

HEC Montréal
Affiliée à l'Université de Montréal

**La Quantification du Risque Opérationnel
des Institutions Bancaires**

par

Hela Dahen

Département Finance

Thèse présentée à la Faculté des études supérieures en vue d'obtention de grade
de Philosophiae Doctor (Ph.D) en Administration

Avril, 2007

Copyright, Hela Dahen, 2006

HEC Montréal
Université de Montréal

Cette thèse intitulée :

La Quantification du Risque Opérationnel des Institutions Bancaires

Présentée par :

Hela Dahen

est évaluée par le jury composé des personnes suivantes :

Dr. Narjess Boubakri
Président-rapporteur

Dr. Georges Dionne
Directeur de recherche

Dr. Bruno Rémillard
Co-directeur de recherche

Dr. François Bellavance
Membre du jury

Dr. Georges Hübner
Examineur externe

RÉSUMÉ

Suite à l'accord de Bâle II, les banques sont invitées à développer leur propre méthode de mesure de capital pour le risque opérationnel. Nous proposons dans cette thèse une méthode de quantification du risque opérationnel en intégrant des données internes et externes de pertes. Cette thèse est composée de trois parties :

La première partie propose des solutions quant au problème d'utilisation des données externes de pertes opérationnelles dans la mesure du capital avec une méthode avancée. Dans une première étape, nous mettons en place un modèle explicatif des montants de pertes incluant la taille de l'entreprise, le lieu de la perte, les lignes d'affaires ainsi que les types de risque. Les résultats de l'estimation par la méthode des moindres carrés ordinaires (MCO) montrent que ces variables ont un pouvoir statistiquement significatif dans l'explication des montants de pertes. Ces dernières vont être retenues pour le développement d'une formule de normalisation. Nous montrons ainsi comment il est possible de calculer la perte équivalente à une perte externe, qui pourrait être subie au niveau d'une banque. D'ailleurs nous validons notre méthode en prenant le cas de la banque Merrill Lynch et nous montrons que les statistiques des pertes observées de cette banque sont sensiblement proches de celles trouvées après mise à échelle.

Dans une seconde étape, nous développons un deuxième modèle pour la mise à l'échelle des fréquences des pertes qui pourraient avoir lieu durant une période déterminée à partir des données externes. Nous proposons deux modèles tronqués avec composante de régression, à savoir le Poisson et le binomial négatif. Des variables estimant la taille et la répartition géographique des activités des banques ont été introduites comme variables explicatives dans le modèle. Les résultats montrent que la distribution binomiale négative domine la distribution Poisson. Ainsi, la mise à l'échelle se fait en calculant les paramètres de la distribution retenue à partir des coefficients estimés et des variables propres à une banque donnée. Il est donc possible de générer les fréquences des pertes extrêmes sur un horizon déterminé, même lorsque la banque étudiée n'a pas suffisamment de données.

Dans la seconde partie, nous étudions le degré d'ajustement de différentes distributions paramétriques aux données de pertes opérationnelles externes mises à l'échelle d'une banque canadienne. En effet, cette étape est très importante pour le calcul de la VaR avec la méthode avancée, à savoir Loss Distribution Approach (LDA). Nous testons les distributions exponentielle, lognormale, Weibull et une famille de distributions à quatre paramètres : la GB2. Les résultats d'estimation des paramètres rejettent l'ajustement de ces distributions aux données. Par ailleurs, nous validons l'existence d'une symétrie pondérée dans les données que nous combinons avec le modèle GB2. La qualité d'ajustement du modèle GB2 fractionné (ainsi construit) aux données est évaluée grâce à des tests d'ajustement avec *bootstrap* paramétrique. Nous montrons que le modèle GB2 fractionné offre un excellent ajustement et décrit mieux le comportement des pertes opérationnelles.

Quant à la troisième partie, nous déterminons la valeur à risque opérationnel (VaR) d'une banque canadienne avec la méthode LDA. Nous développons cette méthode de mesure du capital opérationnel en choisissant la meilleure distribution de sévérité et de fréquence tout en tenant compte du seuil de troncature, puisque les données sont collectées à partir d'un certain seuil. Nous testons la robustesse de notre modèle en le comparant au modèle standard, qui est construit à partir des distributions fréquemment utilisées à savoir la lognormale et la Poisson et qui ne tient pas compte des pertes au-dessous du seuil de collecte. Les résultats montrent que le modèle standard sous-estime significativement la VaR. De plus, nous proposons un algorithme pour la combinaison des données internes et externes mises à l'échelle. Ainsi, la VaR calculée constitue une bonne mesure de l'exposition réelle face au risque opérationnel d'une banque.

Mots clés : Risque opérationnel, institutions bancaires, mise à l'échelle, modèles de comptage tronqués avec composante de régression, approche de la distribution des pertes, symétrie pondérée, distribution tronquée, GB2, valeur à risque opérationnel, *bootstrap* paramétrique.

ABSTRACT

Operational risk has received an increasing emphasis in the recent years. The BIS¹, through Basel Accord II, has urged banks to develop their own advanced capital measurement approach (AMA) to manage operational risk. We propose in this thesis a method to quantify operational risk by combining internal and external loss data. This dissertation is composed of three chapters:

In the first chapter, we propose a severity scaling model that allows us to estimate the equivalent operational loss amount that can occur in a bank, from an external data base. We take into consideration the size of the institution, the location, the business lines and the risk types. The results demonstrate that these variables have a significant impact to explain the observed losses. Considering a normalization formula with the significant variables, we show how we determine the equivalent loss amount that could occur in a bank.

Secondly, we develop a frequency scaling model to determine the number of extreme losses that a bank could have during a specified period. We suggest two count models: the truncated Poisson regression and the truncated negative binomial regression. The explanatory variables of the regression are the size and the geographical distribution of the banking activities. We show that the second model yields to more accurate results. Therefore, the frequency scaling consists on calculating the parameters of the distribution from the estimated coefficients and the bank variables. We can, thus, generate a number of extreme losses during a determined horizon, even if the bank doesn't have sufficient loss data.

In the second chapter, we investigate the fitting of parametric distributions to the operational loss data. Indeed, the quality of the fitting is very important to obtain accurate results of operational risk capital. We propose four severity distributions: exponential, lognormal, Weibull and a four parameters family distribution, the GB2. We demonstrate that the goodness of fit tests reject the fit of all the estimated

¹ Bank for International Settlements

distributions to scaled external data. Moreover, we test the existence of a weighted symmetry in the data. Then, we build up a split GB2 model by introducing the weighted symmetry in the GB2 distribution. We test the goodness of fit of this model by a Kolmogorov-Smirnov test with parametric bootstrap. The results show that the split GB2 model offers an excellent fit to data and describe better the behaviour of operational losses.

The third chapter proposes a methodology to determine the operational value at risk (VaR) with the Loss Distribution Approach (LDA), with an empirical study of a Canadian bank. We determine first the best severity distribution that fits the loss amounts, among the four distributions tested. We take into consideration the truncation issue, since the internal losses are collected above a known threshold. Secondly, we consider two frequency distributions and we select the one that offers the best fitting to the number of losses per day or per week. Moreover, we adjust the estimated parameters to take into account the frequency of losses under the threshold. Then, we define a standard model, often used in practice, which is built from an aggregation of the lognormal and the Poisson distributions, without any consideration of truncation. We compare our model to the standard one and we show that the latter underestimates significantly the VaR comparatively to our model. So, our model reflects better the bank exposure to operational risk since the VaR developed is more accurate and more realistic than the standard one. We also propose an algorithm to calculate the VaR by risk type, by combining internal and scaled external loss data and by aggregating the appropriate severity and frequency distributions.

Key words : Operational risk, banking institutions, scaling models, truncated count model regression, Loss Distribution Approach, weighted symmetry, truncated distribution, GB2, Operational Value at Risk, parametric bootstrap.

TABLE DES MATIERES

LISTE DES TABLEAUX.....	VIII
LISTE DES GRAPHIQUES.....	IX
REMERCIEMENTS	XI
INTRODUCTION GÉNÉRALE	1
1. MODÈLES DE MISE À L'ÉCHELLE DES MONTANTS ET DES FRÉQUENCES DES DONNÉES DE PERTES OPÉRATIONNELLES EXTERNES	4
1.1. INTRODUCTION	4
1.2. MISE EN CONTEXTE	5
1.2.1. <i>Le cadre réglementaire.....</i>	<i>5</i>
1.2.2. <i>Les sources de données externes et leurs biais potentiels.....</i>	<i>6</i>
1.2.3. <i>Revue de la littérature</i>	<i>9</i>
1.3. DESCRIPTION DES DONNÉES EXTERNES.....	14
1.4. HYPOTHÈSES DU MODÈLE.....	16
1.5. MODÈLE DE MISE À L'ÉCHELLE DES MONTANTS DE PERTES EXTERNES	17
1.5.1. <i>Modèle théorique de mise à l'échelle</i>	<i>17</i>
1.5.2. <i>Description des variables</i>	<i>19</i>
1.5.3. <i>Régression linéaire</i>	<i>23</i>
1.5.4. <i>Résultats de la régression.....</i>	<i>24</i>
1.5.5. <i>Tests de robustesse sur la variable taille</i>	<i>25</i>
1.5.6. <i>Formule de normalisation</i>	<i>26</i>
1.5.7. <i>Validation du modèle de mise à échelle.....</i>	<i>28</i>
1.6. MODÈLE DE MISE À L'ÉCHELLE DES FRÉQUENCES DES PERTES EXTERNES	30
1.6.1. <i>Description du modèle.....</i>	<i>31</i>
1.6.2. <i>Description des variables</i>	<i>32</i>
1.6.3. <i>Modèle Poisson tronqué avec composante de régression.....</i>	<i>33</i>
1.6.4. <i>Modèle binomial négatif tronqué avec composante de régression.....</i>	<i>35</i>
1.6.5. <i>Présentation et comparaison des résultats</i>	<i>37</i>
1.7. CONCLUSION	40
ANNEXE 1	52

2. MODÈLE GB2 FRACTIONNÉ POUR UN MEILLEUR AJUSTEMENT DE LA DISTRIBUTION DE SÉVÉRITÉ	54
2.1. INTRODUCTION	54
2.2. REVUE DE LA LITTÉRATURE	55
2.3. DESCRIPTION DES DONNÉES	59
2.4. MODÈLES D'AJUSTEMENT	60
2.4.1. Distributions proposées	60
2.4.2. Estimation des paramètres.....	63
2.4.3. Tests d'ajustement	64
2.4.4. Résultats.....	66
2.5. UN MODÈLE GB2 FRACTIONNÉ	67
2.5.1. Définition de la symétrie pondérée	68
2.5.2. Estimation des paramètres de la symétrie pondérée.....	69
2.5.3. Test de la symétrie pondérée	69
2.5.4. Modèle GB2 fractionné.....	71
2.6. CONCLUSION	74
ANNEXE 2	85
3. DÉTERMINATION DE LA VALEUR À RISQUE OPÉRATIONNEL D'UNE BANQUE CANADIENNE	87
3.1. INTRODUCTION	87
3.2. REVUE DE LA LITTÉRATURE	88
3.3. LES DONNÉES.....	93
3.3.1. Description des données	93
3.3.2. Statistiques descriptives sur les montants de pertes	94
3.3.3. Statistiques descriptives sur les fréquences des pertes	95
3.4. LE MODÈLE LDA	97
3.4.1. Conditions d'application du modèle.....	97
3.4.2. Présentation du modèle LDA.....	98
3.5. ESTIMATION DE LA DISTRIBUTION DE SÉVÉRITÉ	99
3.5.1. Distributions testées.....	100
3.5.2. Modélisation de l'ensemble des données	101
3.5.3. Découpage de la distribution.....	104
3.5.4. Tests de bon ajustement	110
3.5.5. Résultats de l'estimation des paramètres	110
3.6. ESTIMATION DE LA DISTRIBUTION DE FRÉQUENCE	113
3.6.1. Distributions testées.....	114
3.6.2. Estimation des paramètres.....	114
3.6.3. Correction des paramètres	115
3.6.4. Tests de bon ajustement	117
3.6.5. Résultat de l'estimation des paramètres	119

3.7.	CALCUL DE LA VAR OPÉRATIONNEL PAR TYPE DE RISQUE.....	120
3.7.1.	<i>Agrégation des distributions avec les données internes</i>	<i>120</i>
3.7.2.	<i>Comparaison de notre modèle avec le modèle standard</i>	<i>121</i>
3.7.3.	<i>Combinaison des données internes et données externes.....</i>	<i>123</i>
3.7.4.	<i>Détermination de la VaR</i>	<i>125</i>
3.7.5.	<i>Impact de la combinaison des données internes et externes sur les résultats.....</i>	<i>126</i>
3.8.	CONCLUSION	127
	CONCLUSION GÉNÉRALE.....	143
	BIBLIOGRAPHIE	147

LISTE DES TABLEAUX

Tableau 1-1 : Statistiques descriptives sur les données de la base externe	42
Tableau 1-2 : Résultats de l'estimation des paramètres de la régression linéaire	44
Tableau 1-3 : Tests de robustesse.....	46
Tableau 1-4 : Statistiques sur les pertes mises à l'échelle.....	47
Tableau 1-5 : Impact des variables de mise à l'échelle sur les montants des pertes ..	48
Tableau 1-6 : Statistiques descriptives sur les variables introduites dans le modèle de fréquence	49
Tableau 1-7 : Résultats de l'estimation des coefficients dans le modèle de fréquence	50
Tableau 1-8 : Résultats de l'application de la mise à l'échelle des fréquences.....	51
Tableau 2-1: Distributions étudiées.....	61
Tableau 2-2: Statistiques descriptives sur les données.....	77
Tableau 2-3 : Résultats de l'estimation des paramètres	82
Tableau 2-4 : Résultats de l'estimation des paramètres du modèle GB2 fractionné..	84
Tableau 3-1 : Statistiques descriptives	129
Tableau 3-2 : Estimation des distributions de sévérité pour les types de risque FI, ATCE, EPSE.	131
Tableau 3-3: Estimation des distributions de sévérité pour le type de risque CPPC	132
Tableau 3-4 : Estimation des distributions de sévérité pour le type de risque GELP	133
Tableau 3-5 : Estimation des distributions de sévérité pour le type de risque FE ...	135
Tableau 3-6 : Estimation des distributions de fréquences par type de risque	137
Tableau 3-7 : Comparaison entre notre modèle et le modèle standard.	139
Tableau 3-8 : La VaR opérationnel par type de risque.....	140
Tableau 3-9 : Comparaison entre la VaR opérationnel par type de risque calculée avec le modèle Poisson et celle calculée avec le modèle binomial négatif	141
Tableau 3-10 : Impact de l'intégration des données de pertes externes sur la VaR.	142

LISTE DES GRAPHIQUES

Graphique 2-1 : La GB2 et ses cas particuliers	76
Graphique 2-2 : Tests graphiques de la distribution exponentielle	78
Graphique 2-3 : Tests graphiques de la distribution lognormale.	79
Graphique 2-4 : Tests graphiques de la distribution Weibull.....	80
Graphique 2-5 : Tests graphiques de la distribution GB2.....	81
Graphique 2-6 : Tests graphiques du modèle GB2 fractionné.	83

*À ma famille, si lointaine mais si proche de mon cœur,
À Mehdi pour son amour et soutien inconditionnel,
À ma très chère petite fille Ines,*

Je vous aime profondément

REMERCIEMENTS

On dit souvent que le trajet est aussi important que la destination. Les cinq années de doctorat m'ont permis de bien comprendre la signification de cette phrase toute simple. Ce parcours, en effet, ne s'est pas réalisé sans défis et sans soulever de nombreuses questions pour lesquelles les réponses nécessitaient de longues heures de travail.

Je tiens tout d'abord à exprimer ma profonde gratitude à mon directeur de recherche, Dr. Georges Dionne, pour son encadrement exemplaire, pour ses conseils précieux et pour sa grande disponibilité

J'aimerais remercier Dr. Bruno Rémillard, mon co-directeur de recherche, pour ses commentaires pertinents et pour tous les apprentissages mathématiques et statistiques qu'il a su me faire transférer.

Un grand merci à Dr. François Bellavance et Dr. Susan Christoffersen pour leur disponibilité et pour leurs commentaires judicieux.

Je remercie également Dr. Georges Hübner, l'examineur externe, pour tous ses commentaires et remarques très pertinents visant à améliorer cette recherche. Un grand merci à Dr. Narjess Boubakri, président rapporteur, pour toute sa gentillesse et tous ses encouragements.

Je suis hautement reconnaissante à tout le personnel du département de gestion du risque opérationnel de la banque étudiée. Grâce à l'opportunité qu'ils m'ont offert de travailler à la banque, j'ai pu réaliser la partie empirique de cette thèse.

J'aimerais remercier également le ministère tunisien de l'enseignement supérieur, l'Institut de Finance Mathématique (IFM2), le Centre de Recherche en E-Finance (CREF), et le Fonds Québécois de la Recherche sur la Société et la Culture (FQRSC) pour leur soutien financier.

Un grand merci à mon père qui m'a inculqué la valeur du travail sérieux. Grâce à ses encouragements, j'ai pu surmonter les moments difficiles de mon cursus doctoral. J'espère que tu seras fier de moi, papa! Mes pensées vont à ma tendre maman qui n'a pas arrêté de prier pour mon succès. Toutes les larmes qu'elle a versées à cause de la longue séparation m'ont poussée encore à aller de l'avant. Je remercie également mon frère et mes deux sœurs pour leurs encouragements et soutien inconditionnel.

Je veux témoigner ma profonde gratitude à mon cher époux pour son amour, son soutien et encouragement. C'est la personne qui a plus enduré mes frustrations et mes sauts d'humeur dans les moments difficiles de mon cheminement doctoral. Merci beaucoup pour ta patience et d'être toujours à mes côtés. Merci à mon adorable petite fille Ines, qui malgré son jeune âge, m'encourageait à continuer d'étudier.

