \frac{\mu}{\sigma} \sqrt{2\pi}(1 - F(H, \theta))
\end{aligned}$$

## B Problématiques numériques liées à l'OBRE

---

Obtenir la décomposition de Cholesky d'une matrice se fait très facilement sous MATLAB. C'est donc l'approximation numérique des intégrales qui doit faire l'objet d'une attention particulière. Dans la littérature, on peut trouver des approches permettant de contourner en partie ce problème. La plus courante consiste à discrétiser ces intégrales sur l'ensemble des données, en remplaçant la fonction de répartition par sa version empirique. Ainsi, pour une fonction  $\phi$  sur un domaine  $\mathcal{X}$ , on obtient la transformation sur les données  $(x_1, \dots, x_n)$  suivante

$$\int_{\mathcal{X}} \phi(x) dF(x, \theta) \rightarrow \int_{\mathcal{X}} \phi(x) d\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \phi(x_i)$$

Cette écriture est applicable aux calculs des matrices  $M_k$  (2). En revanche, il faut veiller à garder l'intégration pour le calcul de  $a(\theta)$  (cf article [4]). Cela permet donc de réduire de deux le nombre d'intégrales que l'on doit approximer numériquement. On notera également que cette approche est applicable aux termes de la matrice d'information de Fisher initialisant la valeur initiale de  $A(\theta)$ .

La seconde démarche consiste à approximer ces intégrales<sup>1</sup>. Afin d'obtenir des bornes finies, on effectue le changement de variables  $u(x) = \frac{1}{x}$  (qui est bien  $\mathcal{C}^1$  bijectif sur  $[H, \infty[$ ) pour aboutir à une intégrale de la forme

$$\int_H^\infty \phi(x) \frac{1}{\sqrt{2\pi\sigma x}} \exp\left[-\frac{1}{2} \left(\frac{\ln x - \mu}{\sigma}\right)^2\right] dx = \int_0^{1/H} \phi\left(\frac{1}{t}\right) \frac{1}{\sqrt{2\pi\sigma t}} \exp\left[-\frac{1}{2} \left(\frac{\ln t + \mu}{\sigma}\right)^2\right] dt$$

dans le cas de la loi log-normale.

On peut donc définir deux versions différentes d'implémentation de l'algorithme de l'estimateur OBRE. La première faisant intervenir la fonction de répartition empirique associée aux données, est appelée *méthode empirique*, la seconde utilisant des méthodes d'approximations numériques d'intégrales, est appelée *méthode reparamétrée*.

Ces deux programmes prennent les mêmes arguments, à savoir :

- *norme* : qui correspond à la norme utilisée dans la fonction de pondération  $W_c$ . Les modalités testées sont la norme euclidienne, la norme infinie et la norme 1.
- $\eta$  : le seuil de précision dont dépend la convergence.
- $H$  : le seuil de collecte.
- $c$  : le paramètre de modulation robustesse/efficacité.
- *donnees* : l'échantillon de données collectées.
- *limite* : le nombre maximal d'itérations autorisées dans l'algorithme.
- $\gamma$  : le paramètre permettant de faire varier le pas  $\Delta\theta$  dans l'algorithme.

Il peut arriver que, pour certains échantillons, l'algorithme ne converge pas. Il est donc important de définir un nombre limite de boucles dans l'algorithme. Dans la suite, on prend un nombre maximum de *limite* = 1000.

Avant de s'atteler aux propriétés de robustesse ainsi qu'à l'étude de la sensibilité des paramètres d'entrée, on doit s'assurer que l'estimateur vérifie les propriétés de convergence énoncés en début de partie dans le paragraphe 5.5. Il est intéressant de comparer ces propriétés pour les deux méthodes proposées afin de sélectionner la meilleure.

---

1. Plusieurs routines sont disponibles sous Matlab, chacune d'entre elles utilisant des méthodes mathématiques différentes. On peut citer à ce titre, la méthode des trapèzes, la quadrature de Lobatto, la quadrature de Simpson ou encore celle de Gauss-Kronrod. Quelle que soit la méthode choisie, l'appel de la fonction Matlab nécessite de prendre en arguments des bornes finies  $a$  et  $b$ . Dans notre cas, celles-ci valent respectivement  $H$  et  $+\infty$ .



## Choix de l'algorithme sur critère de convergence

Les figures B.1 et B.2 présentent les biais relatifs moyens et les écarts-type relatifs moyens de l'estimateur pour les méthodes empiriques et reparamétrées respectivement en fonction de la taille d'échantillon. Les paramètres de l'OBRE sont fixés à  $\eta = 0.001$ ,  $c = 5$  (pour ne privilégier ni robustesse, ni efficacité),  $limite = 1000$  et la norme choisie est euclidienne (elle apparaissant comme la plus intuitive). Nous prenons également  $\gamma = 1$ . Les paramètres initiaux sont ici obtenus grâce à une estimation par maximum de vraisemblance des données sans prendre en compte le seuil. Les courbes sont obtenues pour les quatre profils de risque mentionnés ci-dessus, les données étant tronquées au seuil  $H = 1000$  €. Les graphiques sont obtenus pour différentes tailles d'échantillon, celles-ci balayant l'intervalle  $[20, 50000]$  en se focalisant principalement sur les échantillons de petites tailles, et les biais et écarts-type relatifs moyens sont obtenus à partir de 500 simulations.

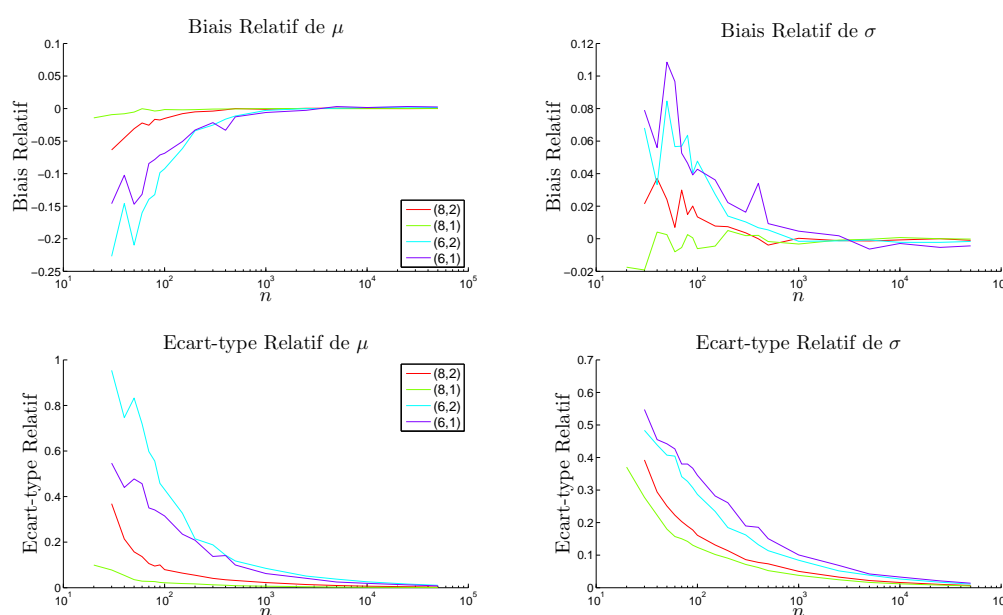


Figure B.1 – Convergence de l'estimateur OBRE par méthode empirique pour différents profils de risque

La première chose que l'on peut conclure à l'étude de ces graphiques est que **l'estimateur OBRE est bien convergent pour les deux méthodes** puisque, quel que soit le profil de risque, le biais et l'écart-type relatifs moyens convergent vers 0 lorsque la taille de l'échantillon augmente. Dans l'ensemble, le comportement de cette convergence est similaire entre les deux modèles, les modèles les plus sévères semblent également converger plus rapidement. Du reste, la convergence apparaît comme légèrement plus rapide pour le modèle reparamétré. Dans l'ensemble, **un seuil satisfaisant de convergence en termes de biais et de dispersion est atteint pour une taille d'échantillon de 500 données, valeur que nous utilisons par la suite pour calibrer les paramètres du modèle.**

La figure B.3 présente le taux de non convergence ainsi que le taux d'erreur pour ces deux méthodes. On considère que l'algorithme est non convergent lorsque le nombre d'itérations maximal, est atteint sans que la condition d'arrêt ne soit respectée. Par ailleurs, une erreur est rencontrée dans l'algorithme lorsque l'une des matrices  $M_k$  devient non inversible ou à valeurs complexes, ou bien encore lorsque la matrice  $M_2^{-1}$  n'est pas définie positive, et donc qu'il n'est pas possible d'utiliser la factorisation de Cholesky.

On peut observer deux comportements bien distincts selon la méthode employée. D'un côté, l'approche empirique fournit un faible taux d'erreur (inférieur à 0.5% pour la plupart des profils) mais un taux élevé de non convergence, celui-ci variant entre 10 et 25% selon les profils de risque pour les petits échantillons. De l'autre, la méthode reparamétrée donne un taux de non convergence quasi nul mais un très fort pourcentage d'erreurs rencontrées dans la procédure. Aucun de ces deux estimateurs ne semble donc se dégager puisqu'un fort taux de non convergence ou d'erreur lors de l'estimation d'échantillons de faible taille n'est pas recevable, d'autant plus que cette étude se

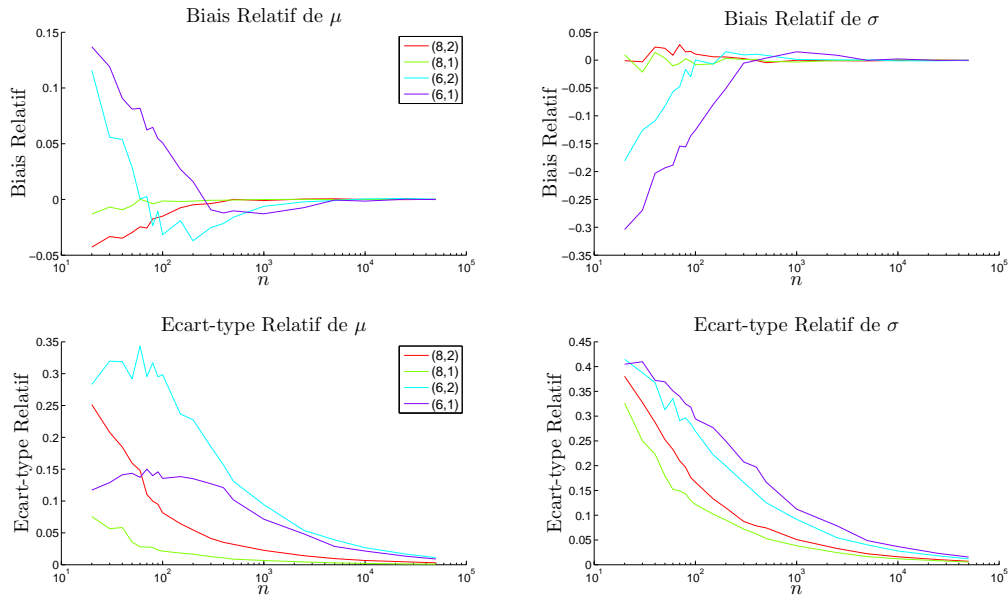


Figure B.2 – Convergence de l'estimateur OBRE par méthode reparamétrée pour différents profils de risque

fait sur données simulées. En effet, pour les échantillons de petite taille, les conditions optimales utilisées pour la simulation ne sont plus respectées.

Cependant, il semble surprenant que l'algorithme ne converge pas pour 1000 itérations, alors que dans les cas de convergence, le nombre moyen d'itérations est très largement inférieur à la centaine. Pour comprendre pourquoi la condition d'arrêt n'est pas atteinte, nous avons tracé les valeurs de  $\mu$  et de  $\sigma$  à chaque itération. Un tel cas de non convergence est présenté figure B.4 à gauche. Le couple d'initialisation est de couleur rouge, les couples représentant les itérations impaires de couleur cyan et les paires de couleur bleue. On peut constater une oscillation entre les itérations paires et impaires sans jamais converger, la condition d'arrêt se basant sur la variation relative des paramètres n'étant jamais atteinte. On peut alors supposer que la "vraie" solution se trouve entre les nuages de points des itérations paires et impaires, le problème étant que la variation  $\Delta\theta$  dans l'équation (5.6) est trop importante pour entraîner la convergence.

Il semble naturel, qu'en de telles circonstances, la variation des paramètres appliqués dans cette équation soit revue à la baisse. C'est pourquoi nous modifions la valeur de  $\gamma$ .

De la sorte, le pas à chaque itération va pouvoir être affiné en augmentant la valeur de  $\gamma$ . Il est important de noter que dans les cas où l'algorithme converge déjà (c'est-à-dire pour  $\gamma = 1$ ), augmenter  $\gamma$  ne va pas altérer la valeur limite des paramètres finaux obtenus. Ainsi, dans le graphe de droite donné figure B.4, la trajectoire verte correspond aux couples de paramètres  $\mu$  et  $\sigma$  obtenus pour le même échantillon de données mais pour une valeur  $\gamma = 2$ . L'algorithme devient alors convergent, sa limite étant bien celle présupposée, et cela en seulement 10 itérations.

Le tableau B.5 présente le pourcentage de convergence de l'algorithme ainsi que le nombre d'itérations moyen suivant différentes valeurs de  $\gamma$  pour 1000 échantillons de taille 50 générés à partir d'une loi lognormale de paramètres  $\mathcal{LN}(8,1)$  pour lesquels l'algorithme par méthode empirique ne convergeait pas initialement, atteignant le nombre maximal de simulations fixé à 1000. On constate que la convergence a lieu à chaque fois quelque soit la valeur de  $\gamma$ , et que le nombre d'itérations moyen est minimal pour la valeur minimale 2, ce qui est logique. La méthode empirique apparaît donc comme fonctionnelle puisque le problème de non convergence se règle de façon simple.

Cette méthode peut également s'employer lorsqu'une erreur est rencontrée dans l'algorithme, typiquement lors de l'emploi de l'approche par reparamétrage. De façon similaire, on obtient le tableau B.6 reportant des résultats obtenus également pour 1000 échantillons de taille 50 générés à partir d'une loi lognormale de paramètres  $\mathcal{LN}(6,1)$  pour lesquels une erreur a été rencontrée dans l'algorithme. On observe cette fois-ci que l'augmentation de  $\gamma$  ne permet pas d'entraîner à coup sûr la convergence. Une valeur de  $\gamma$  faible maintient en effet un fort pourcentage

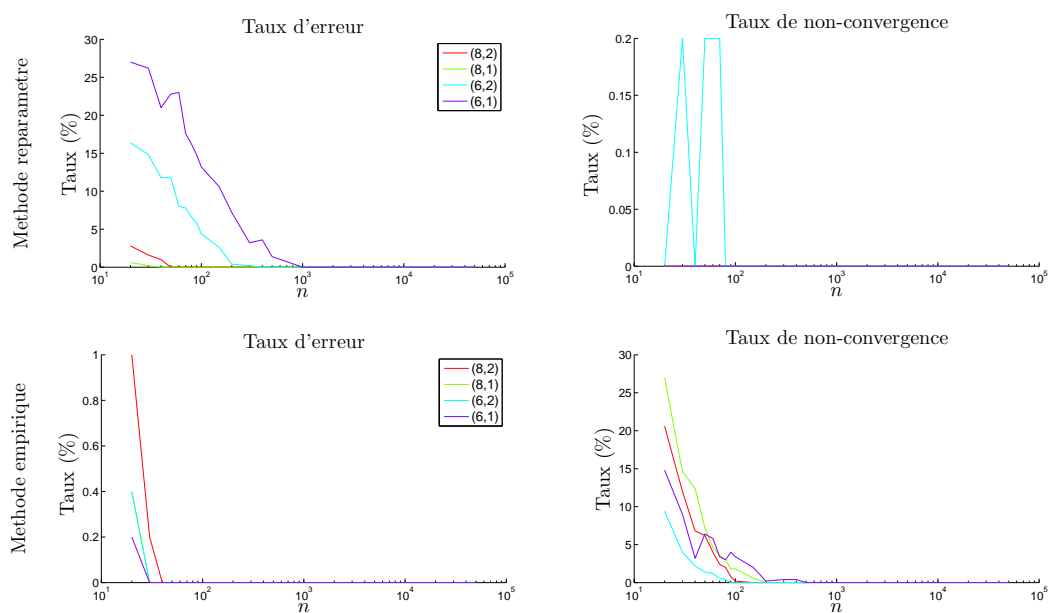


Figure B.3 – Taux d’erreur et de non convergence par méthode empirique (en bas) et reparamétrée (en haut) selon la taille de l’échantillon pour différents profils de risque

d’erreur, cependant l’augmenter entraîne également la non convergence de l’algorithme. Le pourcentage maximal de convergence est seulement de 54%, obtenu pour les valeurs 50 et 100.

On peut donc en conclure que cette version reparamétrée n’est pas opérante puisque pour de petits échantillons, une quantité non négligeable de procédures n’aboutit pas. Par exemple, en prenant un échantillon d’une loi  $\mathcal{LN}(6,2)$  et en prenant le meilleur cas étudié dans le tableau, au moins 10% des échantillons verra l’algorithme d’estimation échouer. Ce constat s’aggravera d’autant plus sur données réelles.

## Sensibilité des paramètres

Dans cette partie, nous testons la sensibilité de l’estimation aux paramètres  $\eta$  et aux paramètres initiaux afin de définir un calibrage optimal.

**Sensibilité au paramètre  $\eta$ .** Comme la condition d’arrêt dépend du paramètre  $\theta$ , il est intéressant de connaître le comportement du critère de convergence lorsque  $\eta$  varie. C’est pourquoi nous avons tracé figures B.7 et B.8 les informations relatives à la convergence de l’estimateur et celles relatives aux problèmes algorithmiques. Les résultats sont obtenus en faisant varier  $\eta$  de  $10^{-6}$  à 0.5 en 20 valeurs également réparties à l’échelle logarithmique, et en fixant les autres paramètres à  $n = 500$ ,  $c = 5$ ,  $limite = 1000$ ,  $H = 1000\text{€}$ , et  $divTheta = 1$ .

On constate que, d’une part, lorsque le seuil  $\eta$  est trop faible, le biais relatif des deux paramètres est trop élevé ( $>10\%$ ), d’autre part, quand le seuil est trop fin, la convergence ne s’améliore plus en ce que le biais relatif n’évolue plus. De plus, le temps de calcul est logarithmiquement décroissant avec la valeur de  $\eta$ . En se référant à ces graphiques, on peut calibrer notre seuil de convergence pour optimiser le rapport biais minimal/temps de calcul. **Nous choisissons donc  $\eta = 0,001$**  afin de garder en plus une marge de sécurité, la convergence étant assurée pour les quatre modèles pour  $\eta$  de l’ordre de 0,05.