Je voudrais remercier également ma belle famille pour son aide précieuse et pour sa compréhension. Elle m'a beaucoup aidé à concilier travail, étude et famille.

Un grand merci à mes amis et collègues de la Chaire du Canada en gestion des risques, pour leur soutien et commentaires judicieux, notamment, Nadia, Olfa, Khemais, Denitsa, Hakim, Oussama et Fatoumata.

INTRODUCTION GÉNÉRALE

Durant ces dernières années, nous remarquons un intérêt croissant des institutions financières pour identifier les pertes associées au risque opérationnel, et ce, suite à des considérations réglementaires d'une part, et suite à l'occurrence de pertes opérationnelles colossales dans le secteur financier d'autre part. Citons des exemples de pertes opérationnelles énormes subies dans le secteur financier : 2,4 milliards de dollars attribuables aux poursuites subséquentes à l'affaire Enron et une perte de 690 millions de dollars causée par une transaction non autorisée à Allied Irish Bank. Ajoutons le cas de la plus vieille banque du Royaume-Uni (233 ans), la Barings, qui a fait faillite à la suite d'activités non autorisées ayant occasionné une perte de 1,3 milliard de dollars. Ces exemples montrent l'ampleur de ce risque. Ils constituent également un signal pour alerter les institutions financières qui doivent impérativement le définir, le mesurer et le gérer afin d'éviter les éventuelles pertes colossales qui peuvent en découler.

Outre ces pertes importantes, le risque opérationnel touche toutes les activités et les opérations des institutions financières de différentes manières. En effet, nous trouvons des événements opérationnels attribuables aux personnes, aux processus, aux systèmes et aux événements externes. En revanche, les unités ne sont pas touchées de la même façon par le risque opérationnel. L'impact varie selon la nature des activités et les différents intervenants. Le risque opérationnel prend donc de plus en plus d'envergure et sa gestion devient une nécessité.

Conscientes de ce grand risque, les autorités réglementaires ont lancé le débat sur la définition, l'identification, la mesure et la gestion du risque opérationnel à partir de juin 1999. Elles introduisent ainsi de la pression sur les banques afin qu'elles mettent en place un cadre de gestion propre au risque opérationnel (système de gestion de risque, senior management, suffisamment de ressources dédiées à la gestion du risque opérationnel dans les lignes d'affaires). Ce cadre permet, entre autres, l'identification des pertes et la mesure d'un capital opérationnel. Une façon de couvrir l'exposition

au risque opérationnel est de détenir un capital permettant de couvrir les pertes non anticipées, comme c'est le cas pour le risque de marché et de crédit.

Plusieurs approches de mesure de capital risque opérationnel ont été proposées par les autorités réglementaires, sauf que les banques sont invitées à développer leur propre méthode, une méthode de mesure avancée qui reflètera mieux le niveau de risque opérationnel. Le développement d'une telle méthode de mesure est au centre de cette thèse. Nous étudions en effet différentes facettes de la quantification du risque opérationnel des institutions bancaires dans le but de développer une mesure qui tient compte de l'exposition réelle d'une banque. Ainsi, l'objectif de cette thèse est de proposer une méthode robuste pour le calcul de la valeur à risque opérationnel qui sera la plus réaliste et la plus représentative du niveau de risque opérationnel d'une banque.

Comme les recherches dans ce domaine sont encore en phase embryonnaire, tous les développements et les outils de quantification que nous proposons dans cette thèse seront ainsi d'une grande utilité pour les institutions financières à court terme, étant donné qu'elles ont des exigences à remplir, mais également pour les autres industries qui, à moyen terme, trouveront que l'exercice leur est profitable.

Cette thèse portant sur la quantification du risque opérationnel et le développement d'une mesure de capital comporte trois parties.

La première partie consiste à développer des modèles pour mettre la sévérité et la fréquence des pertes externes de plus d'un million de dollars à l'échelle d'une banque donnée. Le recours à des données de pertes opérationnelles externes s'avère essentiel pour compléter les données internes et, surtout, pour aller chercher les pertes importantes très rares qui, généralement, manquent dans les bases internes des banques. Nous introduisons dans notre modèle des variables de lieu, de ligne d'affaires où la perte a eu lieu et de type de risque en plus de la variable taille. Nous montrons ainsi que notre modèle vient améliorer les modèles existants dans la littérature qui se basent uniquement sur la taille. D'autre part, nous avons mis en place un modèle original pour la mise à l'échelle et l'ajustement du nombre de pertes

externes de plus d'un million de dollars sur une période déterminée. Les paramètres du modèle retenu à savoir le binomial négatif tronqué avec composante de régression dépendent de la taille de l'institution financière et de la répartition géographique de ses activités.

La deuxième partie de la thèse met l'accent sur l'estimation de la distribution de sévérité. Le choix de la distribution est en effet d'une importance cruciale pour bien décrire le comportement des pertes et estimer correctement la valeur à risque opérationnel. Nous mettons en place à cet effet un modèle GB2 fractionné en introduisant la notion de symétrie pondérée dans les données de pertes opérationnelles. Les résultats montrent que ce modèle offre un excellent ajustement à la distribution empirique des pertes mises à l'échelle d'une banque canadienne. Alors que l'ajustement des distributions usuelles ainsi que celui de la distribution GB2 à quatre paramètres n'ont pas été retenues pour modéliser ces pertes.

Quant à la troisième partie, nous portons notre attention à la Valeur à risque opérationnel et nous procédons à toutes les étapes nécessaires à son calcul. En effet, nous étudions l'ajustement des distributions paramétriques de sévérité et de fréquence tout en considérant le montant et la fréquence des pertes au-dessous du seuil de collecte, et ce, avec les données réelles d'une banque canadienne. Ces éléments sont en effet négligés dans le modèle standard fréquemment utilisé, construit à partir des distributions lognormale et Poisson. Une comparaison des deux modèles montre que le modèle standard sous estime énormément la Valeur à risque opérationnel par rapport à celle calculée avec notre modèle. De plus, dans le but d'avoir une mesure rigoureuse et réaliste de la perte opérationnelle non anticipée, nous intégrons des pertes opérationnelles externes mises à l'échelle de la banque à l'étude. Ceci permet de tenir compte de certaines pertes extrêmes éventuelles n'ayant pas encore été subies. En effet, une analyse de l'impact de la combinaison des pertes externes et internes sur le calcul de la VaR a été faite dans cette partie et a révélé que les VaR, à quantiles élevés, sont considérablement sous estimées lorsqu'elles sont calculées avec les données de pertes internes seulement.

1. MODÈLES DE MISE À L'ÉCHELLE DES MONTANTS ET DES FRÉQUENCES DES DONNÉES DE PERTES OPÉRATIONNELLES EXTERNES

1.1. Introduction

L'une des approches proposées dans l'accord de Bâle II pour la quantification du risque opérationnel est l'approche avancée. Le développement d'une telle approche exige une large base de données. Les données peuvent provenir de différentes sources. En fait, les données internes sont d'une grande utilité pour refléter le degré d'exposition réelle face au risque opérationnel. Cependant, l'historique de collecte est court et les pertes opérationnelles observées sont loin d'être représentatives des pertes qu'une institution bancaire pourrait subir. En effet, les données internes d'une banque n'incluent pas suffisamment de pertes rares² avec une haute sévérité, d'autant plus que le processus de collecte des pertes en est encore à ses débuts.

Ainsi, le recours à des données de pertes opérationnelles externes s'avère essentiel pour compléter les données internes et, surtout, pour aller chercher les pertes importantes très rares qui, généralement, manquent dans les données internes. Il y a lieu donc d'inclure ces éventuelles pertes importantes dans la base interne d'une banque pour réduire l'effet « surprise » (le non anticipé) et calculer le capital adéquat. Il est évident que nous ne pouvons pas prédire les pertes extrêmes exactes lorsque celles-ci ne sont pas survenues. Cependant, il nous est possible, à partir des pertes enregistrées dans le secteur bancaire, de faire une projection pour une banque en tenant compte de certains facteurs pour la mise à l'échelle.

² Une perte rare est définie comme étant une perte découlant d'un événement à fréquence très faible.

Compte tenu de ce contexte et de la nécessité d'utiliser une base de données externe dans une approche avancée de calcul de capital, l'objectif de ce chapitre est de développer une méthode robuste pour prévoir, à partir des données externes, la sévérité ainsi que les fréquences des pertes qu'une banque pourrait subir. Plusieurs facteurs seront pris en considération pour expliquer les montants de pertes ainsi que leur nombre sur une période déterminée. Il s'agit donc d'une projection des pertes externes survenues dans l'industrie au niveau d'une banque.

La méthode développée dans ce chapitre a été testée sur les données d'une base externe de pertes opérationnelles de plus de 1 million de dollars. Cependant, elle reste applicable à n'importe quelle base de données externe. Une combinaison des pertes externes mises à l'échelle avec les données internes d'une banque permet de refléter l'exposition face au risque opérationnel.

Ce chapitre est présenté comme suit. La deuxième section décrit les différentes approches de mesure de capital, les sources de données, leurs caractéristiques, ainsi qu'un bref survol des méthodes de mise à l'échelle dans la littérature. Une description des données de la base externe est présentée dans la troisième section. La quatrième section expose les hypothèses du modèle. Ensuite, le modèle de mise à l'échelle de la sévérité est développé dans la cinquième section. L'avant-dernière section est consacrée au développement du modèle de mise à l'échelle des fréquences. Enfin, une conclusion et une discussion sur les avenues possibles de recherches sont présentées à la fin de ce chapitre.

1.2. Mise en contexte

1.2.1. Le cadre réglementaire

En 2001, le Comité de Bâle a défini le risque opérationnel comme étant le risque de pertes résultant de l'inadéquation ou de la défaillance des processus, d'individus et de systèmes, ou résultant d'événements externes. Le risque juridique est aussi inclus, mais la définition ne prend pas en compte le risque de réputation ni le risque d'affaires.

Les autorités réglementaires ont identifié trois approches différentes pour le calcul du capital du risque opérationnel. Elles vont de l'approche la plus simple à une autre plus complexe, plus sensible au risque. Nous présentons ici l'approche avancée reliée à notre problématique de recherche.

Il s'agit d'une approche plus sophistiquée qui repose sur des méthodes internes de calcul de capital adoptées par les banques. Les autorités réglementaires offrent une grande flexibilité quant au choix de la méthode, pourvu qu'elle combine adéquatement des critères qualitatifs et quantitatifs et estime raisonnablement la perte non anticipée en se basant sur la combinaison de données internes, des données externes pertinentes, des analyses par scénarios ainsi que des facteurs internes sur l'environnement de contrôle). La méthode retenue doit refléter le niveau d'exposition aux risques de l'institution financière et doit être approuvée par les régulateurs avant son implantation.

Trois méthodes sont proposées dans le cadre de l'approche avancée, à savoir : 1) la méthode interne de mesure (*Internal Measurement Approach, IMA*) ; 2) la méthode des distributions des pertes (*Loss Distribution Approach, LDA*) ; et 3) la méthode par carte de pointage (*Scorecard*).

Nous nous intéressons dans cette étude à la méthode des distributions des pertes (*LDA*) qui est la plus populaire. En effet, elle s'inspire énormément des méthodes actuarielles développées dans le domaine de l'assurance. Elle repose sur l'estimation de la perte non anticipée à partir de la modélisation des montants et des fréquences des pertes opérationnelles. La bonne combinaison des données des pertes internes et externes constitue donc une étape importante à considérer dans une approche avancée.

1.2.2. Les sources de données externes et leurs biais potentiels

Nous portons notre attention sur l'utilisation des données externes. Une mise à l'échelle de ces dernières permet de les combiner avec les données internes pour obtenir une base de données représentative du risque opérationnel d'une banque. Ceci

fait partie de l'objectif de l'implantation d'une méthode avancée. Les sources de données externes sont encore limitées. Nous citons :

- Des données publiques obtenues à partir des rapports médiatisés et des magazines. Il s'agit des pertes de plus de 1 million de dollars. Il existe deux bases de données externes sur le marché (tel que Fitch). Le problème, avec ce type de données, est que la base ne contient que les pertes de haute sévérité survenues dans de grandes institutions financières. Le recours à cette base ne résout pas le problème de manque de données pour certains types de risque (perturbation des affaires et défaillance des systèmes), mais il permet de compléter la base avec des données extrêmes survenant rarement. Ces pertes formeront les queues des distributions, puisque les données internes de la plupart des institutions financières n'ont pas un historique représentatif des pertes importantes qui pourraient survenir. Le regroupement des données internes et externes requiert un certain traitement pour corriger le biais lié aux données.
- Des données fournies par les courtiers d'assurance (tel que Willis, Aon et Marsh). Il s'agit des pertes opérationnelles réclamées par les institutions financières. L'avantage majeur de cette source est sa fiabilité. En effet, comme les données sont collectées directement à partir des institutions financières, le biais de sélection est minime. Cependant, l'inconvénient de cette source est la différence des seuils de collecte liés aux différentes franchises des polices d'assurance, pas toujours observables. La deuxième limite réside dans la spécificité des types de risque collectés. En effet, seules les pertes de nature assurable seront comprises dans la base.
- Données non publiques obtenues à partir d'un rassemblement des données internes des banques. Ces dernières se sont entendues pour partager des informations sur leurs données. Elles constituent donc un consortium tel que ORX (*Operational Riskdata eXchange Association*). Cependant, vu la confidentialité des informations, les statistiques et les analyses sur les pertes

ne sont disponibles qu'aux participants. L'avantage de cette source de données est sa fiabilité. Le seuil de collecte est beaucoup plus bas que celui des sources précédentes. Ainsi, les montants de pertes sont plus comparables, surtout si les banques membres sont de taille moyenne. Toutefois, l'inconvénient majeur de cette source de données est qu'il est impossible d'accéder aux pertes événement par événement. Il n'est donc pas possible de construire une base avec les données internes et externes.

Les données externes contiennent beaucoup de biais, parmi lesquels :

- Le biais de sélection : Seules les pertes très importantes sont publiées. Ce biais est lié à la nature des bases disponibles et il est donc difficile à corriger.
- Le biais de contrôle : Les pertes proviennent des banques ayant différents environnements de contrôle. Il n'existe malheureusement pas de variables estimant la qualité de contrôle au niveau des banques dans les bases externes. Ainsi, avec les informations disponibles, il n'est pas possible de corriger ce biais.
- Le biais de collecte : Lorsque les données proviennent de différentes sources, des seuils différents peuvent entraîner des biais. Frachot et Roncalli (2002) et Baud, Frachot et Roncalli (2002) décrivent comment des données internes peuvent être comparées avec des données externes ayant des distributions de sévérité différentes.
- Le biais d'échelle : les pertes proviennent de différentes banques avec différentes tailles (actifs, revenus, nombre d'employés...) et localisées dans différents pays. La correction de ce biais fait l'objet de notre recherche.

Ainsi, nous proposons dans cette étude d'améliorer la méthode de mise à échelle existant dans la littérature. Il est donc à signaler que nous ne visons pas à corriger le biais de sélection ou les autres biais éventuels

1.2.3. Revue de la littérature

Peu de recherches ont été faites pour trouver une solution au problème d'échelle. Shih, Samad-Khan et Medapa (2000) ont introduit la taille de l'institution comme principal facteur de mise à l'échelle. Ils ont montré que la relation entre les pertes opérationnelles et la taille de la firme est non linéaire. En effet, la relation entre le logarithme du facteur d'échelle et les montants des pertes est plus importante que celle entre les pertes et la variable d'échelle brute. D'ailleurs, une banque deux fois plus grande qu'une autre ne va pas occasionner, en moyenne, des pertes deux fois plus importantes que les pertes subies par l'autre firme. Shih, Samad-Khan et Medapa (2000) supposent effectivement que la relation est la suivante :

$$L = R^{\alpha} \times F(\theta)$$

où :

L : le montant des pertes;

R : le revenu total de la firme où la perte a eu lieu;

α : un facteur de mise à l'échelle;

θ : un vecteur représentant tous les facteurs de risque non expliqués par R. $F(\theta)$ est donc un terme résiduel multiplicatif non expliqué par les fluctuations du revenu.

En prenant le logarithme de cette équation, nous obtenons une relation linéaire. Il est donc possible d'estimer le facteur de la mise à l'échelle α et le logarithme de la fonction des autres facteurs $F(\theta)$ qui constitue la constante de la régression.

Le seul facteur de risque inclus dans le modèle et estimant la taille de la firme est le revenu total. Il est donc probable que la majorité de la variabilité des pertes soit causée par d'autres facteurs tels que le type de la ligne d'affaires, la qualité de la gestion et l'efficacité de l'environnement de contrôle. Dans cette même étude, il a été montré que la taille n'explique qu'une petite portion (5 % environ) des montants de pertes.

Dans la même lignée, Hartung (2004) a développé une formule de normalisation qui permet de calculer l'équivalent d'une perte externe pour une banque donnée. La formule utilisée est la suivante :

$$Loss_{adj} = loss_{org} \cdot \left\{ 1 + a \cdot \left[\left(\frac{Scal.Param(Loss_{adj})}{Scal.Param(Loss_{org})} \right)^b - 1 \right] \right\}$$

où :

$Loss_{adj}$: le montant de la perte ajustée pour une banque donnée;

$Loss_{org}$: le montant original de la perte pour une banque de référence;

$Scal.Param(Loss_{adj})$: un paramètre de mise à l'échelle pour une banque donnée;

$Scal.Param(Loss_{org})$: un paramètre de mise à l'échelle pour une banque de référence;

a, b : des facteurs d'ajustement tels que $a \in [-1; 1]$, $b \in [0; 1]$.