**Sensibilité aux paramètres d’initialisation.** Dans l’algorithme, nous avons choisi d’initialiser les paramètres par la valeur obtenue par maximum de vraisemblance sur les données non tronquées. Cela signifie que, même

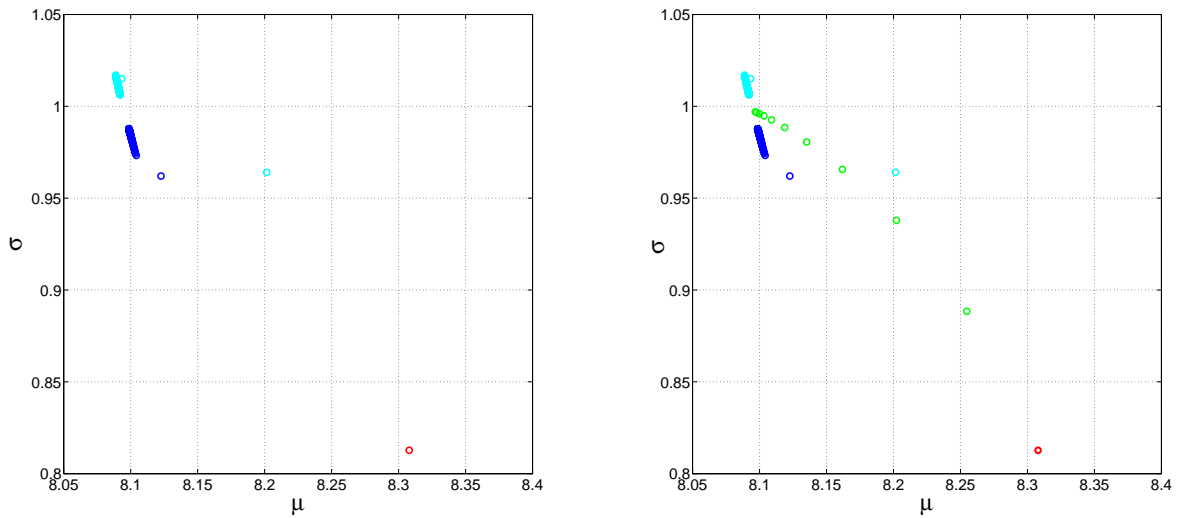


Figure B.4 – Valeurs des couples  $(\mu, \sigma)$  à chaque itération de l'algorithme pour un jeu de données, dans un cas de non convergence  $\gamma = 1$  (à gauche), dans un cas de convergence pour  $\gamma = 2$  (à droite)

$\gamma$	2	5	10	20
% convergence	99,9	100	99,9	99,9
Nombre d'itérations moyen	11,78	29	58,075	116,152

Figure B.5 – Taux de convergence et nombre d'itérations moyen selon  $\gamma$  pour la méthode empirique

si le seuil est de  $H = 1000$ , cette initialisation se fait pour un seuil  $H = 0$ . En effet, cela permet d'obtenir des paramètres d'ordre de grandeur convenable, et de passer outre les estimations parfois aberrantes que peut fournir le maximum de vraisemblance sur données tronquées (un  $\mu$  négatif et un  $\sigma$  très élevé).

L'objectif de ce paragraphe est de montrer que cette approche ne contrarie en rien la convergence de l'algorithme. Pour prouver cela, nous proposons d'étudier 3 initialisations :

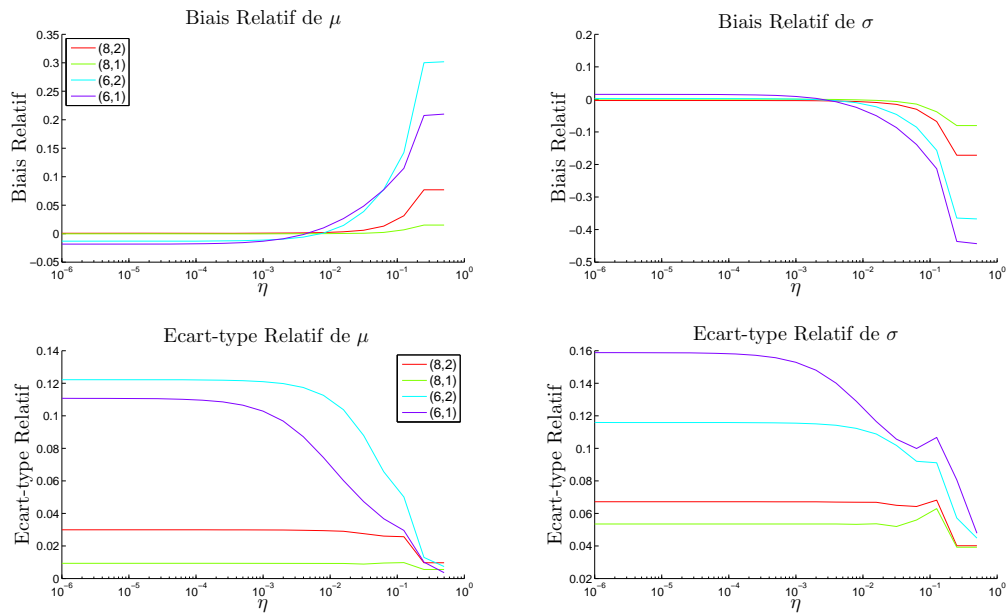
- Une initialisation par ML sur données non tronquées sans modification de ces données. C'est ce qui a été fait dans notre algorithme. On la note *Init1*.
- Une initialisation par ML sur données non tronquées en translatant les données de la valeur du seuil. Ainsi, une perte  $x$  générée au-dessus du seuil devient  $x - H$ . On la note *Init2*.
- Une initialisation par ML sur données tronquées. On la note *Init3*.

Pour ces trois initialisations et pour différentes tailles d'échantillon, nous calculons le biais moyen obtenu sur un modèle  $\mathcal{LN}(8, 2)$  pour 500 simulations, ainsi que le nombre d'erreurs algorithmiques rencontrées, le nombre moyen d'itérations et le nombre de fois où la limite a été atteinte. Ces résultats sont fournis par la table B.1 pour les paramètres  $\eta = 0.001$ ,  $c = 5$ ,  $limite = 1000$  et  $\gamma = 1$ .

On peut remarquer que l'initialisation n'importe que très peu. En effet, pour ces 3 initialisations, le biais relatif moyen est nul pour  $n = 100$  et  $n = 1000$ , le nombre d'erreurs algorithmiques reste nul, le nombre d'itérations reste bas et le nombre de fois où la limite est atteinte reste très faible (d'autant plus que ce problème est remédiable en augmentant la valeur de  $\gamma$ ). L'initialisation que nous proposons, *Init1*, apparaît légèrement meilleure que l'initialisation *Init2*. Même si le nombre moyen d'itérations est deux fois plus faible pour l'initialisation *Init3*, nous évitons de l'utiliser, puisque l'estimateur ML sur données tronquées peut être très sensible sur certains échantillons (typiquement une valeur de  $\mu$  négative et une valeur de  $\sigma$  très importante).

**En somme, rien n'invalide notre façon d'initialiser. Dans la suite, nous initialisons donc sur données non modifiées et non tronquées.**

$divTheta$	2	5	10	20	50	100	500
% erreurs	86	72	60	47	42	38	40
% non convergence	0	0	2	6	2	6	56

Figure B.6 – Taux d'erreur et taux de convergence selon  $\gamma$  pour la méthode reparamétréeFigure B.7 – Convergence de l'estimateur OBRE par méthode empirique selon  $\eta$  pour différents profils de risque

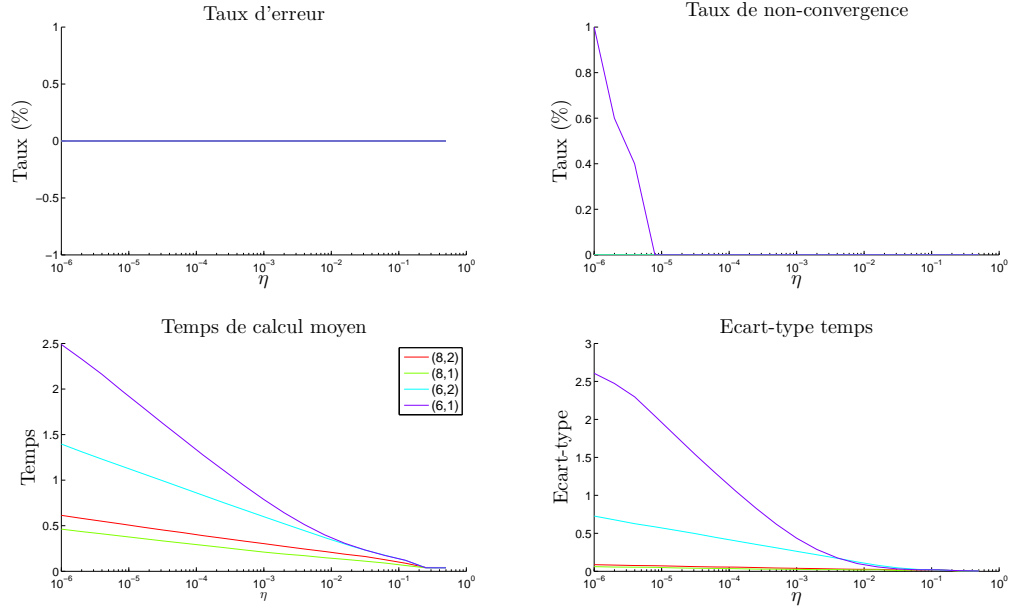


Figure B.8 – Taux d’erreurs et de non convergences, et temps de calcul nécessaire par méthode empirique selon  $\eta$  pour différents profils de risque

		$\mathbb{B}_{rel}(\mu)$	$\mathbb{B}_{rel}(\sigma)$	Nb erreurs	Nb boucles moyen	Nb limite atteinte
$n = 1000$	<i>Init1</i>	0%	-0.2%	0	4.9	0
	<i>Init2</i>	0%	-0.3%	0	7.5	0
	<i>Init3</i>	0%	-0.2%	0	2.7	0
$n = 100$	<i>Init1</i>	-0.7%	-0.1%	0	17.7	5
	<i>Init2</i>	-0.7%	-0.3%	0	19.3	5
	<i>Init3</i>	-0.7%	-0.7%	0	9.0	3

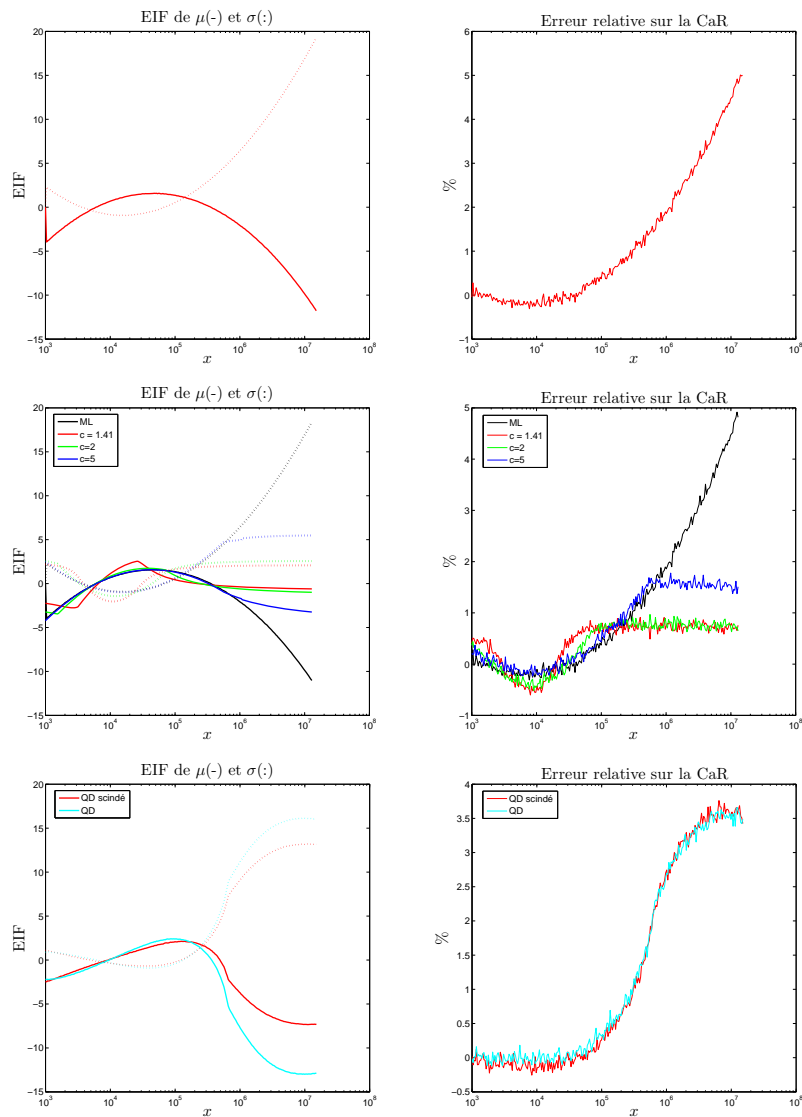
Table B.1 – Analyse de différentes initialisations de l’algorithme sur plusieurs critères pour deux tailles d’échantillon

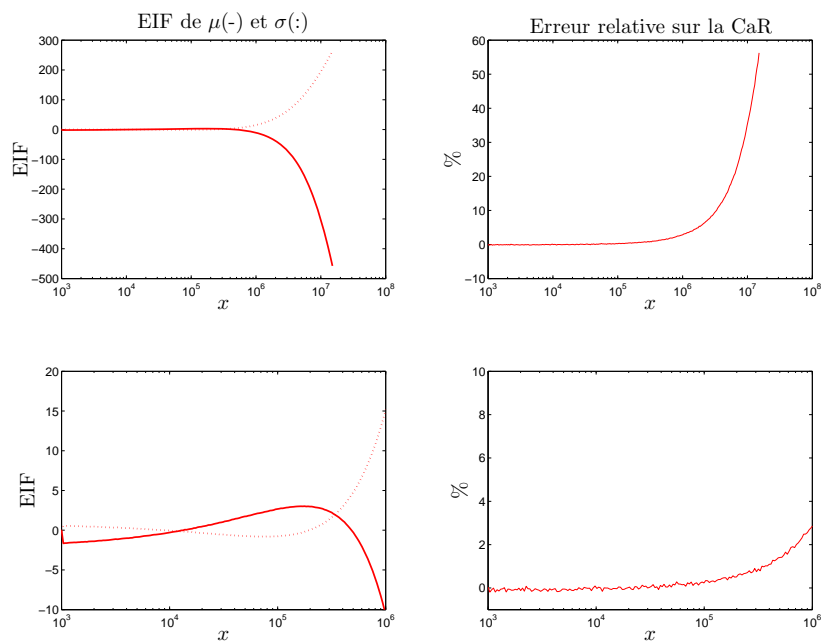
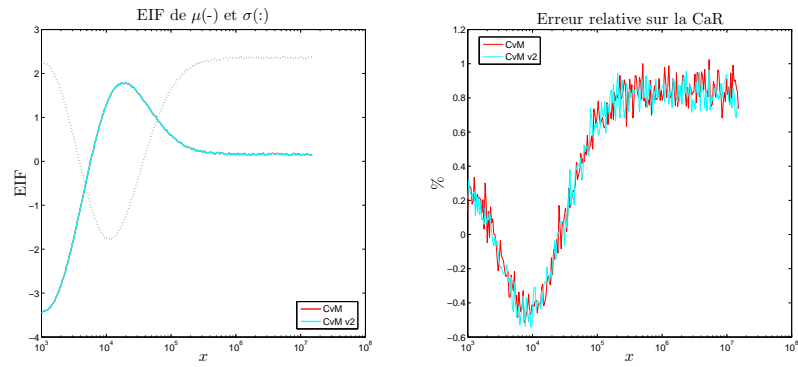
# C Résultats de robustesse complémentaires

## Fonction d'influence des estimateurs

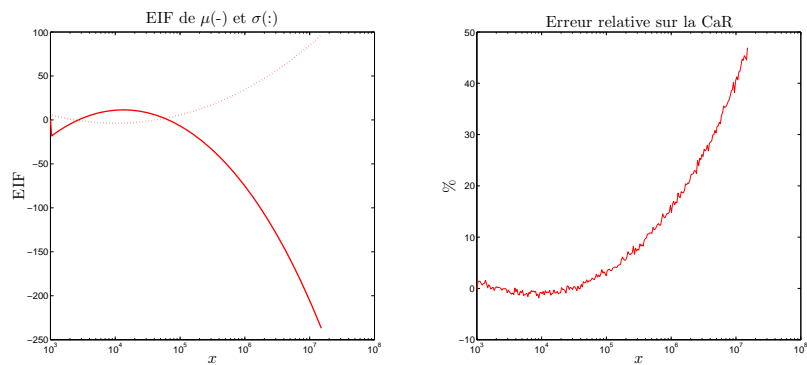
Les résultats présentés ici sont faits dans les mêmes conditions que le cas  $\mathcal{LN}(8, 2)$  étudié dans le corps du rapport. Les graphiques sont arrangés dans l'ordre des estimateurs suivant : ML, OBRE, QD, CvM et GMM.

### EIF d'un modèle $\mathcal{LN}(8, 1)$

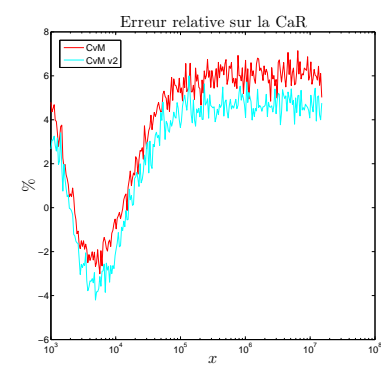
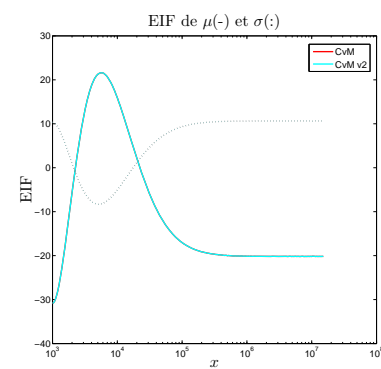
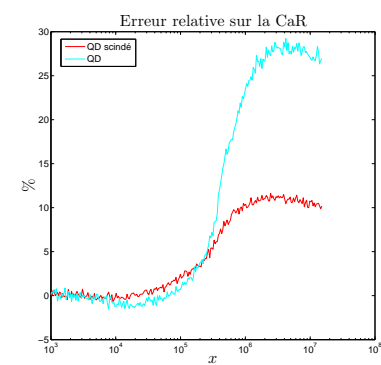
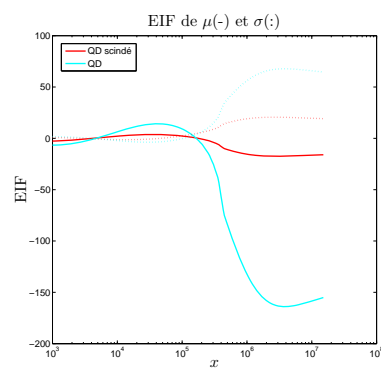
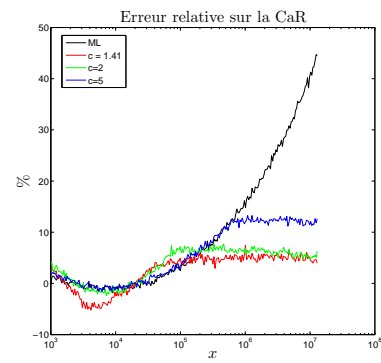
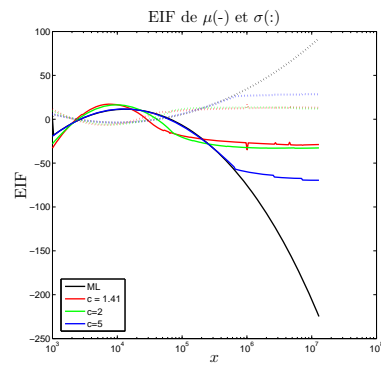