Le paramètre de mise à l'échelle a été attribué selon la cause de l'événement. Des exemples de ce paramètre sont les revenus, le nombre d'employés ou la qualité de la gestion des risques. Concernant les facteurs d'ajustement, des hypothèses ont été émises quant à leur valeur selon le paramètre de mise à l'échelle. Les limites de ce modèle de mise à l'échelle consistent, d'une part, dans la non-justification théorique de la formule utilisée, et, d'autre part, dans l'absence de méthode développée pour estimer les facteurs d'ajustement.

Par ailleurs, le montant de la perte peut être décomposé en une composante commune et une composante idiosyncratique, selon l'étude de Na (2004). En effet, la composante commune à toutes les banques ou lignes d'affaires capte tous les changements dans l'environnement macroéconomique, géopolitique et culturel, alors que la composante idiosyncratique comprend tous les facteurs spécifiques à la ligne d'affaires ou propres à l'événement de perte. Une relation de puissance (*power relationship*) a été mise en place entre cette dernière composante et un estimateur de taille. Une formule de normalisation a été développée pour trouver le montant de

perte équivalent d'une perte prise comme référence, dans une banque donnée au niveau des lignes d'affaires B1 et B2. La formule est la suivante :

$$\frac{L_{T,B1}}{\left(R_{\text{idiosyncratique}}\right)_{T,B1}^{\lambda}} = \frac{L_{T,B2}}{\left(R_{\text{idiosyncratique}}\right)_{T,B2}^{\lambda}}$$

où :

$L_{T,Bi}$: un montant de perte survenue à la date T, à la banque ou la ligne d'affaires Bi;

$\left(R_{\text{idiosyncratique}}\right)_{T,Bi}$: le revenu de la banque ou de la ligne d'affaires Bi à la date T, et qui constitue le seul estimateur de la composante idiosyncratique;

λ : un facteur d'échelle.

Ce modèle peut être amélioré en introduisant d'autres facteurs pour la mise à l'échelle, autres que la taille de l'entreprise. Nous nous basons sur cette étude pour développer un modèle de mise à l'échelle de la sévérité des pertes en prenant en considération la taille, le lieu, la ligne d'affaires et le type de risque de perte.

Une fois que la mise à l'échelle de la sévérité est accomplie, il est également important de déterminer la fréquence des pertes normalisées sur un horizon de temps déterminé. Très peu de recherches ont essayé cette mise à l'échelle. En effet, certaines études ont développé des modèles de normalisation de la sévérité mais n'ont fait aucune considération de mise à l'échelle des fréquences (Shih, Samad-Khan et Medapa, 2001; Hartung, 2004). Hartung (2004) regroupe la fréquence des pertes de quatre banques sur un horizon de neuf ans. La banque prise comme référence aura une distribution identique à la distribution des quatre banques regroupées ensemble. Or, ceci n'est pas toujours le cas, vu que toutes ces banques ne sont pas nécessairement comparables. Plusieurs facteurs entrent donc en jeu dans la détermination de la fréquence des pertes.

Par ailleurs, Na (2004) a développé un modèle de mise à l'échelle des fréquences, équivalent à celui des sévérités. En effet, ce modèle stipule que, comme pour la sévérité, la fréquence des pertes peut être décomposée en composante commune et composante idiosyncratique estimée par la taille. Il conclut que la taille est un facteur significatif pour expliquer la variabilité du nombre de pertes. Cependant, la principale

limite du modèle est qu'il ne tient pas compte du caractère discret des données de fréquences comparativement aux données sur les montants de pertes.

Il existe certaines méthodes d'intégration des données internes et externes qui ont été développées dans la littérature telles que :

- Une estimation séparée des deux distributions basée sur des données internes et des données externes. La combinaison des deux distributions est faite par des techniques bayésiennes (Alexander, 2003)
- Création d'un échantillon large d'observations contenant des données internes et des données externes (Frachot et al., 2002)
- Améliorer la précision de la queue de la distribution basée sur l'information contenue dans la base externe (Chapelle et al., 2004)

Les deux premières méthodes exigent un seuil de collecte non élevé, Ce qui n'est pas le cas des bases externes. Dans notre recherche, nous allons nous inspirer de la méthode développée par Chappelle et al. (2004) mais en utilisant les distributions identifiées dans le premier et second chapitre de la thèse, estimant les montants et les fréquences des pertes mises à l'échelle.

Dans la présente étude, nous développons un modèle pour mettre à l'échelle d'une banque les fréquences de pertes extrêmes observées dans l'industrie. Nous proposons un modèle de comptage avec composante de régression. Un tel modèle permet de tenir compte du caractère discret et non négatif des données. Deux modèles, à savoir Poisson et binomial négatif, seront donc testés (Klugman, Panjer et Willmot, 1998; Cruz, 2001). La composante de régression contenue dans le modèle permet de tenir compte de certains facteurs pour la mise à échelle.

Dans les modèles considérés dans la littérature pour décrire les variables discrètes (Cox et Lewis, 1966; El Sayyad, 1973; Frome, Kutner et Beauchamp, 1973; Hausman, Hall et Griliches, 1984; Gouriéroux, Monfort et Trognon, 1984), les variables endogènes sont supposées avoir une distribution Poisson avec composante

de régression. Le paramètre de cette distribution est une fonction des valeurs des variables exogènes. Le choix de ce modèle est justifié lorsque la variable dépendante compte l'occurrence d'un événement donné durant une période déterminée et lorsque les hypothèses usuelles de la distribution Poisson sont satisfaites.

Plusieurs applications de ce modèle ont été faites dans la littérature, telles que la modélisation du nombre de brevets reçus par une firme (Hausman, Hall et Griliches, 1984), du nombre de consultations d'un médecin (Cameron, Trivedi, Milne et Piggott, 1988) ou du nombre d'accidents automobiles ou aériens (Dionne et Vanasse, 1989 et 1992; Dionne et al. 1997). L'application de Dionne et Vanasse (1989) est la première à introduire une composante de régression dans le domaine de l'assurance, un domaine qui présente beaucoup de similitudes avec le risque opérationnel. Le nombre d'accidents par individu est supposé suivre une distribution Poisson dont le paramètre varie d'une unité d'exposition à une autre. Ce dernier dépend effectivement des caractéristiques des unités exposées. Les coefficients de ces variables sont estimés par la méthode du maximum de vraisemblance telle que discutée par Maddala (1983) et Cameron et Trivedi (1986).

Le modèle Poisson avec composante de régression suppose l'équidispersion (l'égalité entre la moyenne et la variance conditionnelles). Cette restriction peut être non compatible avec les données. Le recours à la distribution binomiale négative³ permet de pallier à ce problème, puisqu'elle permet la surdispersion, caractéristique fréquemment observée dans les données. Les études de Dionne et Vanasse (1989 et 1992) et Boyer, Dionne et Vanasse (1991) ont montré la supériorité du modèle binomial négatif avec composante de régression par rapport au modèle Poisson avec composante de régression pour les accidents automobiles. D'ailleurs, le modèle binomial négatif avec composante de régression est maintenant fréquemment utilisé dans la littérature.

Il est possible d'estimer les paramètres des distributions par maximum de vraisemblance lorsque certaines conditions sont vérifiées. Dans le cas où la densité

³ C'est une distribution Poisson dont le paramètre est aléatoire suivant une distribution gamma.

est mal spécifiée, les estimateurs trouvés par maximum de vraisemblance ne sont pas cohérents. Gourieroux, Monfort et Trognon (1984a et 1984b) ont proposé d'autres méthodes pour pallier à ce problème telles que la pseudo maximum de vraisemblance (PML), la pseudo maximum de vraisemblance quasi-généralisée (QGPML). Ils ont spécifié les conditions sous lesquelles ces estimateurs PML et QGPML des modèles de famille exponentielle linéaire sont cohérents, dans le cas de modèles non tronqués. Toutefois, si la densité de la binomiale négative est correctement spécifiée, les estimateurs du maximum de vraisemblance seront plus efficaces que PML et QGPML (Dionne et Vanasse, 1992).

Nous allons donc appliquer ces modèles très utilisés dans la littérature pour la modélisation du nombre de pertes opérationnelles. Ces modèles permettront ainsi d'introduire des informations sur l'institution financière où la perte a eu lieu. Des variables exogènes de lieu et de répartition géographique de l'entreprise permettront la mise à l'échelle. À notre connaissance, c'est la première application de ces modèles au risque opérationnel et, plus précisément, pour la mise à l'échelle des fréquences des pertes opérationnelles. Cependant, nous observons que les fréquences sont supérieures à zéro. Nous développons donc des modèles Poisson et binomial négatif tronqués au point zéro avec composante de régression. La présentation des densités tronquées de ces modèles a été faite par Cameron et Trivedi (1998) et les tests de surdispersion pour les mêmes modèles ont été développés par Gurmu (1991) et Gurmu et Trivedi (1992).

1.3. Description des données externes

Il s'agit de la base de données OpVaR de Fitch, constituée de pertes opérationnelles de 1 million de dollars américains et plus. La base de données contient des pertes de toutes les industries. Une première sélection a donc été faite pour avoir seulement les pertes opérationnelles provenant d'institutions financières, puisque nous ciblons seulement les banques. La base de données contient les informations suivantes :

1. Le type d'événement, niveau 1 : il s'agit des types de risque définis par les autorités réglementaires. Cette rubrique comprend :

- fraudes externes;
- fraudes internes;
- clients, produits et pratiques commerciales;
- emploi, pratiques et sécurité environnementale;
- gestion de l'exécution, de la livraison et des processus;
- actifs tangibles et corporels endommagés;
- perturbation des affaires et défaillance des systèmes.

Nous avons également les types d'événements de niveau 2 et 3, qui offrent plus de précision et plus de granularité. Citons à titre d'exemple : la discrimination et la diversité comme sous-événement du type de risque 'emploi, pratiques et sécurité environnementale'. Comme événements de type 3 du sous-type de risque diversité et discrimination, nous citons, à titre d'exemples la discrimination due à l'âge, au sexe, à la race, à l'orientation sexuelle et au harcèlement sexuel.

2. Le nom de la compagnie mère et celui de la filiale.
3. Une description détaillée de l'événement de perte.
4. Le montant de la perte en devise locale, en dollars américains et sa valeur actuelle (en tenant compte de l'inflation).
5. La date d'événement. Cependant, nous retenons seulement l'année de l'événement, puisque nous avons des doutes quant à la date exacte de l'événement (jour et mois).
6. L'industrie : il s'agit soit des services financiers, soit de l'administration publique.
7. L'unité d'affaires, niveaux 1, 2 et 3. Le premier niveau distingue les institutions financières et non financières. Dans notre cas, nous nous intéressons seulement au secteur financier. Le niveau 2 concerne les institutions financières, faisant la distinction entre les banques, les assurances, les banques d'investissement et les autres institutions. Quant au niveau 3 des

unités d'affaires, nous retrouvons les lignes d'affaires définies par le comité de Bâle.

8. Le lieu, c'est-à-dire le pays où la perte a eu lieu.
9. Un code d'identification propre à chaque perte.
10. Des informations sur l'institution où la perte a eu lieu comme le total des actifs, le total des fonds propres, le total des dépôts, le total des revenus et le nombre d'employés.

Il est à mentionner qu'il existe des observations pour lesquelles les informations sur la firme manquent. Comme ces informations sont très importantes pour le modèle de mise à l'échelle plus loin, nous sommes obligés de retenir seulement les données de pertes pour lesquelles des informations spécifiques sur la firme sont disponibles. De plus, 1,8 % des pertes sont survenues entre 1981 et 1994 et ont une moyenne de 130,31 M\$ par événement, alors que les pertes ayant lieu après cette date ont une moyenne de 67,15 M\$. Nous écartons donc les événements ayant lieu avant 1994 de la base externe à cause d'un biais de collecte. Ainsi, 1 056 observations de pertes de plus d'un million de dollars américains sont retenues dans la base de données.

1.4. Hypothèses du modèle

Nous appliquons un modèle théorique de mise à l'échelle pour les sévérités et les fréquences. L'application empirique des modèles de mise à l'échelle sur les données externes est d'une grande utilité pour montrer la simplicité du modèle et pour pouvoir dégager des résultats. Il est clair que la base de données utilisée est sujette à plusieurs critiques pour les raisons citées auparavant. Comme il n'existe pas, à présent, de meilleures sources de données, nous émettons les hypothèses suivantes pour pouvoir mener notre recherche. Force est de constater que la méthodologie reste applicable à d'autres bases, à condition qu'elles contiennent les informations requises.

- Nous supposons que les montants de pertes rapportés dans la base à partir des médias sont précis et factoriels. L'évaluation des pertes est ainsi non basée sur des rumeurs ou des prédictions.

- Nous supposons que tous les types de pertes ont la même probabilité d’être rapportés dans la base; il n’y a donc pas d’effet médiatique relié à certains types de risque.
- Nous supposons que nous disposons de toutes les pertes de plus de 1 million de dollars dans la base externe pour les institutions financières existantes dans la base.
- Nous supposons qu’il n’y a pas de corrélation entre le montant de la perte et la probabilité que cette dernière soit rapportée. De plus, les distributions de la sévérité et des fréquences sont donc supposées être indépendantes.

1.5. Modèle de mise à l’échelle des montants de pertes externes

1.5.1. Modèle théorique de mise à l’échelle

Le mécanisme de mise à l’échelle repose sur trois hypothèses fondamentales; la première est la décomposition de la perte monétaire en deux composantes communes et idiosyncratiques ou spécifiques; la seconde stipule une relation non linéaire entre la partie idiosyncratique et les différents facteurs qui la composent; enfin, la dernière hypothèse énonce qu’outre les facteurs que nous contrôlons pour la mise à l’échelle, les autres facteurs non observables, telle que la qualité de l’environnement de contrôle, sont supposés être les mêmes entre les banques.

Concernant la première hypothèse, nous pouvons supposer que la perte opérationnelle peut se décomposer en deux composantes (Na, 2004): une composante commune à toutes les banques et une composante idiosyncratique spécifique à chaque montant de perte.

La composante commune contient tous les facteurs indépendants des activités de la banque et peut avoir le même impact sur toutes les banques; c’est donc une composante constante pour tous les événements de perte. Elle fait référence à l’environnement macroéconomique, géopolitique, culturel ou à la nature humaine en général (Na, 2004 et Na et al., 2006). Nous allons supposer cette composante

constante pour toutes les banques comme c'est fait dans l'étude de Na, (2004). Toutefois, il est toujours possible de considérer cette commonalité comme une fonction de variables d'état identiques pour toutes les banques, mais qui varient dans le temps.

La composante idiosyncratique, quant à elle, fait référence au risque spécifique de l'institution financière ou à la ligne d'affaires. Certains éléments de cette composante sont observables. Ils pourraient ainsi être quantifiables ou mesurables, comme la taille de la banque, le type de risque, la ligne d'affaires ou le lieu de l'événement de perte. Par contre, il existe des éléments non observables relatifs à l'environnement de contrôle, qui sont donc difficiles à quantifier. Ces éléments ne font pas l'objet d'étude dans ce travail.

Nous pouvons ainsi identifier un montant de perte comme une fonction de ces deux composantes :

$$\text{Perte}_i = f((\text{Comp}_{\text{commune}}), (\text{Comp}_{\text{idiosyncratique}})_i). \quad (1)$$

La deuxième hypothèse stipule que la fonction f est non linéaire. En effet, Na (2004) suppose que la fonction f est le produit d'une fonction de la composante commune et d'une fonction de la composante idiosyncratique. Or, comme la composante commune est constante, nous pouvons la modéliser par un paramètre.

$$\text{Perte}_i = \text{Comp}_{\text{commune}} \times g(\text{Comp}_{\text{idiosyncratique}})_i. \quad (2)$$

En ce qui concerne la fonction g , nous nous inspirons de l'étude de Shih, Samad-Khan et Medapa (2000), qui suppose une relation de puissance entre le montant de perte et la taille de la firme. Nous ne nous limitons cependant pas à la taille, estimée par l'actif total, comme facteur déterminant de la sévérité des pertes; nous ajoutons d'autres facteurs exprimés dans la fonction h qui suit :

$$g(\text{Comp}_{\text{idiosyncratique}}) = \text{Actifs}^a \times h(\text{facteurs}).$$

Nous pouvons donc réécrire (2) comme suit :

$$\text{Perte}_i = \text{Comp}_{\text{commune}} \times (\text{Actifs}_i^a \times h(\text{facteurs}_i)).$$

Pour simplifier l'étude, nous supposons que :

$$h(\text{facteurs}_i) = \exp\left(\sum_j b_j \times \text{facteur}_{ij}\right).$$

Ainsi, nous obtenons la relation linéaire suivante :

$$\text{Log}(\text{Perte}_i) = \text{Log}(\text{Comp}_{\text{commune}}) + a \times \text{Log}(\text{actifs}_i) + \left(\sum_j b_j \times \text{facteur}_{ij}\right). \quad (3)$$

Afin d'expliquer la variabilité des pertes et de construire le modèle de mise à l'échelle, il faut identifier les différents éléments de la composante idiosyncratique, à savoir les facteurs qui entrent en jeu pour expliquer la sévérité des pertes.

1.5.2. Description des variables

La variable dépendante est le logarithme des pertes opérationnelles. Les statistiques descriptives du tableau 1-1a montrent que la moyenne par événement de perte est évaluée à 67 millions de dollars, avec un écart-type de 521 millions de dollars. Le maximum des pertes est de 16 milliards de dollars. Les valeurs du premier quantile (2,73 M \$), de la médiane (8,10 M \$) ainsi que le troisième quantile (33,8 M \$) montrent que la distribution des pertes est très asymétrique. Les montants de pertes varient ainsi beaucoup allant des pertes moyennement importantes aux pertes catastrophiques.