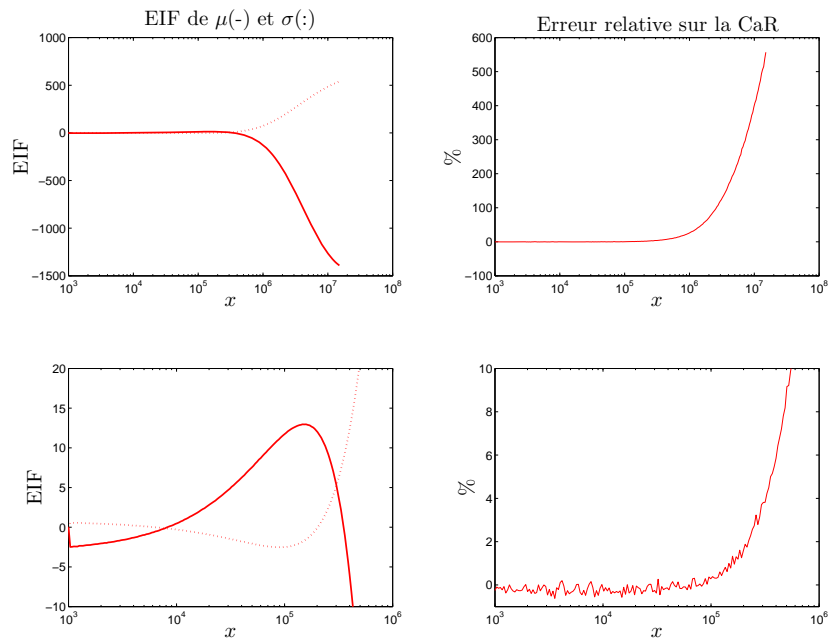


## EIF d'un modèle $\mathcal{LN}(6, 2)$

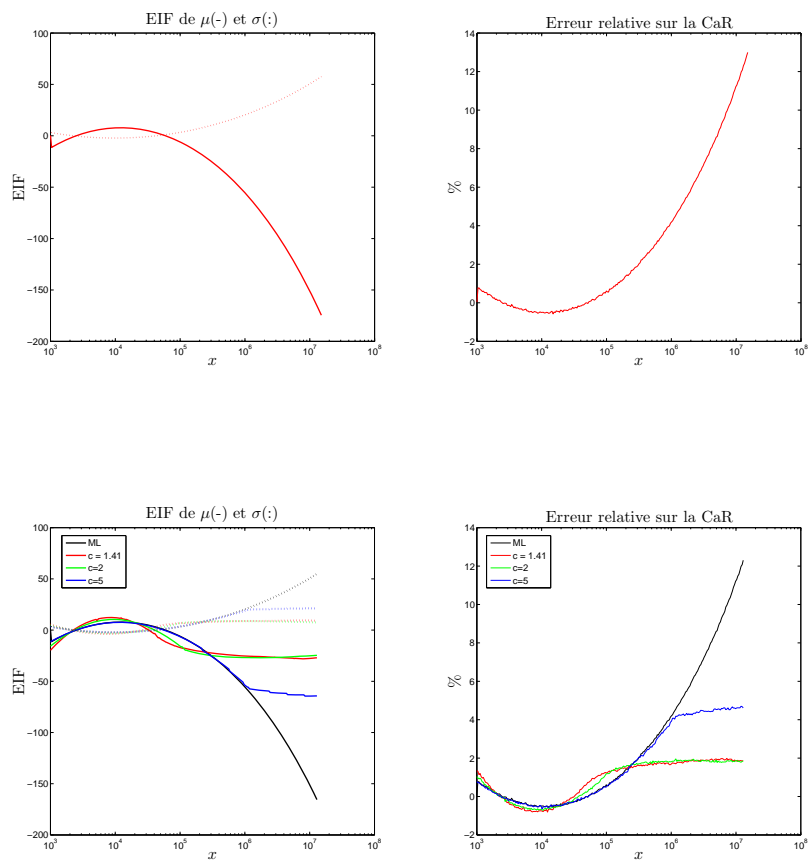


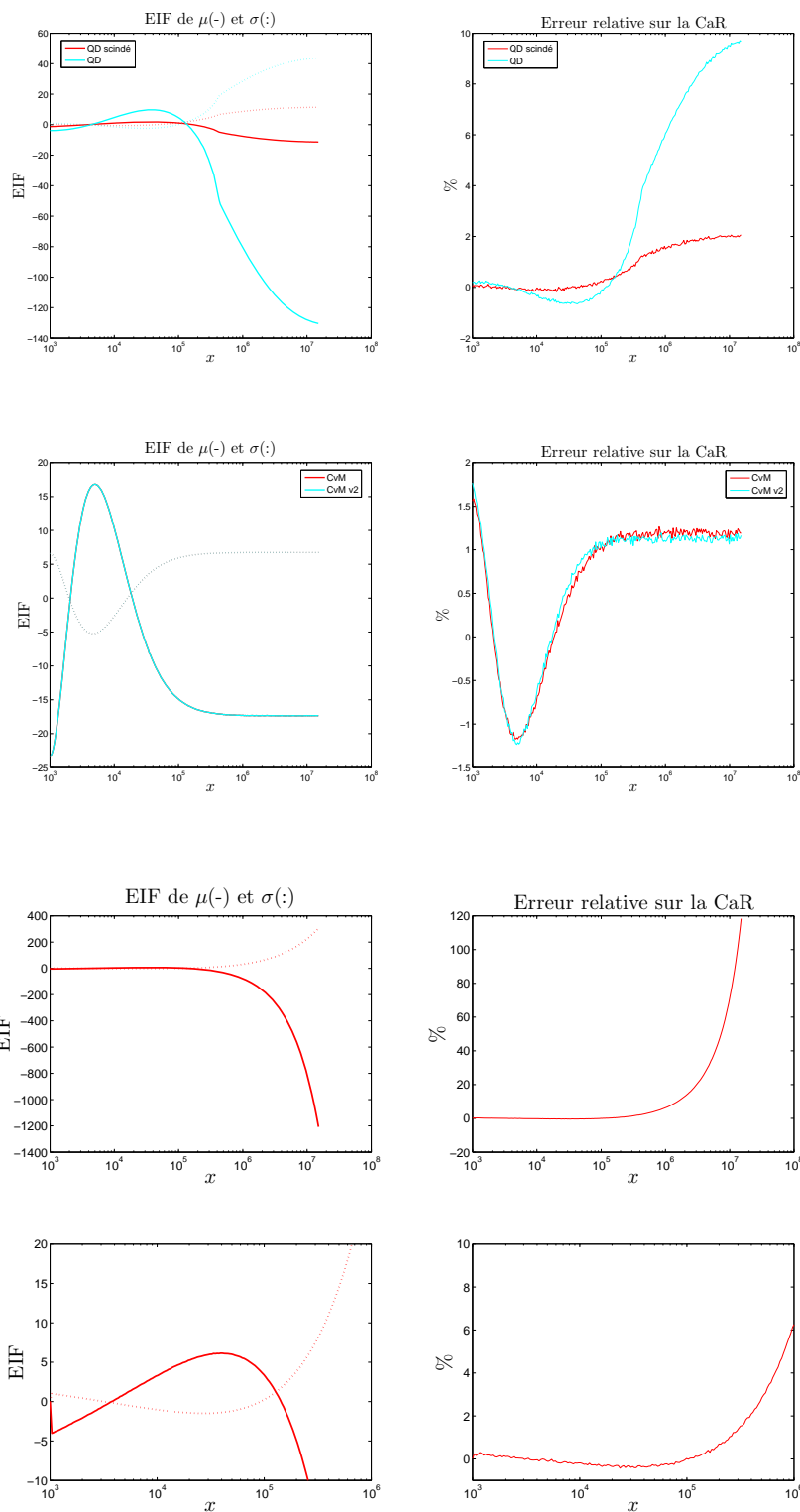






### EIF d'un modèle $\mathcal{LN}(6, 1)$

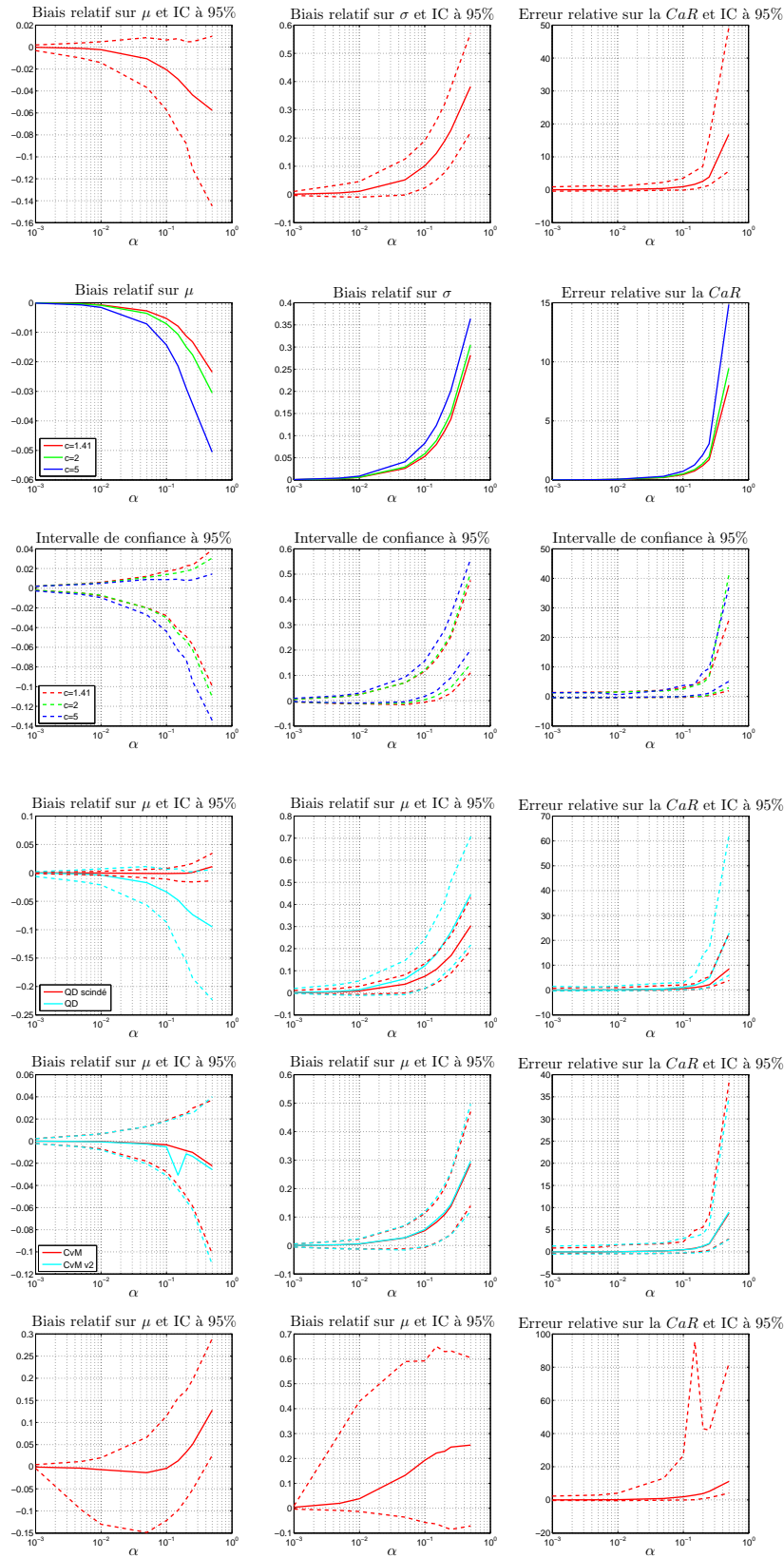




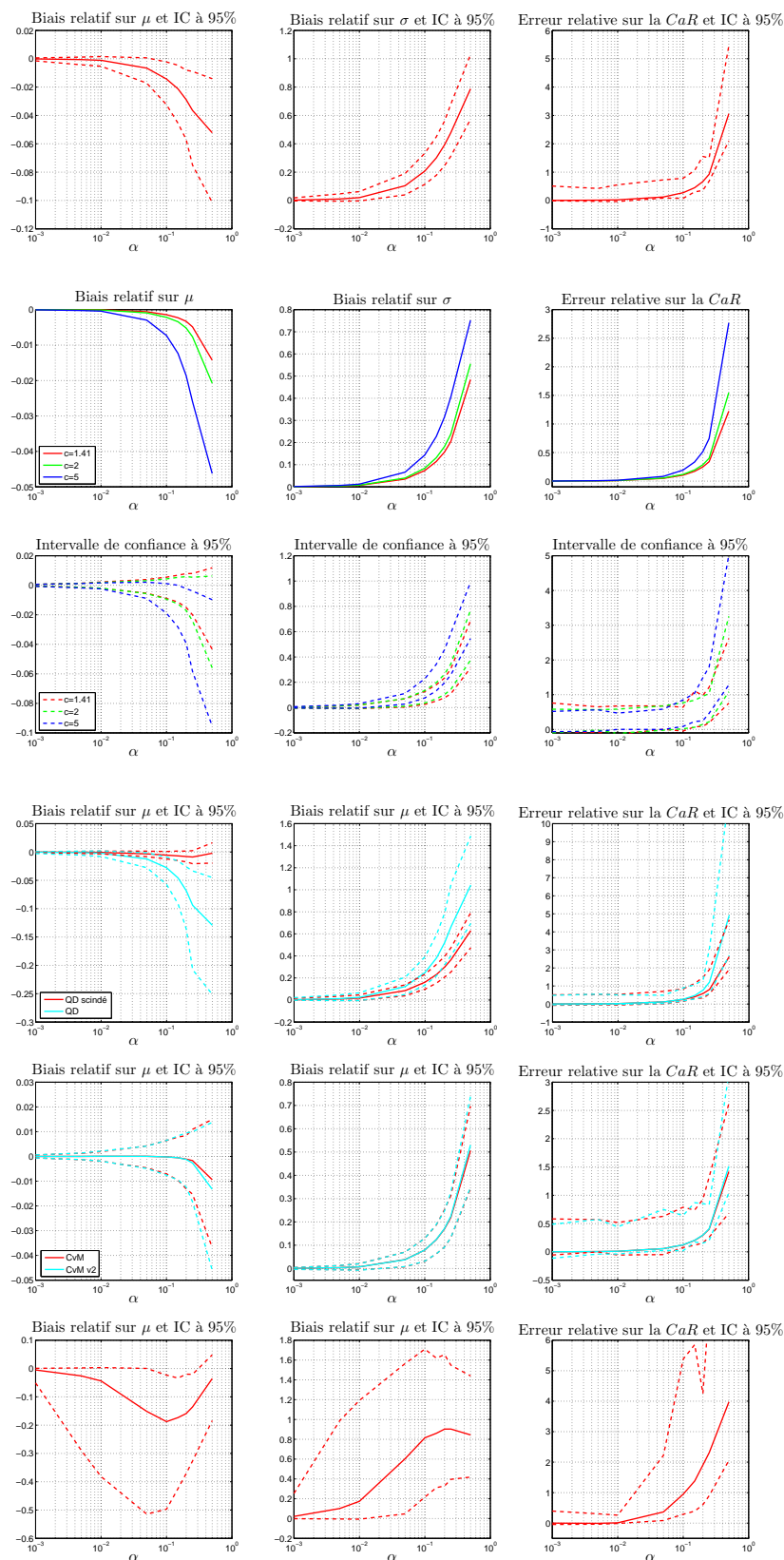
## Contamination par un modèle log-normal

Les résultats présentés ici sont faits dans les mêmes conditions que le cas  $\mathcal{LN}(8, 2)/\mathcal{LN}(8, 4)$  étudié dans le corps du rapport. Les graphiques sont arrangés dans l'ordre des estimateurs suivant : ML, OBRE, QD, CvM et GMM.

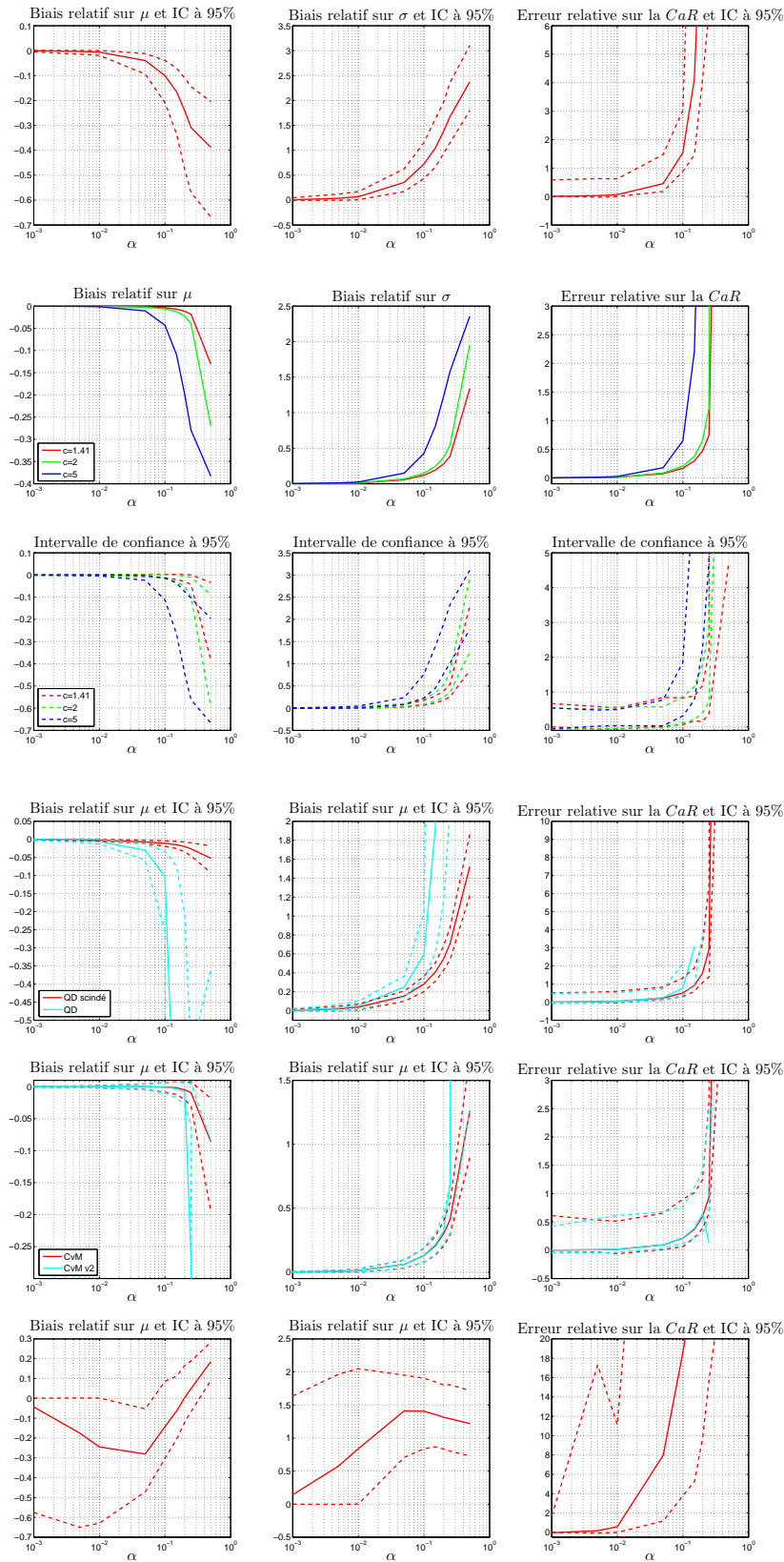
## Contamination d'un modèle $\mathcal{LN}(8, 2)$ par un modèle $\mathcal{LN}(8, 3)$