Quant aux variables explicatives qui vont être incluses dans le modèle pour expliquer la variation du logarithme des pertes, elles sont décrites dans ce qui suit⁴. Le tableau 1-1 présente des statistiques descriptives des pertes en fonction de ces variables.

⁴ Nous avons testé également des variables binaires annuelles. Nous n'avons pas présenté les résultats de l'estimation de ces dix variables puisque aucune de ces variables n'a un impact significativement différent de zéro.

- La taille : La base dispose des variables caractérisant la taille de la compagnie. Selon les résultats de l'étude de Shih, Samad-Khan et Medapa (2000), la taille est faiblement reliée au montant de la perte. D'autres variables doivent expliquer cette variabilité.

Plusieurs informations sur la taille sont disponibles telles que le total des revenus, le total des actifs, le total des dépôts, le nombre d'employés et le total des fonds propres. Cependant, toutes ces variables sont corrélées entre elles. Le choix d'une d'entre elles est suffisant pour capter l'effet de taille. Nous avons choisi le total des actifs comme estimateur de la taille (la variable la plus corrélée avec les pertes). Les institutions financières ayant eu des pertes reportées dans la base de données utilisée sont de tailles très variables, allant de la plus petite banque, avec un total des actifs de 43 millions de dollars, à la plus grande institution, ayant un actif de 1 533 036 millions de dollars. Le total moyen des actifs est évalué à 270 681 millions de dollars.

Dans le tableau 1-1b, nous avons calculé le nombre d'événements, la moyenne et l'écart type des pertes en fonction de la taille. Nous avons donc classé les banques en trois grandeurs : celles ayant des actifs de moins de 400 milliards (petite et moyenne tailles), celles ayant des actifs entre 400 et 800 milliards (grande taille) et les très grandes banques, dont les actifs excèdent 800 milliards (très grande taille). Les résultats du tableau montrent que la moyenne des pertes est beaucoup plus importante dans les très grandes banques par rapport aux autres. Par contre, les institutions financières de grande taille ont une moyenne de pertes moins importante que les petites et moyennes banques.

Nous nous attendons que plus une institution financière soit grande, plus les pertes subies sont importantes. En effet, une banque ayant un total d'actifs plus élevé qu'une autre pourrait subir des pertes de type dommages aux actifs plus importantes suite à une catastrophe naturelle,

par exemple. La taille pourrait donc avoir un impact positif sur la sévérité des pertes.

- Le lieu : Comme les pertes n'ont pas toutes eu lieu dans un même pays, il faut ainsi incorporer une variable qui capte l'effet du lieu. En effet, nous nous attendons à ce que cette variable soit significativement liée aux montants des pertes, vu la différence d'environnement, de législation, etc. Il est à remarquer que 60 % des pertes ont eu lieu aux États-Unis, comparativement à seulement 4 % au Canada. Cette variation peut être expliquée par le nombre de banques, beaucoup plus élevé aux États-Unis qu'au Canada. L'autre proportion des pertes a été partagée entre les pays de l'Europe et le reste du monde.

Le tableau 1-1c présente des statistiques des pertes en fonction du lieu. Nous remarquons que la moyenne des pertes opérationnelles diffère selon le lieu de l'occurrence de la perte, montrant ainsi des environnements différents. Il est à mentionner que la moyenne des pertes aux États-Unis est plus élevée de celle au Canada (38 M\$ versus 9 M\$). De plus l'environnement autre (pays différents du Canada, États-Unis et de ceux situés en Europe) présente la moyenne des pertes la plus élevée (163 M\$). Il constituera donc l'environnement le plus risqué.

Dans l'objectif de cerner le lien entre la taille de l'institution financière et le lieu, nous présentons dans le tableau 1-1d des statistiques sur la taille des institutions en fonction du lieu de l'occurrence des pertes. Nous remarquons que le total des actifs moyen des institutions au Canada est moins élevé par rapport aux États-Unis et l'Europe. Toutefois, bien que l'environnement représenté par les autres pays soit plus risqué (moyenne des pertes la plus élevée), les institutions ayant subi ces pertes sont en moyenne plus petites que celles localisées dans les autres environnements. Ainsi, il n'y a pas de lien direct entre la taille de l'institution financière et le lieu de l'occurrence des pertes.

Des variables binaires (États-Unis, Canada, Europe, Autres) captent l'effet de la localisation des pertes dans l'un des pays ou continent cité précédemment.

- La ligne d'affaires : Nous nous attendons à ce que la nature de la ligne d'affaires ait un impact sur la sévérité des pertes. Certaines unités enregistrent, en moyenne, des pertes plus élevées que d'autres. La mise en relief de la ligne d'affaires, où la perte a eu lieu, peut expliquer la sévérité des pertes extrêmes. D'après les statistiques présentées dans le 1-1e, nous remarquons que deux lignes d'affaires, services bancaires commerciaux (25 %) et services bancaires détail (33 %), sur sept contiennent 58 % des pertes. De plus, la moyenne des pertes est beaucoup plus élevée dans la ligne d'affaires services bancaires commerciaux par rapport aux autres unités.

À partir des informations collectées du LDCE et QIS-4⁵ de 27 institutions financières, la ligne d'affaires services bancaires détail contient 44 % des pertes opérationnelles des 177 données de plus de un million de dollars collectées sur la période 2001-2004, alors que la ligne d'affaires services bancaires commerciaux n'en contient que 9 %. Cet écart peut être expliqué par la période de collecte (l'étude de LDCE couvre 4 années de pertes, alors que la base externe couvre 11 années d'historique) ou par le nombre différent d'institutions financières où les pertes ont eu lieu.

Ainsi, des variables dichotomiques pour chaque ligne d'affaires capteront l'effet de la nature de l'activité de chacune dans la détermination des montants de pertes.

- Les types de risque : Certains types de risque sont à faible fréquence et à haute sévérité, alors que d'autres sont à haute fréquence et à sévérité

⁵ *Loss Data Collection Exercise (LDCE) et Quantitative Impact Study 4 (QIS-4)* : deux études menées par la Banque Fédérale Américaine et des agences de réglementation économique dans le but d'évaluer l'impact de Bâle II sur le capital minimum réglementaire requis.

relativement faible. Le tableau 1-1f montre que 44 % des pertes sont de type *clients, produits et pratiques commerciales* et que plus de 40 % des pertes sont réparties entre les *fraudes internes* et *fraudes externes*. Toutefois, moins de 0,5 % des pertes sont de type *actifs tangibles, corporels et endommagés* et 0,5 % des pertes sont de type *perturbation des affaires et défaillance des systèmes*. La moyenne des pertes est la plus élevée pour le type de risque *actifs tangibles, corporels et endommagés* (115 M\$) alors qu'elle est la plus faible pour *perturbation des affaires et défaillance des systèmes* (5 M\$).

Les résultats des études LDCE et QIS-4 montrent une très grande différence par rapport à cette répartition. En effet, 49 % des pertes sont de type gestion d'exécution, 31 % sont des pertes de type clients, produits et pratiques commerciales, 7 % des fraudes externes, et 3 % sont des fraudes internes. Quant aux types de risque dommages aux actifs et perturbation des affaires et défaillance des systèmes, la proportion des pertes est aussi faible que celle dans la base externe.

L'introduction de variables dichotomiques captant cet aspect des types de risque, peut être pertinente dans l'explication de la variabilité des montants de pertes. Ainsi, 7 variables capteront l'effet type de risque dans notre modèle.

Il était préférable de considérer le modèle de mise à l'échelle par ligne d'affaires et/ou par type de risque au lieu de prendre des variables binaires. Cependant, pour certains types de risque et/ou ligne d'affaires, les données sont peu nombreuses ne nous permettant pas de faire la régression.

1.5.3. Régression linéaire

Pour expliquer le degré de variabilité des montants de pertes externes, nous allons estimer les coefficients de la régression suivante, qui nous permettront d'évaluer les

composantes communes et spécifiques pour chaque montant de perte. La régression suivante découle de l'équation (3).

$$Y_i = a_0 + a_1 \text{Taille}_i + a_2 \text{ÉU}_i + a_3 \text{Canada}_i + a_4 \text{Europe}_i + \sum_{j=5}^{11} a_j \text{UA}_{ij} + \sum_{j=12}^{17} a_j \text{TR}_{ij} + e_i \quad (4)$$

avec :

Y_i : Log(pertes)_i;

a_0 : Log(composante commune);

Taille_i : Log(actifs)_i;

ÉU_i : variable binaire qui prend la valeur 1 si la perte a eu lieu aux États-Unis, 0 sinon;

Canada_i : variable binaire qui prend la valeur 1 si la perte a eu lieu au Canada, 0 sinon;

Europe_i : variable binaire qui prend la valeur 1 si la perte a eu lieu en Europe, 0 sinon;

C'est la variable *Autres* qui est la catégorie de référence;

UA_{ij} : variable binaire qui prend la valeur 1 si la perte a eu lieu dans l'unité d'affaires j , 0 sinon;

C'est la variable relative à la ligne d'affaires *Paiements et règlements* qui est la catégorie de référence;

TR_{ij} : variable binaire qui prend la valeur 1 si la perte est de type de risque j , 0 sinon;

C'est la variable relative au type de risque *Perturbation des affaires et défaillance des systèmes* qui est la catégorie de référence;

e_i : variable d'écart qui représente la composante spécifique non observable supposée suivre une loi normale $(0, \sigma^2)$.

1.5.4. Résultats de la régression

L'estimation des paramètres est faite par la méthode des moindres carrés ordinaires. Les résultats de l'estimation sont présentés dans le tableau 1-2. Le coefficient de détermination ajusté $R^2_{\text{ajusté}}$ est de l'ordre de 10,63 %. Bien que le coefficient soit faible, il est meilleur que celui de 5 % trouvé dans la littérature jusqu'à date (Shih et

al, 2000). Rappelons qu'il est difficile de capter certains facteurs non observables qui ne sont pas présents dans la base externe.

La variable taille, estimée par le logarithme du total des actifs, est significativement différente de 0. Le coefficient est positif, confirmant le fait que plus une firme est grande, plus le niveau des pertes est important.

Les coefficients des variables binaires *ÉU* et *Canada* sont significativement différentes de 0. Le signe négatif doit être interprété par rapport au coefficient *Autres* (catégorie de référence). En comparant les coefficients *ÉU* et *Canada*, nous remarquons que les États-Unis constituent un environnement plus risqué que le Canada. De plus, pour s'assurer que cette différence est statistiquement significative, nous avons refait la régression mais en prenant la variable *ÉU* comme catégorie de référence. Les résultats montrent effectivement que le coefficient de la variable binaire *Canada* est négatif et statistiquement différent de 0. Ce constat confirme notre affirmation à savoir que le Canada constitue un environnement moins risqué que les États-Unis.

La variable *Services bancaires commerciaux* est la seule variable liée à la nature de la ligne d'affaires ayant un impact significativement non nul à un degré de confiance de 99 %. Le coefficient, interprété par rapport à celui de la variable *paiements et règlements* (variable omise), est positif, montrant que les pertes de ce type sont plus importantes en montants que pour les autres lignes d'affaires.

Enfin, le type de risque *Clients, produits et pratiques commerciales* a un pouvoir explicatif statistiquement significatif à un seuil de 10%. Ceci montre que les pertes sont plus importantes pour ce type de risque comparativement aux autres.

1.5.5. Tests de robustesse sur la variable taille

Nous commençons par une régression simple incluant seulement la variable taille, modèle semblable à celui de Shih et al (2000). Nous ajoutons chaque fois une catégorie de variable à ce même modèle pour capter l'effet additionnel éventuel de chaque catégorie et pour tester la stabilité des paramètres. Les résultats montrent que

la taille n'explique qu'une très faible partie du niveau des pertes. En effet, le modèle 1 du tableau 1-3 montre un $R^2_{\text{ajusté}}$ de 0,6 %. Cette statistique est nettement améliorée (4,32 %) suite à l'introduction des catégories de la variable de lieu. Il est à mentionner que les valeurs des coefficients estimés restent stables et demeurent significativement différentes de 0 et ce, par rapport au modèle de base. Le modèle 3 ajoute des variables estimant la nature des lignes d'affaires où la perte a eu lieu. Le coefficient de détermination ajusté passe de 4,32 % (modèle 2) à 7,16 % (modèle 3). La variable *services bancaires commerciaux* reste toujours significativement différente de 0. Ainsi, chaque catégorie de variable a un pouvoir statistiquement significatif pour expliquer la sévérité des pertes opérationnelles. Les coefficients des variables sont relativement stables.

Dans un deuxième temps, nous retenons seulement les variables statistiquement non nulles à un degré de confiance de 90 % et plus. Ensuite, nous faisons la régression du log des pertes sur les 5 variables retenues. Le modèle ainsi construit permet de tester si les variables significatives dans le modèle de base conservent leur pouvoir explicatif lorsqu'elles sont testées seules. Les résultats présentés dans le modèle 4 du tableau 1-3 montrent que toutes les variables restent significativement non nulles à un degré de confiance de 90 %. Le coefficient de détermination ajusté est de l'ordre de 9,38 %. Les signes et l'ampleur des variables ne changent pas. Ces 5 variables statistiquement significatives seront retenues pour développer la formule de normalisation.

1.5.6. Formule de normalisation

Nous nous intéressons à une banque A donnée et nous voulons trouver la valeur équivalente d'une perte ayant lieu dans une autre institution financière. Une formule de normalisation permet ainsi de mettre une perte survenue dans une banque B à l'échelle d'une banque A .

Selon l'équation 2, une perte i est le produit d'une composante commune et d'une fonction de la composante spécifique. L'analyse par régression faite précédemment nous a permis d'identifier ces deux composantes.

$$\text{Log}(\text{perte}_i) = \underbrace{\hat{a}_0}_{\log(\text{Comp. Comm})} + \underbrace{\hat{a}_1 \text{Taille}_i + \hat{a}_2 \text{ÉU}_i + \hat{a}_3 \text{Canada}_i + \hat{a}_4 \text{SBC}_i + \hat{a}_5 \text{CPPC}_i}_{\log g(\text{Comp Idiosyncratique})}$$

où :

SBC : fait référence à la ligne d'affaires *Services bancaires commerciaux*

CPPC : fait référence au type de risque *Clients, produits et pratiques commerciales*

Comme la composante commune est constante pour tous les montants de pertes, il est possible de réécrire l'équation 2 comme suit :

$$\text{Comp}_{\text{comm}} = \frac{\text{Perte}_A}{g(\text{Comp}_{\text{idio}})_A} = \frac{\text{Perte}_B}{g(\text{Comp}_{\text{idio}})_B} = \dots = \frac{\text{Perte}_N}{g(\text{Comp}_{\text{idio}})_N}. \quad (5)$$

Supposons que nous avons une perte survenue dans une banque *B* et que nous voulons connaître sa valeur équivalente si elle survenait à une banque *A*. De l'analyse précédente, il nous est possible de déterminer la composante idiosyncratique de la perte *B* ainsi que celle de *A*. En effet, nous multiplions les coefficients déjà estimés par la valeur correspondante des différentes variables pour trouver la composante idiosyncratique ou spécifique.

$$\text{Perte}_A = \frac{g(\text{Comp}_{\text{idio}})_A}{g(\text{Comp}_{\text{idio}})_B} \times \text{Perte}_B \quad (6)$$

avec

$$g(\text{Comp}_{\text{idio}})_A = \exp(\hat{a}_1 \text{Taille}_A + \hat{a}_2 \text{ÉU}_A + \hat{a}_3 \text{Canada}_A + \hat{a}_4 \text{SBC}_A + \hat{a}_5 \text{CPPC}_A).$$

L'équation (6) suppose qu'outre les variables retenues pour faire la mise à l'échelle, la partie non expliquée par le modèle de régression qui peut être attribuable aux facteurs qualitatifs non observables tels que la qualité de gestion, l'environnement de contrôle, etc., est supposée être la même entre la banque *A* et la banque *B* (troisième hypothèse).

Ainsi, pour calculer le montant équivalent d'une perte survenue dans l'industrie bancaire pour une banque donnée, il faut tout d'abord calculer les composantes idiosyncratiques des deux pertes avec l'équation précédente. Ensuite, nous

appliquons la formule (6) pour trouver la perte équivalente pour la banque A. En appliquant cette même méthode à toute la base externe, nous aurons une base de pertes extrêmes ayant eu lieu dans d'autres institutions bancaires mais mises à l'échelle d'une banque donnée. Ainsi, la sévérité des pertes a été ajustée en tenant compte de plusieurs facteurs tels que la taille, le lieu, la ligne d'affaires et le type de risque.

1.5.7. Validation du modèle de mise à l'échelle

Afin de concrétiser le modèle de mise à l'échelle, nous choisissons la banque américaine *Merrill Lynch*⁶ de la base externe dont le nombre d'événements de pertes sur la période 1994-2004 est de 52. Nous allons tout d'abord déterminer les pertes opérationnelles de la base externe mises à l'échelle de cette banque. Nous allons ensuite comparer les statistiques des pertes réellement observées à celles trouvées après la mise à l'échelle.

Dans une première étape, nous allons déterminer l'équivalent des 1 056 événements de pertes de la base externe pour la banque Merrill Lynch. Nous calculerons donc le montant de perte qui pourrait survenir à Merrill Lynch découlant d'un même type de risque, survenant à une même ligne d'affaire et à la même année d'occurrence que celle de la base externe. Par contre, nous prendrons le total des actifs de la banque Merrill Lynch relatif à l'année de l'événement pour toutes les pertes. De plus, nous prendrons les États-Unis comme seul lieu d'occurrence de toutes les pertes externes vu que toutes les pertes observées de Merrill Lynch ont eu lieu aux États-Unis. Une fois que nous aurons identifié les variables explicatives de la régression (4), nous calculerons la composante idiosyncratique équivalente à chaque perte dans la base externe, tel que montré dans l'annexe 1, et ce, avec les coefficients des variables estimés précédemment. La formule de normalisation présentée à la section précédente nous permet de déterminer la perte mise à l'échelle de la banque en question.