## Contamination d'un modèle $\mathcal{LN}(8, 1)$ par un modèle $\mathcal{LN}(8, 2)$



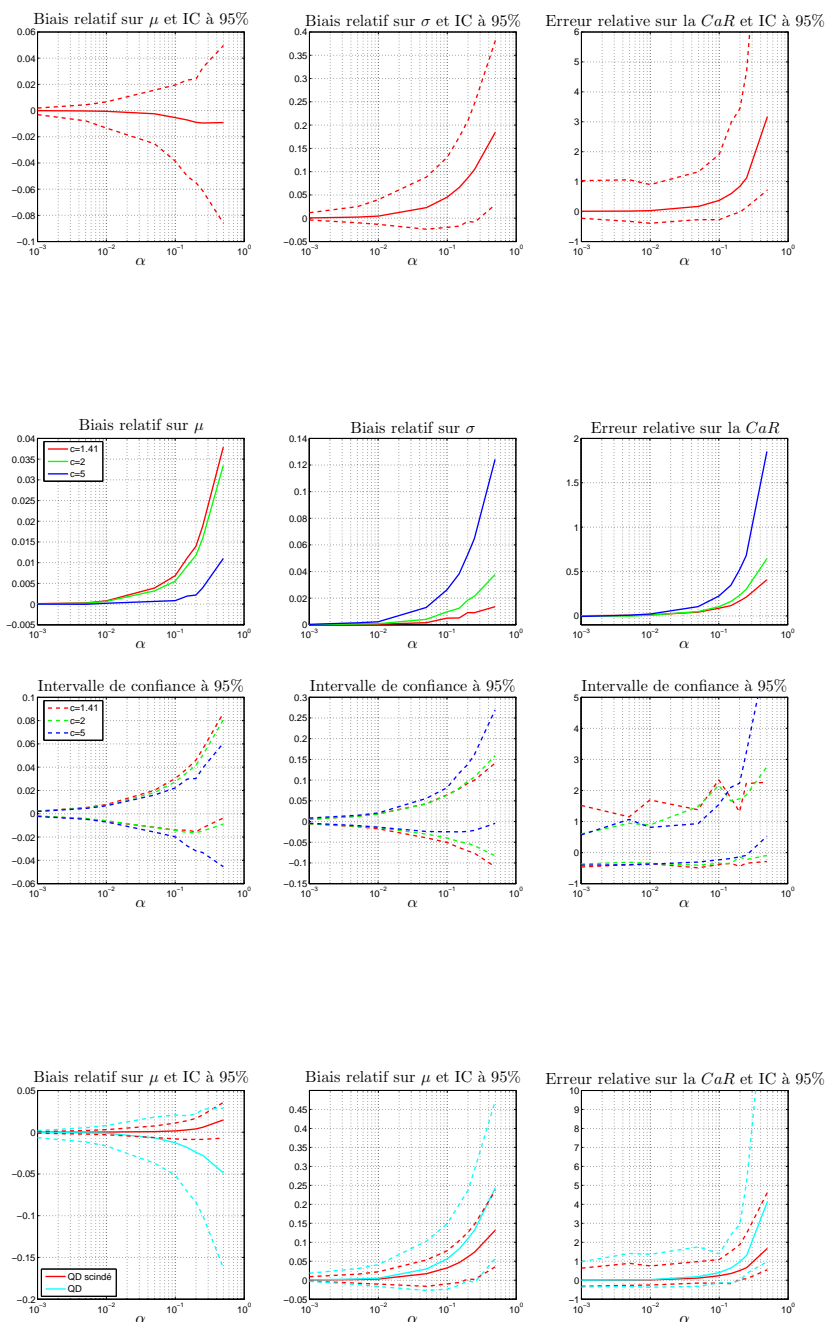
## Contamination d'un modèle $\mathcal{LN}(8, 1)$ par un modèle $\mathcal{LN}(8, 3)$

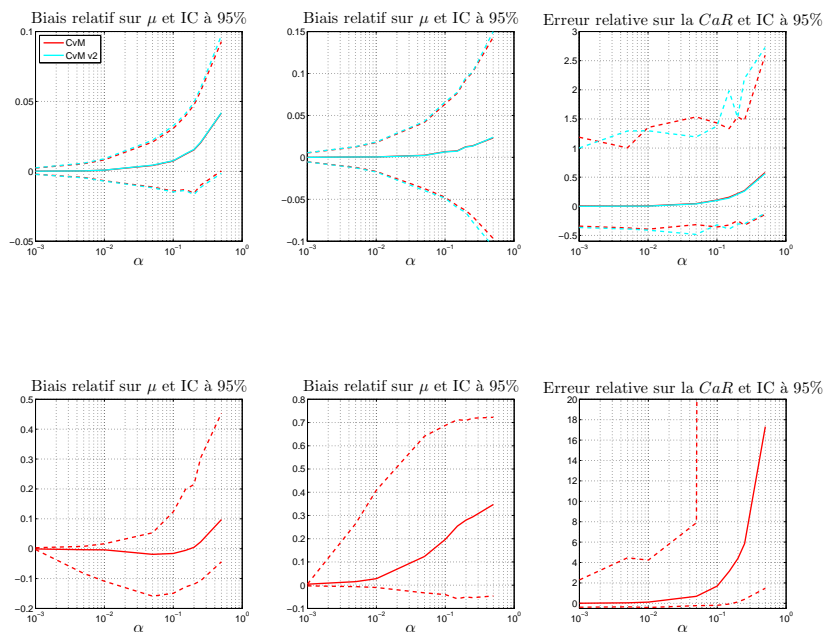


## Contamination par un modèle de Pareto

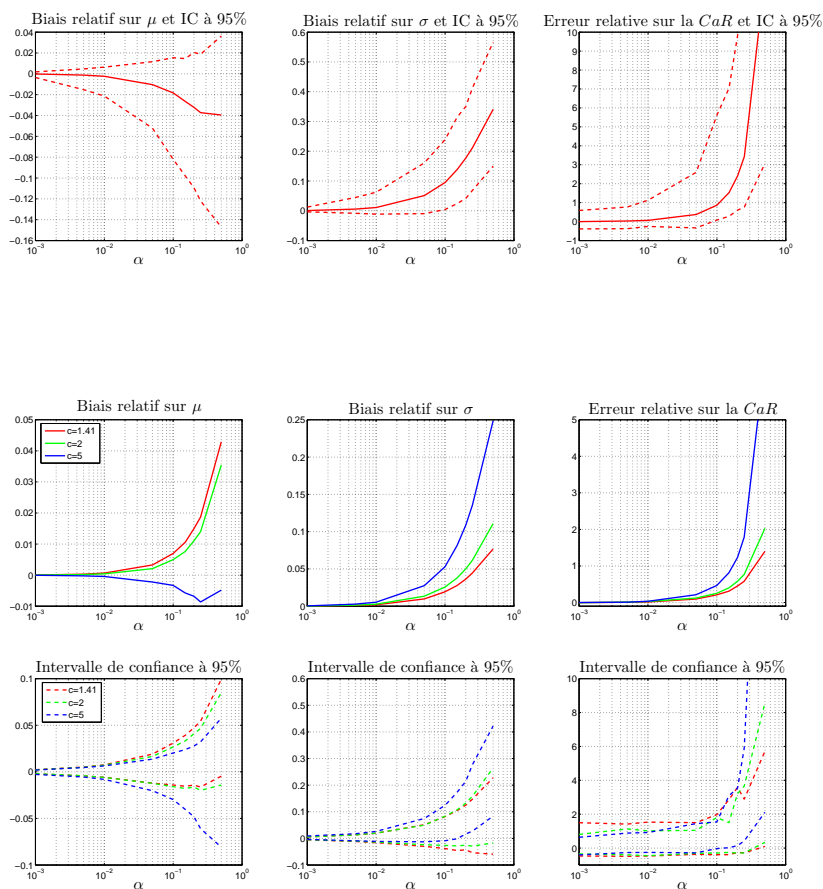
Les résultats présentés ici sont faits dans les mêmes conditions que le cas  $\mathcal{LN}(8, 2)$  étudié dans le corps du rapport. Les graphiques sont présentés dans l'ordre des estimateurs suivant : ML, OBRE, QD, CvM et GMM.

### Contamination d'un modèle $\mathcal{LN}(8, 2)$ par un modèle $\mathcal{P}(0.7, 4000)$

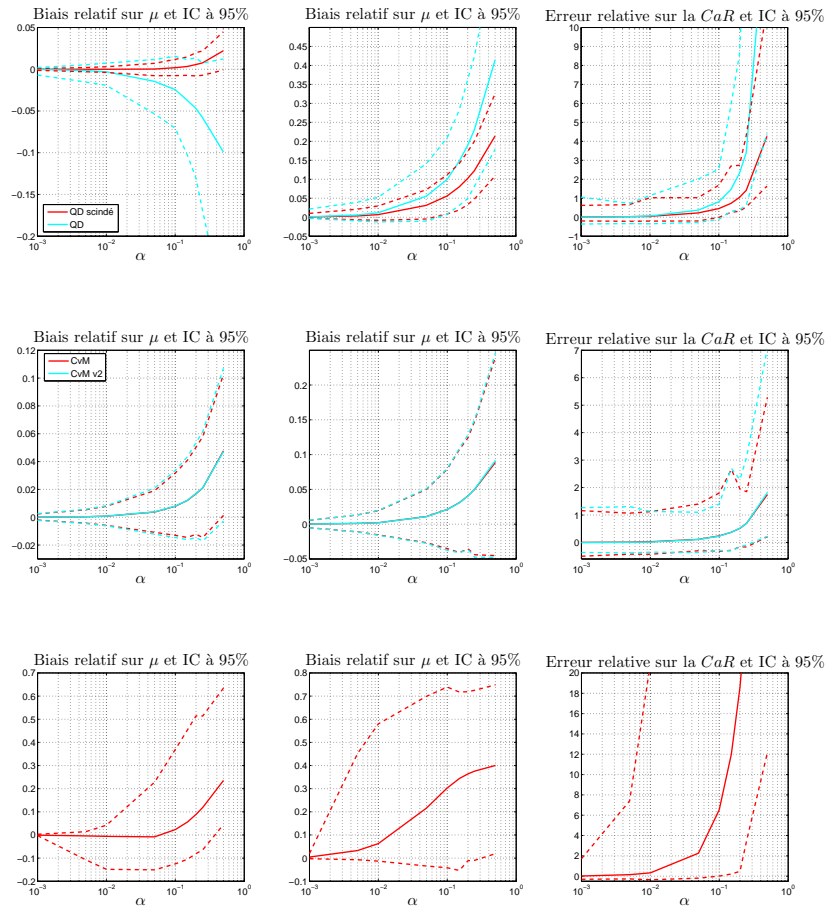




## Contamination d'un modèle $\mathcal{LN}(8, 2)$ par un modèle $\mathcal{P}(0.6, 4000)$







## D Résultats sur données réelles

	$\hat{\mu}$	$\hat{\sigma}$	$F_{\hat{\theta}}(H)$	KS	CVM	AD2	AD2up	
E1	4,49	3,24	77%	0,271	0,086	0,012	0,295	9,47E+07
E2	6,57	2,03	57%	0,013	0,009	0,003	0,774	1,12E+07
E3	5,86	1,68	0%	0	0	0	0,114	9,14E+05
E4	8,22	1,94	25%	0	0	0	0,021	2,09E+07
E5	-10,95	5,29	100%	0,001	0,004	0,015	0,204	-
E6	1,75	3,39	94%	0	0,001	0,001	0,056	6,37E+07
E7	-3,96	3,73	100%	0,048	0,072	0,097	0,105	9,18E+06
E8	-30,01	6,09	100%	-	-	-	-	-
E9	-77,39	10,11	100%	-	-	-	-	-
E10	7,55	1,00	26%	0,146	0,062	0,056	0,172	1,96E+05
E11	11,42	1,72	0%	0,155	0,205	0,207	0,49	7,97E+07
E12	-21,14	8,31	100%	-	-	-	-	-
E13	4,64	2,52	82%	0,412	0,117	0,009	0,074	2,29E+07
E14	2,64	3,06	92%	0	0	0	0,004	1,84E+08