⁶ Nous avons choisi cette banque parce que c'est celle qui a eu le maximum de pertes de plus de 1million de dollars sur la période 1994-2004 dans la base externe.

Ensuite, nous comparons des statistiques calculées, d'une part, sur l'échantillon de 52 pertes réellement observées à la Merrill Lynch, et sur les 1 050⁷ pertes qui pourraient avoir lieu à la même banque, équivalentes à celles de la base externe, d'autre part. Ces statistiques sont présentées au tableau 1-4. Elles montrent que les moyennes des deux échantillons sont sensiblement proches. En effet, un test d'hypothèse confirme ce constat et montre que les deux moyennes ne sont pas statistiquement différentes à un seuil de 95 %. Il est également à remarquer que les écarts type des deux échantillons sont proches (83,1 versus 84,3).

Dans une deuxième étape de l'analyse, nous regardons l'impact des variables prises en considération dans la mise à l'échelle sur les montants de pertes après normalisation. Le tableau 1-5 présente un exemple d'événement de perte extrait de la base externe ayant lieu à la banque de New York, ainsi que les montants de pertes mis à l'échelle de banques fictives. Nous modifions tout d'abord une seule caractéristique de l'événement à la fois, pour voir son impact monétaire sur la perte. Nous analysons ensuite l'effet agrégé de plusieurs variables sur le montant de perte.

Nous remarquons que si l'événement a eu lieu dans une banque plus grande, tout autre facteur étant égal par ailleurs, le montant de la perte serait légèrement plus élevé (de 8,26 millions de dollars à 9,27 millions, alors que le total des actifs a plus que quadruplé). Toutefois, si le même événement avait lieu au Canada plutôt qu'aux États-Unis, la perte aurait une ampleur moindre (elle passerait de 8,26 millions à 4,97 millions). Comme nous l'avons déjà montré, le Canada présente donc un environnement moins risqué que les États-Unis.

Nous remarquons également que la ligne d'affaires services bancaires commerciaux où l'événement de perte a eu lieu, présente plus de risque d'occasionner des pertes importantes que les autres lignes d'affaires. La perte passera d'ailleurs de 8,26 millions à 16,06 millions si elle a lieu dans la ligne d'affaires services bancaires commerciaux au lieu de la ligne d'affaires services bancaires détail. D'autre part, le

⁷ Nous avons exclu 6 pertes de l'analyse parce qu'elles présentent des valeurs trop extrêmes 'outliers'.

type de risque a un impact important sur l'ampleur des pertes. En effet, si la perte est de type clients, produits et pratiques commerciales, le montant passera à 15,56 millions de dollars, vu que ce type de risque a un impact significatif sur la sévérité des pertes, tel que conclu précédemment à la section 4.4.

Nous présentons enfin, dans les trois dernières lignes du tableau 1-5, l'impact agrégé de deux ou plusieurs variables. Nous faisons la mise à l'échelle en modifiant la taille de la banque et le lieu de l'événement. Il est à remarquer que même si la taille de la banque est plus importante, l'impact du lieu (Canada) l'emporte et la mise à l'échelle donne un montant moins élevé que l'événement d'origine. En modifiant aussi la ligne d'affaires et le type de risque, nous trouvons que la perte a plus que doublé. Il est à remarquer que la perte qui en résulte est très différente de celle trouvée suite à la seule modification de la taille. Ainsi, cette analyse nous montre que l'effet de la taille est peu important comparativement aux autres facteurs de la mise à échelle. Notre modèle vient donc améliorer les modèles existants dans la littérature qui se basent uniquement sur la taille.

1.6. Modèle de mise à l'échelle des fréquences des pertes externes

Rappelons que notre objectif est de corriger le biais d'échelle afin de combiner les données internes et externes dans le but de mesurer le capital opérationnel. Nous avons déterminé, à la section précédente, la mise à l'échelle des montants de pertes. Il est ainsi possible de déterminer plusieurs pertes extrêmes susceptibles d'arriver à notre banque de référence A. La question qui reste à poser est : À quelles fréquences une banque subira-t-elle ces pertes ?

La mise à l'échelle des fréquences est une notion peu abordée dans la littérature. En effet, certains chercheurs ont développé des modèles pour la mise à l'échelle des sévérités, mais le nombre de pertes externes à combiner avec les données internes n'est pas encore modélisé. Nous proposons dans ce qui suit un modèle permettant d'ajuster le nombre de pertes externes par banque et de le ramener à l'échelle d'une banque A.

1.6.1. Description du modèle

Le modèle développé dans cette section permet, d'une part, la mise à l'échelle du nombre des pertes externes et, d'autre part, la détermination de la distribution théorique s'ajustant aux fréquences. Nous nous attendons à ce que le nombre de pertes par institution financière sur un horizon déterminé dépende de certains facteurs décrivant les caractéristiques de l'institution financière. La taille de l'institution peut effectivement jouer un rôle important dans la détermination des fréquences des pertes. Il est à remarquer que plus une banque est grande, plus elle sera exposée à des risques opérationnels. En effet, si une banque fait plus de transactions, a plus d'actifs, d'employés et de revenus qu'une autre banque, elle aura probablement plus de pertes opérationnelles de différents types (fraudes, dommages aux actifs, etc.) De plus, la répartition géographique des activités de l'institution peut nous donner une idée sur le degré d'efficacité des contrôles. En effet, plus les activités sont réparties sur plusieurs pays, plus les mesures de contrôles sont dispersées et moins efficaces.

Il est donc possible d'expliquer le nombre de pertes par une régression sur les différentes variables citées ci-dessus. Cependant, comme il s'agit de fréquences, donc de variables discrètes, il est plus opportun de modéliser ces nombres par des distributions de comptage telles que la Poisson et la binomiale négative. Ainsi, l'application de modèles de comptage avec composante de régression est appropriée dans ce contexte. Ces modèles ont l'avantage de trouver la distribution paramétrique s'ajustant aux données de fréquences, mais également d'avoir des paramètres flexibles propres à chaque observation. En d'autres termes, les paramètres de la distribution dépendent des variables identifiant les caractéristiques de l'institution financière où la perte a eu lieu. Une fois l'estimation des paramètres faite, il est possible de calculer ceux propres à une banque donnée.

D'autre part, comme nous disposons seulement d'institutions qui ont subi des pertes, les fréquences sont donc non nulles. Il y a alors lieu de corriger ce biais, en prenant des distributions tronquées au point zéro. Dans ce qui suit, nous allons décrire les variables qui seront incluses dans le modèle, pour ensuite présenter et tester chacun

des deux modèles, à savoir le modèle Poisson tronqué avec composante de régression et le modèle binomial négatif tronqué avec composante de régression.

1.6.2. Description des variables

Nous créons une variable décrivant le nombre de pertes de plus d'un million de dollars par institution financière sur la période de 1994 jusqu'à 2004⁸. Nous avons donc un échantillon de 323 institutions financières ayant subi des pertes de plus de un million de dollars rapportées dans la base externe. Nous expliquerons la fréquence par la taille de la banque, ainsi que le degré de répartition géographique de ses activités. En effet, nous nous attendons à ce que plus la taille de la banque est grande, plus le nombre de ses pertes augmente. D'autre part, plus les activités d'une institution financière sont réparties géographiquement, plus les coûts de contrôle sont élevés, engendrant des pertes d'efficacité et plus le nombre de pertes est élevé.

La taille sera donc estimée par le logarithme du total moyen des actifs de l'entreprise sur la période 1994-2004. La répartition géographique sera estimée par des variables binaires telles que :

- *ÉU* : variable binaire prenant la valeur 1 si l'institution a eu des pertes aux États-Unis sur la période 1994-2004, 0 sinon;
- *Canada* : variable binaire prenant la valeur 1 si l'institution a eu des pertes au Canada sur la période 1994-2004, 0 sinon;
- *Europe* : variable binaire prenant la valeur 1 si l'institution a eu des pertes en Europe sur la période 1994-2004, 0 sinon;
- *Autres* : variable binaire prenant la valeur 1 si l'institution a eu des pertes dans un autre pays ou continent sur la période 1994-2004, 0 sinon.

Contrairement au modèle de sévérité, ces variables ne sont pas mutuellement exclusives.

Le tableau 1-6a présente des statistiques descriptives sur le nombre de pertes par banque, ainsi que le total des actifs par institution financière sur la période 1994-2004. Le nombre moyen des pertes de plus d'un million est de 3,3 événements par

⁸ Nous retenons cette période durant laquelle la collecte des pertes a été plus exhaustive.

institution sur un horizon de 11 années, avec un maximum de 52. La taille des institutions financières est très variable, le total moyen des actifs étant 123 174 millions de dollars. Quant à la répartition géographique des activités des banques, nous trouvons que les pertes sont plus concentrées aux États-Unis et dans les pays autres que le Canada, États-Unis et ceux de l'Europe.

Le tableau 1-6b montre que les banques de petite et moyenne tailles (la moyenne des actifs est inférieure à 400 000 M\$) ont subi en moyenne moins de pertes sur la période 1994-2004 par rapport aux grandes banques (la moyenne des actifs est entre 400 000 M\$ et 800 000 M\$). De plus, les très grandes banques, dont la moyenne des actifs excède les 800 000 M\$, ont eu un nombre moyen de pertes (19) plus élevé que les autres banques. Ces statistiques descriptives montrent donc un lien entre la taille de l'institution financière et le nombre de pertes de plus de 1 million de dollars.

Le tableau 1-6c présente des statistiques descriptives sur le nombre de pertes par banque selon la répartition géographique des activités de l'institution en question. Les résultats montrent que si les activités sont concentrées dans un même pays, la moyenne du nombre de pertes est de 2 par banque; par contre, elle varie entre 8 et 10 lorsque les activités sont réparties sur deux ou trois pays. Le nombre moyen des pertes par institution bancaire atteint 28 lorsque les activités sont très dispersées géographiquement. Il est donc intéressant de tenir compte de la répartition géographique des activités de la banque dans la modélisation des fréquences.

1.6.3. Modèle Poisson tronqué avec composante de régression

Si Y_i , le nombre de pertes par compagnie i sur la période 1994-2004, suit une distribution Poisson, alors la probabilité d'avoir y pertes sera :

$$P(Y_i = y) = \frac{e^{-\lambda} \lambda^y}{y!} \quad y = 0, 1, 2, \dots \text{ et } \lambda > 0$$

où λ est le paramètre de la Poisson. La caractéristique principale de cette distribution est $E(Y_i) = \text{Var}(Y_i) = \lambda$.

Toutefois, comme nous sommes en présence de données tronquées au point zéro, il faut donc estimer la probabilité du nombre de pertes conditionnelle au fait que les fréquences observées sont strictement supérieures à zéro. La probabilité conditionnelle est :

$$P(Y_i = y | Y_i > 0) = \frac{e^{-\lambda} \lambda^y}{y!(1 - e^{-\lambda})} \quad y = 1, 2, \dots \text{ et } \lambda > 0.$$

D'autre part, nous pouvons permettre au paramètre λ de varier d'une observation à une autre. Soit $\lambda_i = \exp(X_i \beta)$, où X_i est un vecteur de $(1 \times m)$ variables exogènes (caractéristiques des entreprises où la perte a eu lieu) et β un vecteur de $(m \times 1)$ coefficients. La fonction exponentielle permet d'assurer la non-négativité du paramètre λ_i .

Dans notre contexte, le paramètre λ_i est de la forme :

$$\lambda_i = \exp(\beta_0 + \beta_1 \log(\text{Actifs})_i + \beta_2 \text{ÉU}_i + \beta_3 \text{Canada}_i + \beta_4 \text{Europe}_i + \beta_5 \text{Autre}_i). \quad (7)$$

Ainsi, la probabilité qu'une institution financière i ait y pertes sur un horizon de 11 années, connaissant les caractéristiques qui lui sont propres, est :

$$\Pr(Y_i = y | Y_i > 0, X_i) = \frac{e^{-\exp(X_i \beta)} \exp(X_i \beta)^y}{y!(1 - e^{-\exp(X_i \beta)})}. \quad (8)$$

Il est à remarquer que la distribution Poisson tronquée ne présente pas d'équidispersion comme c'est le cas de la Poisson. En effet, la moyenne et la variance sont :

$$E(y_i | y_i > 0, X_i) = u_i = \lambda_i + \gamma_i$$

$$E(y_i | y_i > 0, X_i) = \frac{\lambda_i}{1 - e^{-\lambda_i}}$$

et

$$\begin{aligned}V(y_i | y_i > 0, X_i) &= \sigma_i^2 = \lambda_i - \gamma_i (u_i - 1) \\V(y_i | y_i > 0, X_i) &= u_i (1 - \gamma_i)\end{aligned}$$

avec

$$\gamma_i = \frac{\lambda_i}{e^{\lambda_i} - 1}$$

γ_i est appelé un facteur d'ajustement. Il est interprété comme étant la différence entre les moyennes des distributions tronquées et non tronquées. La moyenne de la distribution tronquée est en effet plus élevée que celle de la distribution non tronquée. Toutefois, la variance de la distribution tronquée est plus petite, comme le montrent les expressions précédentes.

Le vecteur des paramètres β peut être estimé par la méthode du maximum de vraisemblance (voir Maddala, 1983 et Cameron et Trivedi, 1986, pour une discussion détaillée sur les méthodes d'estimation des modèles Poisson avec composante de régression). Il s'agit donc d'estimer les six coefficients de l'équation (7) et de calculer le paramètre λ_i pour chaque unité d'exposition i (entreprise dans notre cas). Il est ainsi possible de calculer le paramètre λ propre à une banque donnée. Nous retenons donc les coefficients estimés et nous prenons les variables de taille et de répartition géographique de la banque en question.

1.6.4. Modèle binomial négatif tronqué avec composante de régression

Dans le modèle précédent, l'espérance conditionnelle de y étant donné les variables exogènes X_i et la variance conditionnelle correspondante ne peuvent pas varier indépendamment. Pour palier à cette restriction, nous introduisons un terme d'erreur à la définition du paramètre de la Poisson qui s'écrira :

$$\lambda_i^* = \exp(\beta_0 + \beta_1 \log(\text{Actifs})_i + \beta_2 \text{ÉU}_i + \beta_3 \text{Canada}_i + \beta_4 \text{Europe}_i + \beta_5 \text{Autre}_i + \varepsilon_i). \quad (9)$$

Le terme ε_i constitue l'erreur de spécification due à des variables explicatives omises non observées et indépendantes des variables exogènes X_i du modèle. Contrairement au modèle précédent, λ_i^* est une variable aléatoire. La distribution conditionnelle de Y_i devient :

$$\Pr(Y_i = y | Y_i > 0, X_i) = \int_{-\infty}^{+\infty} \frac{e^{-\exp(X_i\beta + \varepsilon_i)} \exp(X_i\beta + \varepsilon_i)^{y_i}}{y_i! (1 - e^{-\exp(X_i\beta + \varepsilon_i)})} g(\varepsilon_i) d\varepsilon_i \quad y = 1, 2, \dots$$

avec $g(\varepsilon_i)$ la fonction de densité de ε_i .

Dans un objectif de commodité, nous supposons, dans ce qui suit, que $\mu_i = \exp(\varepsilon_i)$ suit une distribution gamma de paramètre α telle que;

$$f(\mu_i) = \frac{\mu_i^{\frac{1}{\alpha}-1} \exp\left(-\frac{\mu_i}{\alpha}\right)}{\alpha^{\frac{1}{\alpha}} \Gamma\left(\frac{1}{\alpha}\right)} \quad \alpha > 0$$

avec $E(\mu_i) = 1$ et $V(\mu_i) = \alpha$.

La distribution conditionnelle de Y_i , ainsi obtenue, est donc une binomiale négative tronquée dont la densité est :

$$\Pr(Y_i = y | Y_i > 0, X_i) = \frac{\Gamma\left(y + \frac{1}{\alpha}\right)}{\Gamma(y+1)\Gamma\left(\frac{1}{\alpha}\right)} \times \frac{1}{(1 + \alpha \exp(X_i\beta))^{\frac{1}{\alpha}} - 1} \times \left(\frac{\alpha \exp(X_i\beta)}{1 + \alpha \exp(X_i\beta)}\right)^y \quad (10)$$

La moyenne et la variance de la distribution binomiale négative tronquée avec composante de régression sont les suivantes :

$$E(y_i | y_i > 0, X_i) = u_i^* = \lambda_i + \gamma_i^*$$

et

$$V(y_i | y_i > 0, X_i) = \sigma_i^{*2} = \lambda_i + \alpha \lambda_i - \gamma_i^* (u_i^* - 1)$$

avec

$$\gamma_i^* = \left[\frac{\lambda_i}{(1 + \alpha)^{\alpha^{-1} \lambda_i} - 1} \right].$$

De même que la distribution Poisson, la moyenne de la distribution binomiale négative tronquée est plus élevée de celle de la distribution non tronquée. Alors que la troncature au point zéro a réduit la variance par le facteur d'ajustement γ_i^* . Bien que la distribution Poisson tronquée ne présente plus la caractéristique d'équidispersion ($E(y_i | y_i > 0 \text{ et } X_i) \neq V(y_i | y_i > 0 \text{ et } X_i)$), le modèle binomial négatif tronqué introduit la surdispersion dans le sens où sa variance est plus élevée de celle de la Poisson.

Sous l'hypothèse que la densité est correctement spécifiée et d'autres conditions usuelles, une estimation par maximum de vraisemblance des paramètres β et α peut être faite. Il s'agit donc d'estimer les six coefficients de l'équation (9) ainsi que le coefficient α . Il est ainsi possible de calculer le paramètre λ_i^* pour une banque donnée. Nous retenons donc les coefficients estimés et nous prenons les variables de taille et de répartition géographique de la banque en question. Cependant, si la densité est mal spécifiée, les estimateurs par maximum de vraisemblance du modèle binomial négatif avec composante de régression ne seront pas cohérents.

1.6.5. Présentation et comparaison des résultats

Les résultats de l'estimation des paramètres pour les deux modèles sont résumés dans le tableau 1-7. Il est à remarquer que les coefficients estimés ont les mêmes signes pour les deux modèles. Leurs signes confirment nos attentes. En effet, des coefficients de taille et de répartition géographique positifs montrent que le nombre de pertes augmente avec la taille de la banque et la répartition des activités dans des environnements différents.

Dans le modèle Poisson tronqué avec composante de régression, toutes les variables explicatives introduites dans la régression sont statistiquement significatives à des degrés de confiance situés entre 90 % et 99 %. Elles ont donc un impact explicatif sur le nombre de pertes. Quant au modèle binomial négatif tronqué avec composante de régression, les coefficients des variables *Europe* et *Autres* ne sont pas statistiquement différents de zéro.

Dans l'objectif de comparer les deux modèles et de tester l'hypothèse de l'existence de la surdispersion dans les données, nous implantons un test de surdispersion propre aux modèles tronqués avec composante de régression tel que développé par Gurmu (1991). Le test consiste alors à tester l'hypothèse nulle $H_0 : \alpha = 0$ contre l'hypothèse alternative $H_1 : \alpha > 0$.

Soient $\rho = \begin{pmatrix} \beta \\ \alpha \end{pmatrix}$ le vecteur des coefficients et $\hat{\rho} = \begin{pmatrix} \hat{\beta} \\ 0 \end{pmatrix}$ l'estimateur par maximum de vraisemblance de ρ sous l'hypothèse nulle. Le symbole $\hat{\cdot}$ indique dans ce qui suit que les paramètres sont évalués au point $\hat{\rho}$ sous l'hypothèse stipulant que $\alpha = 0$.

La statistique du 'score test' ou du multiplicateur de Lagrange qui vise à tester la surdispersion dans le modèle Poisson tronquée avec composante de régression est calculée comme suit :

$$\tau = \frac{\sum_{i=1}^N \hat{\lambda}_i^{-1} [\hat{v}_i^2 - y_i + (\hat{v}_i + y_i) \hat{\gamma}_i]}{2 [\mathfrak{V}_{\alpha\alpha}(\hat{\rho}) - \mathfrak{V}_{\alpha\beta}(\hat{\rho}) \mathfrak{V}_{\beta\beta}^{-1}(\hat{\rho}) \mathfrak{V}_{\beta\alpha}(\hat{\rho})]^{\frac{1}{2}}} \stackrel{a}{\sim} N(0,1)$$

où :

$$\hat{v}_i = y_i - \hat{u}_i;$$

^a
 \sim : indique que cette statistique est asymptotiquement distribuée comme la distribution normale centrée réduite.

$$\mathfrak{I}(\rho) = -E\left(\frac{\delta^2 L(\rho)}{\delta\rho\delta\rho'}\right)\Big|_{H_0} = \begin{bmatrix} \mathfrak{I}_{\beta\beta}(\rho) & \mathfrak{I}_{\beta\alpha}(\rho) \\ \mathfrak{I}_{\alpha\beta}(\rho) & \mathfrak{I}_{\alpha\alpha}(\rho) \end{bmatrix} \text{ est la matrice de l'information}^9.$$

sous l'hypothèse $H_0 : \alpha = 0$.

La valeur de la statistique τ basée sur la distribution binomiale négative tronquée au point zéro est 34,637. Le test permet donc de rejeter l'hypothèse nulle. Ainsi, le résultat montre l'existence d'une surdispersion significative dans les données qu'il est plus adéquat de la modéliser avec la distribution binomiale négative.

Pour trouver les paramètres de la distribution retenue relatifs à une banque donnée, il suffit de calculer le paramètre λ_i^* selon l'équation (9) en prenant les coefficients déjà estimés et la valeur des variables relatives à cette banque sur la période en question. Il est ainsi possible de générer un nombre de pertes de plus de 1 million de dollars qu'une banque pourrait subir à partir de sa propre distribution de fréquences.

Afin de concrétiser la mise à l'échelle des fréquences, nous prenons un exemple d'une banque dont le total actif est de 100 000 M\$ et dont les activités sont réparties entre le Canada et les États-Unis sur la période 1994-2004. Nous calculons le paramètre de la distribution Poisson selon l'équation (7), en prenant les coefficients déjà estimés et les valeurs appropriées des variables. Le tableau 1-8a présente le résultat ainsi que les détails du calcul du paramètre. Ainsi, si nous choisissons le modèle Poisson, le nombre de pertes de plus de 1 million de dollars pour la banque en question suit une distribution Poisson de paramètre 7,35 sur 11 années. Si nous supposons que les nombres annuels de pertes de plus de 1 million de dollars sont indépendants et identiquement distribués, le nombre annuel de pertes suivra alors une distribution Poisson de paramètre 0,67 (7,35/11 années).

Concernant la distribution binomiale négative, il est possible d'en calculer les paramètres, tel que décrit dans le tableau 1-8b. L'équation (9) permet effectivement de calculer le nombre moyen des pertes sur un horizon de 11 années pour la banque

⁹ Se référer à l'étude de Gurmu (1991) pour plus de détails sur l'évaluation de la matrice d'information au point $\hat{\rho}$.

prise en exemple, à savoir 8,97. De plus, si les nombres de pertes annuelles sont indépendants et identiquement distribués, la fréquence des pertes de plus de 1 million de dollars suivra une distribution binomiale de paramètres $r=0,23$ et $p=0,002$ ($p=0,025/11$ années).

Ainsi, grâce à la distribution de fréquence retenue pour une banque donnée, nous pouvons générer un nombre de pertes au-dessus de 1 million de dollars que cette banque pourrait avoir annuellement. Nous pouvons ensuite faire un tirage aléatoire de montants de pertes de plus de 1 million de dollars à partir de la base mise à l'échelle de la banque en question, et ce, selon la fréquence générée. En calculant la perte totale annuelle et en répétant ces étapes plusieurs fois, nous générons ainsi une distribution des pertes extrêmes annuelle qu'une banque pourrait subir. Il est ainsi possible de calculer des statistiques sur la distribution agrégée, telles que la moyenne et les différents quantiles.

1.7. Conclusion

L'utilisation de données externes est une exigence réglementaire pour les banques désirant développer une méthode avancée de calcul de capital. Cette utilisation doit cependant être pertinente et non biaisée. Comme nous croyons que les banques sont toutes exposées à des pertes opérationnelles peu fréquentes mais potentiellement lourdes, il est donc primordial de compléter les données de pertes internes par des pertes externes extrêmes. Ceci permet donc d'avoir une meilleure estimation des queues des distributions.

L'objectif de cette étude était de mettre en place un modèle de mise à l'échelle de la sévérité et de la fréquence des pertes externes afin de corriger le biais d'échelle, et ce, pour une meilleure utilisation des pertes externes. Les résultats de l'estimation par MCO ont montré que la taille, le lieu (États-Unis et Canada) ainsi que la ligne d'affaires (services bancaires commerciaux) et le type de risque (clients, produits et pratiques commerciales) entrent en considération dans l'explication d'une partie des montants de pertes externes. Une formule de normalisation permet de donner l'équivalent d'une perte observée dans l'industrie à l'échelle d'une banque prise

comme référence. Nous avons validé notre modèle en comparant les pertes réellement observées dans une banque avec celles obtenues après mise à l'échelle. Les résultats montrent que les deux échantillons ont des moyennes statistiquement égales et des écart-types très proches.

D'autre part, un modèle de mise à l'échelle et d'ajustement des fréquences a été développé. En effet, l'originalité de ce modèle, non abordé auparavant dans la littérature, est qu'il permet la mise des fréquences externes à l'échelle d'une banque donnée sur un horizon déterminé. De plus, ce même modèle permet de générer un nombre aléatoire de pertes de plus de 1 million qu'une banque pourrait subir. Deux modèles de comptage tronqués avec composante de régression ont été testés, à savoir le Poisson et le binomial négatif. Les résultats montrent que ce dernier domine le premier. La mise à l'échelle des fréquences se fait en calculant les paramètres de la distribution propres à la banque en question. Ces paramètres dépendent effectivement des caractéristiques de l'institution financière telles que sa taille et la répartition géographique de ses activités.

Tableau 1-1: Statistiques descriptives sur les données de la base externe

Les tableaux suivants présentent des statistiques descriptives des pertes en fonction de la taille de l'institution financière, du lieu de l'événement, du type de risque de la perte et de la ligne d'affaires où l'événement a eu lieu.

Tableau 1-1a : Statistiques descriptives de l'ensemble des pertes de la base externe en millions de dollars.

	Moyenne (M\$)	1^{er} quantile (M\$)	Médiane (M\$)	3^{ème} quantile (M\$)	Écart-type (M\$)	Nombre de pertes
Pertes externes	67,15	2,73	8,10	33,8	521,04	1056

Tableau 1-1b : Statistiques des pertes en fonction de la taille de l'institution financière estimée par le total des actifs en millions de dollars. Nous avons classé les banques en trois catégories : celles ayant des actifs inférieurs à 400 000 M\$, celles dont les actifs sont entre 400 000 M\$ et 800 000 M\$ et celles de très grande taille, avec un total des actifs de plus de 800 milliards.

	Total des actifs (Millions de dollars)		
	Actifs < 400 000	400 000 ≤ Actifs <800 000	Actifs ≥ 800 000
Moyenne des pertes (M \$)	69,08	51,42	115,57
Nombre des pertes	781	231	44
Écart-type des pertes (M \$)	594,90	114,63	408,97

Tableau 1-1c : Statistiques des pertes opérationnelles en fonction du lieu de l'événement. Nous avons classé les lieux en quatre catégories : États-Unis, Canada, Europe et Autres pays.

	Lieu			
	États-Unis	Canada	Europe	Autres pays
Moyenne des pertes (M \$)	38,42	8,96	75,20	162,51
Nombre de pertes	635	42	177	202
Écart type des pertes (M \$)	147,22	14,99	168,61	1 148,55

Tableau 1-1d : Statistiques des actifs totaux des institutions financières en fonction du lieu de l'événement. Nous avons classé les lieux en quatre catégories : États-Unis, Canada, Europe et autres pays.

	Lieu			
	États-Unis	Canada	Europe	Autres pays
Moyenne des actifs (M \$)	272 673,04	220 107,56	357 057,40	199 247, 95
Écart-type des actifs (M \$)	277 926, 19	288 945,74	283 704,46	225 372, 72

Tableau 1-1e : Statistiques des pertes opérationnelles en fonction des lignes d'affaires où l'événement de perte a eu lieu. Nous retenons la classification proposée par Bâle II, à savoir 8 lignes d'affaires : CD, PR, FC, GA, NV, SA, SBC, SBD*.

	Lignes d'affaires							
	CD	PR	FC	GA	NV	SA	SBC	SBD
Moyenne des pertes (M \$)	23,23	35,08	87,50	50,55	89,29	29,43	133,33	39,55
Nombre de pertes	174	52	55	56	82	24	268	345
Écart type des pertes (M \$)	63,16	79,55	361,28	90,87	280,86	41,79	993,64	131,20

* CD : Courtage détail / PR : Paiements et règlements / FC : Finance corporative / GA : Gestion d'actifs / NV : Négociation et vente / SA : Services d'agence / SBC : Services bancaires commerciaux / SBD : Services bancaires détail.

Tableau 1-1f : Statistiques des pertes opérationnelles en fonction du type de risque des pertes. Nous retenons la classification proposée par Bâle II, à savoir 7 types de risque : ATCE, CPPC, EPSE, FE, FI, GELP, PADS**.

	Types de risque						
	ATCE	CPPC	EPSE	FE	FI	GELP	PADS
Moyenne des pertes (M \$)	114,60	61,01	11,59	97,86	64,20	67,53	4,99
Nombre de pertes	4	460	57	222	227	80	6
Écart type des pertes (M \$)	168,83	179,29	18,50	1074,49	194,35	306,82	4,02

**ATCE : Actifs tangibles Corporels endommagés / CPPC : Clients, produits et pratiques commerciales / EPSE : Emploi, pratiques et sécurité environnementale / FE : Fraude externe / FI : Fraude interne / GELP : Gestion d'exécution et livraison des processus / PADS : Perturbation des affaires et défaillance des systèmes.

Tableau 1-2 : Résultats de l'estimation des paramètres de la régression linéaire

Le tableau suivant présente les résultats de l'estimation des coefficients de la régression linéaire par la méthode des moindres carrés ordinaires MCO. Les chiffres en italiques sont les seuils expérimentaux (p-values). La régression est la suivante :

$$Y_i = a_0 + a_1 \text{Taille}_i + a_2 \text{ÉU}_i + a_3 \text{Canada}_i + a_4 \text{Europe}_i + \sum_{j=5}^{11} a_j \text{UA}_{ij} + \sum_{j=12}^{17} a_j \text{TR}_{ij} + e_i$$

Variables	Modèle de base
Constante	0,799 <i>(0,273)</i>
Log (actifs)	0,077*** <i>(0,005)</i>
États-Unis	-0,541*** <i>(0,000)</i>
Canada	-1,087*** <i>(0,000)</i>
Europe	-0,053 <i>(0,755)</i>
Courtage détail	0,115 <i>(0,641)</i>
Finance corporative	0,386 <i>(0,199)</i>
Gestion d'actifs	0,398 <i>(0,184)</i>
Négociation et ventes	0,417 <i>(0,140)</i>
Services bancaires commerciaux	0,797*** <i>(0,001)</i>
Services bancaires détail	0,043 <i>(0,858)</i>
Service d'agence	0,478 <i>(0,213)</i>
Actifs tangibles corporels endommagés	1,300 <i>(0,196)</i>
Clients, produits et pratiques commerciales	1,123* <i>(0,079)</i>
Emploi, pratiques et sécurité environnementale	0,088 <i>(0,895)</i>
Fraudes externes	0,360 <i>(0,578)</i>
Fraudes internes	0,731 <i>(0,257)</i>

Gestion de l'exécution, de la livraison des processus	0,577 (0,381)
F(18, 1037)	8,38 (0,000)
R²	12,07 %
R²_{ajusté}	10,63 %

*** coefficient significatif à un degré de confiance de 99 %

** coefficient significatif à un degré de confiance de 95 %

* coefficient significatif à un degré de confiance de 90 %

Tableau 1-3 : Tests de robustesse

Ce tableau présente les résultats des tests de robustesse. Les modèles (1) à (3) permettent de tester la stabilité de chaque catégorie de variable dans le modèle de base. Le modèle (4) contient seulement les catégories des variables significativement non nulles dans le modèle de mise à l'échelle. Les chiffres en italiques sont les seuils expérimentaux (p-values). Les équations des modèles sont les suivantes :

Modèle (1) : $Y_i = a_0 + a_1 \text{Taille}_i + e_i$

Modèle (2) : $Y_i = a_0 + a_1 \text{Taille}_i + a_2 \text{ÉU}_i + a_3 \text{Canada}_i + a_4 \text{Europe}_i + e_i$

Modèle (3) : $Y_i = a_0 + a_1 \text{Taille}_i + a_2 \text{ÉU}_i + a_3 \text{Canada}_i + a_4 \text{Europe}_i + \sum_{j=5}^{11} a_j \text{UA}_{ij} + e_i$

Modèle (4) : $Y_i = a_0 + a_1 \text{Taille}_i + a_2 \text{ÉU}_i + a_3 \text{Canada}_i + a_4 \text{SBC}_i + a_5 \text{CPPC}_i + e_i$

	Modèle 1	Modèle 2	Modèle 3	Modèle 4
Constante	1,495*** <i>(0,000)</i>	1,797*** <i>(0,000)</i>	1,713*** <i>(0,000)</i>	1,379*** <i>(0,000)</i>
Log (actifs)	0,076** <i>(0,007)</i>	0,084*** <i>(0,003)</i>	0,077*** <i>(0,006)</i>	0,082*** <i>(0,001)</i>
États-Unis		-0,583*** <i>(0,000)</i>	-0,446** <i>(0,001)</i>	-0,595*** <i>(0,000)</i>
Canada		-1,219*** <i>(0,000)</i>	-1,055*** <i>(0,000)</i>	-1,102*** <i>(0,000)</i>
Europe		0,020 <i>(0,907)</i>	0,083 <i>(0,624)</i>	
Courtage détail			-0,275 <i>(0,271)</i>	
Finance corporative			0,423 <i>(0,166)</i>	
Gestion d'actifs			0,369 <i>(0,225)</i>	
Négociation et ventes			0,148 <i>(0,599)</i>	
Services bancaires commerciaux			0,455* <i>(0,062)</i>	0,665*** <i>(0,000)</i>
Services bancaires détail			-0,200 <i>(0,398)</i>	
Service d'agence			0,383 <i>(0,327)</i>	
CPPC				0,633*** <i>(0,000)</i>
F	7,6 <i>(0,000)</i>	12,91 <i>(0,000)</i>	8,39 <i>(0,000)</i>	22,84 <i>(0,000)</i>
R²	0,69 %	4,68 %	8,12 %	9,81 %
R² ajusté	0,60 %	4,32 %	7,16 %	9,38 %

*** coefficient significatif à un degré de confiance de 99 %

** coefficient significatif à un degré de confiance de 95 %

* coefficient significatif à un degré de confiance de 90 %

Tableau 1-4 : Statistiques sur les pertes mises à l'échelle

Ce tableau présente les statistiques relatives aux 52 pertes réellement observées à la banque Merrill Lynch. Nous présentons également, dans la troisième colonne, des statistiques calculées sur les montants des pertes de la base externe mis à l'échelle de la banque en question. L'annexe 1 montre en détail comment la mise à l'échelle a été effectuée.