Table D.1 – ML

	$\hat{\mu}$	$\hat{\sigma}$	$F_{\hat{\theta}}(H)$	KS	CVM	AD2	AD2up	
E1	9,14	1,43	6%	0	0	0	0,053	3,29E+06
E2	7,35	1,77	40%	0	0	0	0,148	7,80E+06
E3	-5,65	3,27	96%	0	0	0	0	5,76E+04
E4	8,60	1,66	15%	0	0	0	0,012	9,84E+06
E5	6,84	2,19	51%	0	0	0	0,013	3,13E+07
E6	8,05	1,57	23%	0	0	0	0,058	3,50E+06
E7	5,09	2,28	79%	0	0	0,001	0,026	4,10E+06
E8	2,43	2,86	94%	0	0	0	0	5,85E+07
E9	8,49	1,48	14%	0	0	0	0,02	1,52E+06
E10	7,85	0,80	12%	0	0	0	0,149	1,59E+05
E11	11,95	1,24	0%	0	0	0	0,136	2,56E+07
E12	10,96	1,80	1%	0	0	0	0,009	9,64E+07
E13	6,15	2,24	63%	0	0	0	0,021	2,10E+07
E14	5,76	2,65	67%	0	0	0	0,003	2,50E+08

Table D.2 – GMM

	$\hat{\mu}$	$\hat{\sigma}$	$F_{\hat{\theta}}(H)$	KS	CVM	AD2	AD2up	
E1	1,98	4,03	89%	0,355	0,186	0,026	0,111	3,84E+08
E2	6,55	2,05	57%	0,022	0,021	0,003	0,775	1,16E+07
E3	6,15	1,31	0%	0	0	0	0,01	4,05E+05
E4	7,47	3,12	43%	0	0	0	0,006	1,70E+09
E5	-8,97	5,06	100%	0,003	0,001	0,008	0,293	-
E6	-0,02	3,79	97%	0	0,002	0,004	0,048	1,17E+08
E7	-3,52	3,66	100%	0,06	0,094	0,119	0,124	8,84E+06
E8	-5,37	3,66	100%	0,001	0	0,001	0,003	-
E9	-27,84	8,55	100%	-	-	-	-	-
E10	7,18	1,41	42%	0	0,008	0,003	0,033	4,87E+05
E11	11,38	1,85	1%	0,331	0,321	0,26	0,343	1,24E+08
E12	-18,92	8,01	100%	-	-	-	-	-
E13	5,79	2,12	70%	0,471	0,329	0,019	0,014	1,00E+07
E14	5,61	2,17	72%	0	0	0	0	3,32E+07

Table D.3 – OBRE  $c = \sqrt{2}$

	$\hat{\mu}$	$\hat{\sigma}$	$F_{\hat{\theta}}(H)$	KS	CVM	AD2	AD2up	
E1	2,89	3,76	86%	0,285	0,145	0,026	0,174	2,36E+08
E2	6,54	2,04	57%	0,03	0,013	0	0,783	1,11E+07
E3	6,02	1,45	0%	0	0	0	0,016	5,10E+05
E4	8,86	0,62	0%	0	0	0	0	1,26E+06
E5	-9,01	5,06	100%	0,002	0,007	0,006	0,292	-
E6	0,72	3,63	96%	0	0	0,004	0,054	8,93E+07
E7	-3,52	3,66	100%	0,048	0,093	0,108	0,116	-
E8	-5,30	3,64	100%	0	0	0,001	0,003	-
E9	-28,17	8,57	100%	-	-	-	-	-
E10	7,45	1,15	32%	0,076	0,081	0,063	0,081	2,67E+05
E11	11,45	1,91	1%	0,412	0,253	0,162	0,201	1,63E+08
E12	-18,82	8,02	100%	-	-	-	-	-
E13	5,45	2,24	74%	0,481	0,315	0,025	0,015	1,29E+07
E14	4,62	2,51	82%	0	0	0	0	6,32E+07

Table D.4 – OBRE  $c = 2$ 

	$\hat{\mu}$	$\hat{\sigma}$	$F_{\hat{\theta}}(H)$	KS	CVM	AD2	AD2up	
E1	4,23	3,33	79%	0,286	0,076	0,007	0,27	1,06E+08
E2	6,58	2,03	56%	0,02	0,016	0,006	0,755	1,09E+07
E3	5,86	1,67	0%	0	0	0	0,108	8,82E+05
E4	8,41	1,74	19%	0	0	0	0,026	1,09E+07
E5	-9,09	5,06	100%	0,001	0,005	0,008	0,278	-
E6	1,65	3,41	94%	0	0	0,001	0,054	6,53E+07
E7	-3,51	3,66	100%	0,044	0,085	0,103	0,088	8,25E+06
E8	-5,18	3,62	100%	0	0	0	0,001	-
E9	-28,88	8,59	100%	-	-	-	-	-
E10	7,52	1,03	27%	0,168	0,073	0,068	0,146	2,04E+05
E11	11,43	1,74	0%	0,189	0,214	0,216	0,471	8,34E+07
E12	-18,77	8,02	100%	-	-	-	-	-
E13	4,73	2,49	81%	0,444	0,157	0,013	0,086	2,12E+07
E14	2,72	3,04	92%	0	0	0	0,01	1,78E+08

Table D.5 – OBRE  $c = 5$ 

	$\hat{\mu}$	$\hat{\sigma}$	$F_{\hat{\theta}}(H)$	KS	CVM	AD2	AD2up	
E1	7,87	2,18	33%	0	0	0	0,388	1,77E+07
E2	7,67	1,65	32%	0	0	0	0,067	6,44E+06
E3	5,34	1,92	0%	0	0	0	0,027	1,31E+06
E4	8,65	1,49	12%	0	0	0	0,007	5,81E+06
E5	6,94	2,08	49%	0	0	0	0,01	2,20E+07
E6	7,31	1,96	42%	0	0	0	0,12	8,17E+06
E7	7,00	1,53	48%	0	0	0	0	9,85E+05
E8	6,76	1,50	54%	0	0	0	0	7,96E+06
E9	5,07	3,10	72%	0	0	0,005	0,047	4,67E+07
E10	7,64	0,90	21%	0,033	0,01	0,009	0,111	1,66E+05
E11	11,39	1,71	0%	0,07	0,157	0,141	0,424	7,35E+07
E12	6,95	3,06	49%	0	0	0	0,019	4,07E+08
E13	7,32	1,71	40%	0	0	0	0,006	6,90E+06
E14	7,25	1,77	42%	0	0	0	0	2,72E+07

Table D.6 – QD scindé

	$\hat{\mu}$	$\hat{\sigma}$	$F_{\hat{\theta}}(H)$	KS	CVM	AD2	AD2up	
E1	6,67	2,54	54%	0,037	0,043	0	0,472	2,70E+07
E2	6,67	1,99	55%	0,068	0,196	0	0,728	1,01E+07
E3	5,69	1,70	0%	0	0	0	0,056	2,31E+07
E4	8,66	1,47	12%	0	0	0	0,008	5,53E+06
E5	-1,20	3,86	98%	0	0,004	0	0,522	2,39E+08
E6	5,11	2,54	76%	0	0	0	0,135	1,68E+07
E7	2,29	2,66	96%	0,22	0,305	0	0,064	3,38E+06
E8	-41,88	6,92	100%	-	-	-	-	-
E9	-67,45	9,65	100%	-	-	-	-	-
E10	7,75	0,85	16%	0,027	0,02	0,025	0,122	1,62E+05
E11	11,39	1,71	0%	0,132	0,243	0,253	0,401	7,27E+07
E12	-10,78	6,80	100%	0,823	0,847	0,822	0,259	3,39E+10
E13	2,77	2,95	92%	0,08	0,027	0,009	0,28	4,23E+07
E14	-7,88	4,79	100%	0	0	0	0,025	-

Table D.7 – QD direct

	$\hat{\mu}$	$\hat{\sigma}$	$F_{\hat{\theta}}(H)$	KS	CVM	AD2	AD2up	
E1	-9,73	6,35	100%	0,08	0,086	0	0,335	5,83E+09
E2	6,50	2,07	58%	0,001	0,003	0	0,853	1,23E+07
E3	5,98	1,39	0%	0	0	0	0,013	1,81E+07
E4	7,30	2,60	44%	0	0	0	0,041	1,46E+08
E5	-64,42	10,17	100%	-	-	-	-	-
E6	-77,02	11,58	100%	-	-	-	-	-
E7	-2,58	3,55	100%	0,016	0,065	0	0,416	8,59E+06
E8	-3,34	3,38	100%	0	0,007	0	0,008	-
E9	-36,68	5,59	100%	-	-	-	-	-
E10	7,13	1,28	43%	0,037	0,089	0,118	0,264	3,09E+05
E11	11,37	1,87	1%	0,158	0,16	0,266	0,464	1,28E+08
E12	-10,60	6,69	100%	0,846	0,597	0,617	0,373	2,52E+10
E13	5,64	2,18	72%	0,362	0,222	0,026	0,032	1,15E+07
E14	5,87	2,09	69%	0	0	0	0	3,06E+07

Table D.8 – CvM normal

	$\hat{\mu}$	$\hat{\sigma}$	$F_{\hat{\theta}}(H)$	KS	CVM	AD2	AD2up	
E1	-45,97	5,62	100%	-	-	-	-	-
E2	-476,64	57,70	100%	-	-	-	-	-
E3	5,98	1,39	0%	0	0	0	0,013	1,82E+07
E4	-677,88	79,78	100%	-	-	-	-	-
E5	-7,01	1,61	100%	-	-	-	-	-
E6	-911,36	109,83	100%	-	-	-	-	-
E7	-65,81	7,76	100%	-	-	-	-	-
E8	-302,15	33,69	100%	-	-	-	-	-
E9	-22,40	3,45	100%	-	-	-	-	-
E10	-79,32	6,04	100%	-	-	-	-	-
E11	11,37	1,87	1%	0,386	0,279	0,317	0,364	1,34E+08
E12	-47,22	5,68	100%	-	-	-	-	-
E13	-230,81	11,07	100%	-	-	-	-	-

Table D.9 – CvM seuil