	Pertes observées de Merrill Lynch	Pertes de la base externe mises à l'échelle de Merrill Lynch
Moyenne (M\$)	38,868	35,359
Médiane (M\$)	11,053	7,941
Écart type (M\$)	83,106	84,298
Coefficient d'aplatissement	21,112	35,733
Coefficient d'asymétrie	4,282	5,272
Minimum (M\$)	1,081	0,591
Maximum (M\$)	506,154	902,126
Nombre d'échantillons	52	1 050

Tableau 1-5 : Impact des variables de mise à l'échelle sur les montants des pertes

Dans ce tableau, nous présentons un événement de perte extrait de la base externe ainsi que des pertes équivalentes mises à l'échelle. Chaque fois, nous modifions une variable de la mise à échelle pour voir son impact sur le montant des pertes. La cellule en gras et en italique présente la variable qui a été modifiée par rapport à l'événement extrait de la base.

Banque	Montant de la perte	Total des actifs	Lieu	Ligne d'affaires	Type de risque
<i>Événement extrait de la base externe</i>					
<i>Bank of New York</i>	8,26 M\$	48 879 M\$	États-Unis	Services bancaires détail	Fraude externe
<i>Événements mis à l'échelle</i>					
<i>Banque fictive 1</i>	9,27 M\$	200 000 M\$	États-Unis	Services bancaires détail	Fraude externe
<i>Banque fictive 2</i>	4,97 M\$	48 879 M\$	Canada	Services bancaires détail	Fraude externe
<i>Banque fictive 3</i>	16,06 M\$	48 879 M\$	États-Unis	Services bancaires commerciaux	Fraude externe
<i>Banque fictive 4</i>	15,56 M\$	48 879 M\$	États-Unis	Services bancaires détail	Clients, produits et pratiques commerciales
<i>Banque fictive 5</i>	5,58 M\$	200 000 M\$	Canada	Services bancaires détail	Fraude externe
<i>Banque fictive 6</i>	10,86 M\$	200 000 M\$	Canada	Services bancaires commerciaux	Fraude externe
<i>Banque fictive 7</i>	20,46 M\$	200 000 M\$	Canada	Services bancaires commerciaux	Clients, produits et pratiques commerciales

Tableau 1-6 : Statistiques descriptives sur les variables introduites dans le modèle de fréquence

Les tableaux suivants présentent des statistiques descriptives du nombre des pertes par banque selon la taille de l'institution financière et la répartition géographique de ses activités sur la période 1994-2004.

Tableau 1-6a : Statistiques calculées sur le nombre de pertes par banque et le total des actifs par banque sur la période 1994-2004.

	Moyenne	Écart type	1 ^{er} quantile	Médiane	3 ^{ème} quantile	Min	Max
Nombre de pertes par banque	3,27	6,13	1	1	3	1	52
Total des actifs par banque (M\$)	123 174	181 820	5 758	34 003	165 057	43	887 515

Tableau 1-6b : Statistiques du nombre des pertes par banque en fonction de la taille de l'institution financière estimée par le total des actifs en millions de dollars. Nous avons classé les banques en trois catégories : celles dont la moyenne des actifs sur la période 1994-2004 est inférieure à 400 000 M\$, celles dont la moyenne des actifs est entre 400 000 M\$ et 800 000 M\$ et celles de très grande taille, avec une moyenne d'actifs de plus de 800 milliards sur la période en question.

	Total des actifs (millions de dollars)		
	Actifs < 400 000	400 000 ≤ Actifs < 800 000	Actifs ≥ 800 000
Nombre moyen des pertes par banque	2,587	8,538	19
Nombre de banques	293	26	4
Écart type du nombre de pertes par banque	4,647	10,879	18,166

Tableau 1-6c : Statistiques du nombre des pertes opérationnelles en fonction du degré de répartition géographique des activités. On distingue 4 environnements différents : États-Unis, Canada, Europe et Autres pays. Les colonnes du tableau suivant montrent si les activités sont concentrées dans un même environnement ou réparties entre 2, 3 ou 4 environnements.

	Répartition géographique			
	Un environnement	Deux environnements	Trois environnements	Quatre environnements
Nombre moyen des pertes par banque	2	10	7,667	28
Nombre de banques	275	33	12	3
Écart type du nombre de pertes par banque	2,957	12,661	4,559	13,077

Tableau 1-7 : Résultats de l'estimation des coefficients dans le modèle de fréquence

Le tableau suivant présente les résultats de l'estimation des coefficients dans le modèle de mise à l'échelle des fréquences. La méthode d'estimation utilisée est le maximum de vraisemblance. La dernière ligne donne le résultat du score test ou le test du multiplicateur de Lagrange qui permet de comparer les deux modèles. Les chiffres en italiques représentent les *t* de Student.

	Modèle Poisson tronqué avec composante de régression¹⁰	Modèle binomial négatif tronqué avec composante de régression⁹
Constante	-5,876*** <i>(-15,28)</i>	-10,439*** <i>(-8,35)</i>
Log (actifs)	1,176*** <i>(15,30)</i>	1,783*** <i>(8,45)</i>
États-Unis	1,432*** <i>(14,20)</i>	2,000*** <i>(6,60)</i>
Canada	0,559*** <i>(5,51)</i>	1,721*** <i>(2,83)</i>
Europe	0,141* <i>(1,73)</i>	-0,111 <i>(-0,34)</i>
Autres pays	0,191** <i>(2,43)</i>	0,457 <i>(1,52)</i>
α		4,347 <i>(1,53)</i>
Log (fonction de vraisemblance)	-695,314	-439,414
Score test	34,637	

*** coefficient significatif à un degré de confiance de 99 %

** coefficient significatif à un degré de confiance de 95 %

* coefficient significatif à un degré de confiance de 90 %

¹⁰ L'estimation de ces paramètres est faite par les logiciels Matlab et Limdep

Tableau 1-8 : Résultats de l'application de la mise à l'échelle des fréquences

Ce tableau présente une application du modèle de la mise à l'échelle des fréquences. Nous considérons une banque dont le total moyen de ses actifs sur la période 1994-2004 est évalué à 100 milliards de dollars. Ses activités sont réparties principalement entre les États-Unis et le Canada. Dans ce qui suit, nous déterminons les paramètres de la distribution Poisson et la binomiale négative propre à cette banque.

Tableau 1-8a : Modèle Poisson avec composante de régression.

	Constante β_0	Log (actifs) β_1	États-Unis β_2	Canada β_3	Europe β_4	Autres pays β_5	Paramètre de Poisson
Coefficients	-5,876	1,176	1,432	0,559	0,141	0,191	7,352*
Variables		Log(100000)	1	1	0	0	

$$* \lambda_i = \exp(\beta_0 + \beta_1 \log(\text{Actifs})_i + \beta_2 \text{ÉU}_i + \beta_3 \text{Canada}_i + \beta_4 \text{Europe}_i + \beta_5 \text{Autre}_i)$$

Tableau 1-8b : Modèle de la binomiale négative avec composante de régression.

	Constante β_0	Log (actifs) β_1	États-Unis β_2	Canada β_3	Europe β_4	Autres β_5	Paramètre de la gamma α	Paramètres de la binomiale négative (r,p)
Coefficients	-10,439	1,783	2	1,721	-0,111	0,457	4,347	(0,23; 0,025)**
Variables		Log (100000)	1	1	0	0		

$$** \begin{cases} r = \frac{1}{\alpha} \\ p = \frac{1}{1 + \alpha \exp(\beta_0 + \beta_1 \log(\text{Actifs})_i + \beta_2 \text{ÉU}_i + \beta_3 \text{Canada}_i + \beta_4 \text{Europe}_i + \beta_5 \text{Autre}_i)} \end{cases}$$

Annexe 1

Dans ce qui suit, nous présentons trois exemples de pertes extraits de la base externe et nous montrons en détail comment la mise à l'échelle est faite.

Banque	Perte observée (M\$)	Année	Total des actifs (M\$)	Lieu	Ligne d'affaires	Type de risque
BankAmerica Corp.	4,70	1994	169 604	États-unis	Négociation et ventes	Clients, produits et pratiques commerciales
Banque de France	34,32	1996	114 959	Europe	Services bancaires détail	Fraude interne
Banque de Montréal	10,83	2002	157 780	Canada	Services bancaires commerciaux	Fraude interne

Nous mettons ces pertes à l'échelle de la banque Merrill Lynch. Nous conservons la ligne d'affaires, le type de risque et l'année d'occurrence des pertes de la base de données et nous remplaçons les variables total des actifs et lieu par celles propres à la banque Merrill Lynch.

Banque	Perte mise à l'échelle (M\$)	Année	Total des actifs (M\$)	Lieu	Ligne d'affaires	Type de risque
Merrill Lynch	À déterminer	1994	163 749	États-Unis	Négociation et ventes	Clients, produits et pratiques commerciales
Merrill Lynch	À déterminer	1996	213 016	États-Unis	Services bancaires détail	Fraude interne
Merrill Lynch	À déterminer	2002	447 928	États-Unis	Services bancaires commerciaux	Fraude interne

Pour déterminer la perte équivalente, nous devons déterminer les composantes idiosyncratiques des pertes de la base externe ainsi que celles de Merrill Lynch à partir de la régression (4).

Banque	Perte observée (perte externe) (A)	Composante idiosyncratique (perte externe) (B)	Composante idiosyncratique (Merrill Lynch) (C)	Perte mise à l'échelle (Merrill Lynch) (D)=(A)x(C/B)
BankAmerica Corp.	4,70	2,80 ^a	2,79 ^d	4,69
Banque de France	34,32	2,61 ^b	1,51 ^e	19,86
Banque de Montréal	10,83	1,73 ^c	3,13 ^f	19,59

$$^a 2,78 = \exp (0,082 \times \log (\text{Actifs}) - 0,595 \times \text{ÉU} - 1,102 \times \text{Canada} + 0,665 \times \text{SBC} + 0,633 \times \text{CPPC}) \\ = \exp (0,082 \times \log (169\ 604) - 0,595 \times 1 - 1,102 \times 0 + 0,665 \times 0 + 0,633 \times 1)$$

$$^b 2,61 = \exp (0,082 \times \log (\text{Actifs}) - 0,595 \times \text{ÉU} - 1,102 \times \text{Canada} + 0,665 \times \text{SBC} + 0,633 \times \text{CPPC}) \\ = \exp (0,082 \times \log (114\ 959) - 0,595 \times 0 - 1,102 \times 0 + 0,665 \times 0 + 0,633 \times 0)$$

$$^c 1,73 = \exp (0,082 \times \log (\text{Actifs}) - 0,595 \times \text{ÉU} - 1,102 \times \text{Canada} + 0,665 \times \text{SBC} + 0,633 \times \text{CPPC}) \\ = \exp (0,082 \times \log (157\ 780) - 0,595 \times 0 - 1,102 \times 1 + 0,665 \times 1 + 0,633 \times 0)$$

$$^d 2,79 = \exp (0,082 \times \log (\text{Actifs}) - 0,595 \times \text{ÉU} - 1,102 \times \text{Canada} + 0,665 \times \text{SBC} + 0,633 \times \text{CPPC}) \\ = \exp (0,082 \times \log (163\ 749) - 0,595 \times 1 - 1,102 \times 0 + 0,665 \times 0 + 0,633 \times 1)$$

$$^e 1,51 = \exp (0,082 \times \log (\text{Actifs}) - 0,595 \times \text{ÉU} - 1,102 \times \text{Canada} + 0,665 \times \text{SBC} + 0,633 \times \text{CPPC}) \\ = \exp (0,082 \times \log (213\ 016) - 0,595 \times 1 - 1,102 \times 0 + 0,665 \times 0 + 0,633 \times 0)$$

$$^f 3,13 = \exp (0,082 \times \log (\text{Actifs}) - 0,595 \times \text{ÉU} - 1,102 \times \text{Canada} + 0,665 \times \text{SBC} + 0,633 \times \text{CPPC}) \\ = \exp (0,082 \times \log (447\ 928) - 0,595 \times 1 - 1,102 \times 0 + 0,665 \times 1 + 0,633 \times 0)$$

2. MODÈLE GB2 FRACTIONNÉ POUR UN MEILLEUR AJUSTEMENT DE LA DISTRIBUTION DE SÉVÉRITÉ

2.1. Introduction

La « Loss Distribution Approach » (LDA) constitue une méthode avancée de calcul de la perte non anticipée dans le contexte du risque opérationnel. Cette méthode a été inspirée des méthodes déjà utilisées dans le domaine de l'assurance pour estimer les primes ou le MPY¹¹ (Cummins et Freifelder, 1978, Cummins et Wiltbank, 1983). Cette méthode consiste, tout d'abord, à estimer la distribution de sévérité ainsi que celle des fréquences. Ensuite, il faut agréger les deux distributions pour déterminer les montants des pertes annuelles. Enfin, la simulation de ces pertes permet de calculer le 99.9^{ème} centile de la distribution agrégée qui constitue la perte annuelle non anticipée ou la valeur à risque à un niveau de confiance de 99.9%, tel que recommandé par Bâle II.

L'avantage principal d'adopter une telle approche est de refléter le degré d'exposition de la banque face au risque opérationnel. Cependant, pour aboutir à un résultat fiable et précis, il est important de bien choisir la distribution paramétrique susceptible de bien décrire le comportement des pertes. Cette étape est cruciale dans le développement de la méthode puisqu'une mauvaise spécification de la distribution entraîne une sur ou une sous évaluation du capital risque opérationnel. L'approche perdra ainsi son avantage majeur de refléter le niveau de risque opérationnel d'une institution.

Malheureusement, la distribution de sévérité est généralement mal choisie. En effet, en pratique, la plupart des institutions bancaires supposent que la distribution lognormale est la bonne candidate sans justifier leur choix. Dans ce chapitre, nous

¹¹ MPY : Maximum Probable Yearly aggregate loss.

allons mettre l'accent sur l'ajustement des distributions paramétriques aux données de pertes. Nous allons donc décortiquer le processus d'ajustement et proposer un modèle permettant de bien décrire le comportement des pertes.

L'objectif de notre étude est de proposer une méthodologie robuste pour bien estimer la distribution de sévérité. Nous proposons dans un premier temps quatre distributions à tester incluant une famille de distribution à quatre paramètres très flexible. Nous construisons, par la suite, un test Kolmogorov-Smirnov avec *bootstrap* paramétrique permettant de valider l'hypothèse du bon ajustement.

Par ailleurs, nous introduisons la notion de symétrie pondérée que nous validons par des tests non paramétriques (Abdous, Ghoudi et Rémillard, 2003). Ensuite, nous construisons un modèle GB2 fractionné, qui consiste à combiner le modèle GB2 avec la symétrie pondérée. Les résultats montrent que le modèle ainsi construit offre un excellent ajustement aux données de pertes.

Ce chapitre est structuré comme suit. Après cette introduction, nous passons en revue, dans la première section, les principales études faites à ce sujet. Dans la section 2, nous décrivons les données de pertes opérationnelles. Nous proposons dans la troisième section les distributions, susceptibles de décrire le comportement des pertes, que nous testons. Nous introduisons la notion de la symétrie pondérée et nous proposons un modèle GB2 fractionné dans la quatrième section. Enfin, la dernière section est réservée à la conclusion et à la suggestion d'extensions intéressantes au modèle que nous considérons.

2.2. Revue de la littérature

Une mauvaise spécification de la distribution s'ajustant aux données de sévérité peut surévaluer ou sous-évaluer la perte non anticipée. Chernobai, Menn, Rachev et Truck (2005c) considèrent plusieurs distributions pour estimer la distribution de sévérité. La perte anticipée ainsi que la valeur à risque (VaR) à un degré de confiance de 95 et 99 % ont été calculées avec ces différentes distributions. Il est à constater que les résultats sont très différents selon la distribution choisie. La valeur de la perte non

anticipée dépend énormément de l'épaisseur de la queue de distribution. Les distributions de sévérité à queues épaisses donnent effectivement des valeurs très élevées du capital contrairement aux distributions à queues fines. C'est pourquoi, il est très important de trouver la meilleure distribution qui décrit le comportement des pertes opérationnelles. La VaR ainsi calculée reflétera l'exposition réelle de la banque face à ce risque. Ce chapitre mettra l'accent sur une meilleure estimation de la distribution de sévérité.

Fontnouvelle et Rosengren (2004) ajustent plusieurs distributions de sévérité (à un et deux paramètres), avec des épaisseurs de queues variables, à des données de pertes internes de six banques. Les résultats montrent que les distributions à queues épaisses telles que la Pareto s'ajustent bien aux données. Cependant, dans certains cas, les résultats des tests d'ajustement ne sont pas concluants et aucune des distributions testées ne décrit le comportement des pertes opérationnelles. Il y a lieu donc de trouver un modèle qui s'ajuste aux données lorsque les distributions usuelles sont rejetées.

D'autre part, la théorie des valeurs extrêmes (TVE) a été fréquemment utilisée par de nombreux auteurs pour la quantification du risque opérationnel. La TVE consiste à modéliser seulement les événements extrêmes de pertes. En effet, Medova (2001), Ebnother, Vanini, McNeil et Antolinez-Fehr (2001) et Moscadelli (2004) utilisent la GPD (Generalized Pareto Distribution) pour modéliser l'excédent des pertes opérationnelles par rapport à un seuil élevé. D'un autre côté, il y a des études qui ont remis en question l'utilisation de cette théorie dans le contexte du risque opérationnel. Dans ce sens, les travaux de Diebold, Schuermann et Stroughair (1998), de Embrechts, Furrer et Kaufmann (2003), de Embrechts, Kaufmann et Samorodnitsky (2004) et de Nešlehová, Embrechts et Chavez-Demoulin (2006) montrent effectivement les conditions requises pour l'application de la TVE et ses limites dans le contexte du risque opérationnel. Les données doivent, en fait, être stationnaires, répétitives et nombreuses. Toutefois, ces critères ne sont pas satisfaits par les données de pertes opérationnelles. Vu ces limites, nous écartons l'application de la TVE de notre étude sur l'ajustement des distributions aux pertes.

Dans la présente étude, nous proposons l'utilisation d'une distribution flexible à quatre paramètres, la Béta Généralisée de second type ou la GB2, et la notion de symétrie pondérée pour modéliser la sévérité des pertes opérationnelles.

La GB2 est une famille de distribution à quatre paramètres. Cette distribution est extrêmement flexible. Elle comprend plusieurs distributions usuelles selon la valeur de ces paramètres. Nous citons la lognormale, Log-t, Gamma, Burr2 et Log-Cauchy comme cas particuliers, voir graphique 2-1. Sa flexibilité permet de modéliser différentes épaisseurs de queues de distributions. La GB2 a été introduite par McDonald (1984) et a montré jusqu'à présent des excellents résultats de modélisation dans plusieurs applications.

En effet, elle a été utilisée pour modéliser les pertes assurables dues à des incendies (Cummins, Dionne, McDonald et Pritchett, 1990), ou dues à des ouragans et des tremblements de terre (Cummins, Lewis et Phillips, 1999). Cette famille de distribution a été utilisée également pour modéliser les rendements et les prix des titres (Bookstaber et McDonald, 1987, Dutta et Babbel, 2002 et Tunaru, Kadam et Albota, 2005). Il a même été montré par McDonald et Xu (1995) que la GB2 est la meilleure parmi tous les cas de distributions étudiés (GB, GB1, GB2, B, B1, B2, GG, Burr3, Burr12, GA, LN)¹² pour s'ajuster aux distributions de revenus et de rendements des actions. Ayant démontrée sa grande flexibilité, nous allons utiliser cette distribution comme candidate pour l'ajustement aux données de pertes opérationnelles. Il est à noter que cette distribution a été effectivement utilisée par Dutta et Perry (2006) pour la modélisation des pertes opérationnelles. Cependant, les auteurs ont souligné la complexité de la modélisation et la difficulté de générer des variables aléatoires selon cette distribution. Nous allons donc montrer dans ce chapitre comment il est possible de générer des variables à partir de la GB2.

Par ailleurs, les pertes opérationnelles ne sont pas symétriques. En effet, nous observons que les événements à perte extrême sont moins fréquents que les

¹² GB : Beta généralisée, GB1 : Beta Généralisée de premier type, GB2 : Beta Généralisée du second type, B : Beta, B1 : Beta de premier type, B2 : Beta de second type, GG : Gamma généralisée, GA : Gamma, LN : lognormale

événements à petite perte. La notion de symétrie pondérée a été abordée dans la littérature pour s'accommoder à ce type de données qui présentent une certaine asymétrie. Il s'agit d'une généralisation de la notion classique de symétrie où les queues des distributions sont similaires, à un facteur près. Une propriété importante de ce concept consiste à diviser ou fractionner la distribution en deux modèles qui peuvent être construits en attribuant différents poids aux parties droite et gauche de la même distribution.

Ces modèles fractionnés ont été appliqués par Lefrançois (1989) et Scallan (1995) et, plus récemment, par Gouriéroux et Monfort (2002) dans l'évaluation des options. Abdous, Ghoudi et Rémillard (2003) ont développé des statistiques permettant de tester l'existence de la symétrie pondérée. Ils ont notamment validé statistiquement l'hypothèse de la symétrie pondérée pour deux études dont l'une porte sur les fluctuations des prix de l'essence à la pompe alors que l'autre porte sur les fluctuations des rendements des indices TSE300 et Dow Jones. Dans ce chapitre, nous nous basons sur ces statistiques pour tester la présence de la symétrie pondérée dans les données de pertes opérationnelles.

Comme nous l'avons mentionné précédemment, il est important de trouver la distribution qui s'ajuste le mieux aux données. La seule estimation des paramètres n'est pas suffisante pour choisir la bonne distribution. Parfois la valeur des paramètres estimés n'est pas raisonnable (Fontnouvelle et Rosengren, 2004). Des tests de bon ajustement sont donc nécessaires pour nous permettre de décider si l'ajustement est pertinent ou pas. Plusieurs tests graphiques et formels existent dans la littérature pour l'ajustement des distributions continues (Klugman, Panjer et Willomot, 1998 et Cruz, 2002). Parmi les tests formels, nous retenons le test de Kolmogorov-Smirnov. Cependant pour obtenir un meilleur résultat, nous allons construire ce test avec *bootstrap* paramétrique. En effet, une façon de calculer la p-value et les valeurs critiques est par simulation Monte Carlo, pour chaque distribution ajustée hypothétique, Ross (2001) et Chernobai, Rachev et Fabozzi (2005b).

2.3. Description des données

Nous allons prendre la base de données externe OpVaR de Fitch, mise à l'échelle d'une banque canadienne. La mise à échelle a été faite en se basant sur le modèle développé au chapitre précédent. En fait, l'ajustement des distributions aux données de pertes permet de choisir la bonne distribution qui décrit mieux le comportement des pertes. Une fois la distribution choisie, il nous est possible de générer des montants de pertes extrêmes qui pourraient survenir à la banque en question.

Nous écartons de la base des données de pertes mises à l'échelle, les pertes trop extrêmes ou *outliers*. En effet, garder ces pertes trop importantes va fausser l'estimation de la distribution des données de plus de 1M\$. Ainsi, l'idée d'écarter les *outliers* selon les études de Hampel et al. (1986), Huber (1981) et McLachlan et Basford (1988) est d'avoir une estimation plus robuste de la distribution.

Dans le contexte du risque opérationnel, Chernobai et Rachev (2006) montrent que l'estimation des paramètres est plus robuste lorsque les pertes très extrêmes ou *outliers* de la base de données sont écartées. D'ailleurs, les auteurs ont pris deux échantillons de pertes opérationnelles de plus d'un million de dollars, l'un comprend toutes les données alors que le deuxième exclut 5% des pertes les plus extrêmes. Ils ont montré que les statistiques et les paramètres des distributions estimées à partir du premier échantillon sont biaisés puisque les quelques pertes extrêmes vont impacter la moyenne, l'écart type, les coefficients d'aplatissement et d'asymétrie. Ils ont montré également que le premier échantillon surévalue la VaR de 61% par rapport au deuxième. Ainsi, en gardant les pertes très extrêmes, la forme de la distribution paramétrique risque d'être influencée par les quelques grosses pertes et ne va pas nécessairement bien décrire le corps de la distribution.

De plus, à cause du biais de sélection, les pertes trop extrêmes ont une probabilité plus élevée d'être rapportées. En gardant les *outliers*, nous allons leur attribuer un poids encore plus important. Ainsi, nous excluons 5% des pertes les plus élevées de la base de données, ce qui correspond à 52 observations. Nous écartons ainsi toutes

les pertes au-dessus de 100 millions \$ considérées comme étant des *outliers*. Il est à mentionner toutefois qu'il est possible de générer des pertes au dessus de la perte maximale (100 M \$) avec la distribution que nous retiendrons lors de ce chapitre.

Le tableau 2-1 présente des statistiques descriptives des pertes opérationnelles. Nous remarquons que les pertes sont très dispersées, elles varient de 303 746 \$ à 97,8 millions \$. Par ailleurs, nous remarquons que les données ne sont pas symétriques. Une simple comparaison entre la moyenne et la médiane montre un grand écart (10,5 millions versus 3,7 millions), confirmant l'asymétrie dans les données. De plus, ce constat est appuyé par un coefficient d'asymétrie positif de l'ordre de 2,65 et un coefficient d'aplatissement de 7,7. Ainsi, il est évident que les données ne peuvent pas être modélisées par une distribution normale. Dans ce qui suit, nous proposons d'identifier le modèle qui pourrait modéliser le comportement de ces pertes.

2.4. Modèles d'ajustement

Comme nous l'avons indiqué précédemment, le bon ajustement de la distribution aux données est très important pour le calcul du capital risque opérationnel. C'est pourquoi nous proposons plusieurs distributions à tester. Par ailleurs, la notion de symétrie pondérée est également testée plus tard dans le chapitre pour offrir plus de possibilités d'ajustement aux données.

2.4.1. Distributions proposées

Nous commençons par choisir laquelle des diverses distributions paramétriques s'ajustera le mieux aux données de pertes. Théoriquement, toutes les distributions continues, dont le domaine de définition est positif, se présentent comme candidates à la modélisation de la distribution de sévérité. Nous allons tester trois distributions usuelles (exponentielle, lognormale et Weibull), en plus de la famille de distribution GB2. Nous décrivons les caractéristiques de chacune d'elles et nous présentons leur fonction de densité et de répartition dans le tableau récapitulatif suivant.

Tableau 2-1: Distributions étudiées

Distribution	Fonctions de densité	Fonctions de répartition
exponentielle	$f(x \theta) = \frac{e^{-\frac{x}{\theta}}}{\theta}$	$F(x \theta) = 1 - e^{-\frac{x}{\theta}}$
logNormale	$f(x \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ où $z = \frac{\log(x) - \mu}{\sigma}$	$F(x \mu, \sigma) = \Phi(z)$ où $z = \frac{\log(x) - \mu}{\sigma}$
Weibull	$f(x \alpha, \beta) = \frac{\beta \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta}}{x}$	$F(x \alpha, \beta) = 1 - e^{-\left(\frac{x}{\alpha}\right)^\beta}$
GB2	$f(x a, b, p, q) = \frac{ a \left(\frac{x}{b}\right)^{ap}}{\beta(p, q)x \left[1 + \left(\frac{x}{b}\right)^a\right]^{p+q}}$ avec $\forall x \geq 0$ et b, p et $q > 0$	$F(x a, b, p, q) = B_{\text{inc}}(z p, q)$ où $z = \frac{\left(\frac{x}{b}\right)^a}{1 + \left(\frac{x}{b}\right)^a}$ avec B_{inc} : fonction de Beta Incomplète

Tout d’abord, nous testons la distribution exponentielle qui offre l’avantage d’être simple ayant un seul paramètre. Ensuite, la distribution lognormale est testée. Cette distribution à deux paramètres est largement utilisée en pratique. D’ailleurs, cette distribution est prise, par hypothèse, comme distribution s’ajustant aux montants de pertes dans plusieurs études (Chernobai, Menn, Truck et Rachev, 2005b, Frachot, Moudoulaud et Roncalli, 2003). Cette distribution ainsi que la distribution exponentielle sont caractérisées par des queues moyennement épaisses. D’autre part, la distribution Weibull est également testée. Elle est dotée de deux paramètres.

Contrairement aux deux précédentes, c'est une distribution à queue fine. Ainsi, ces trois distributions offrent des épaisseurs de queues différentes.

Enfin, nous testons également, une famille de distribution à quatre paramètres, moins utilisée en pratique que les autres certes, mais elle offre une grande flexibilité. En effet, le nombre élevé de ses paramètres lui permet d'avoir de nombreuses formes ainsi que différentes épaisseurs de queue. Le graphique 2-1 montre des cas particuliers de la famille GB2 qui correspondent à des distributions usuelles. Comme il s'agit d'une distribution moins utilisée en pratique, nous allons faire un survol sur certaines de ses caractéristiques.

2.4.1.1. Certaines caractéristiques de la GB2

Génération de variables :

Il est possible de générer des variables suivant la GB2 à partir de la distribution beta comme suit.

Soit Y une variable aléatoire non nulle suivant la distribution $\text{beta}(p,q)$, nous définissons la variable X telle que

$$X = b \left(\frac{1}{Y} - 1 \right)^{-\frac{1}{a}} \sim \text{GB2}(a, b, p, q)$$

Moments d'ordre h :

Quant aux moments d'ordre h de la GB2, ils ne sont définis que lorsque $aq > h$. Ils ont l'expression suivante :

$$E(X^h) = b^h \frac{B\left(p + \frac{h}{a}, q - \frac{h}{a}\right)}{B(p, q)}$$

Ainsi, seuls les paramètres a et q sont impliqués dans la détermination du nombre de moments finis de la distribution. Ces mêmes paramètres nous donnent une indication sur le comportement de la queue du côté droit. En effet, plus le produit $a \cdot q$ est grand, plus la queue est petite.

Après ce bref survol sur les distributions proposées à des fins de modélisation des pertes opérationnelles, nous passons à l'estimation des paramètres de ces quatre distributions à la section suivante.

2.4.2. Estimation des paramètres

Nous proposons dans ce qui suit une estimation des paramètres de chacune des distributions décrites ci-haut. Ensuite, nous implantons dans la prochaine section des tests d'ajustement qui vont nous permettre de choisir quelle distribution décrit le mieux la variabilité des pertes opérationnelles.

Deux principales approches sont proposées dans la littérature pour estimer les paramètres. La première consiste à estimer les paramètres à partir d'un système d'équations égalisant le nombre de paramètres. Trois méthodes populaires composent cette approche : la méthode des moments, la méthode des centiles ou la méthode des moments pondérés. La limite de cette approche est son incapacité de bien ajuster la distribution à toutes les données. Elle se concentre plutôt à ajuster les moments ou les quantiles.

La deuxième approche consiste à estimer les paramètres par optimisation ou maximum de vraisemblance. Dans ce cas, la fonction de vraisemblance d'un ensemble de n observations indépendantes x est estimée par :

$$L(\theta) = \prod_{j=1}^n f(x_j; \theta)$$

où f est la fonction de densité dont les paramètres sont représentés par le vecteur θ .

Les paramètres estimés sont la solution du programme de maximisation de la fonction de vraisemblance précédente. Le tableau 2-2 présente les résultats de l'estimation des paramètres pour les quatre distributions définies ci-haut. Une fois l'estimation des paramètres est accomplie, il est important de restreindre notre choix à un seul modèle et à des paramètres estimés uniques.

2.4.3. Tests d'ajustement

Dans ce qui suit, nous allons mettre en place des tests d'ajustement afin de permettre de tester l'acceptabilité des différents modèles et de les classer pour pouvoir sélectionner le meilleur. Un test d'hypothèse permettra de tester l'acceptabilité de la distribution. En effet, l'hypothèse nulle et l'hypothèse alternative peuvent être formulées de la sorte :

$$H_0 : F_x(x) = F(x; \theta)$$

$$H_1 : F_x(x) \neq F(x; \theta)$$

Avec : $F_x(x)$: la fonction de répartition empirique

$F(x; \theta)$: une des fonctions de répartition des distributions étudiées

Ainsi, le modèle offre un bon ajustement dans le cas où l'hypothèse nulle ne serait par rejetée.

Nous commençons par des tests informels ou graphiques qui nous permettent d'avoir une idée globale du degré d'ajustement de la distribution paramétrique à la distribution empirique. Nous commençons par le graphique des quantiles (Q-Q plot). Il s'agit de faire le graphique des quantiles de la distribution paramétrique versus les quantiles empiriques. Un bon ajustement consiste en une courbe linéaire proche de la droite de 45 degrés. Ensuite, un graphique de la fonction de répartition théorique versus la fonction de répartition empirique peut être fait afin de voir si les deux distributions ont le même comportement. Généralement, si ces tests montrent que l'ajustement est bon, il vaut mieux confirmer l'évidence visuelle par un test plus formel basé sur le calcul d'une statistique. Par contre, si les graphiques montrent que la distribution empirique est loin d'être dérivée de la distribution paramétrique testée, alors il est invraisemblable que les tests formels donnent des résultats différents.

Ainsi, des tests formels viennent confirmer ou infirmer les tests graphiques. Nous construisons un test Kolmogorov-Smirnov (KS) pour nous assurer des résultats des tests graphiques. Le test KS consiste à mesurer l'écart maximal absolu entre la

fonction de répartition empirique et celle du modèle. La statistique KS se calcule comme suit :

$$KS = \sqrt{n} \sup_x |F_x(x) - F(x; \hat{\theta})|$$

où $\hat{\theta}$ désigne le vecteur des paramètres estimés.

Cette statistique est comparée à une valeur critique tabulée. Si la statistique calculée est inférieure à la statistique tabulée, alors l'hypothèse nulle ne sera pas rejetée. Cependant pour des résultats plus fiables du test, nous proposons une méthode KS avec *bootstrap* paramétrique.

Cette méthode, contrairement au test classique, consiste à calculer une p-value et les valeurs critiques par simulation Monte Carlo. La valeur de la p-value nous permet de décider si nous rejetons ou non l'hypothèse nulle stipulant le bon ajustement de la distribution aux données. Ainsi, les valeurs critiques vont être calculées à partir d'échantillons générés. Il s'agit d'accomplir donc les étapes de l'algorithme suivant pour le calcul de la p-value :

- 1- Calculer la statistique KS_0 , telle que définie précédemment, avec la fonction de répartition dont les paramètres déjà estimés sont $\hat{\delta}$.
- 2- Générer, à partir des paramètres estimés de la distribution à tester, un échantillon de montants de pertes de même taille que l'échantillon de départ.
- 3- À partir de l'échantillon généré, estimer les paramètres $\tilde{\theta}$ de la même distribution avec la méthode du maximum de vraisemblance.
- 4- Calculer la statistique KS_i avec la fonction de répartition dont les paramètres nouvellement estimés sont $\tilde{\theta}$.
- 5- Comparer les statistiques KS_0 et KS_i . Si $KS_0 < KS_i$ alors un compteur j s'incrmente de 1.

6- Répéter les étapes 2, 3, 4 et 5, un grand nombre de fois N (ex. $N=10000$)

7- Calculer la p-value comme étant $pv = \frac{j}{N}$

8- Rejeter l'hypothèse nulle, si la p-value (pv) est plus petite que le niveau de confiance (5 %).

2.4.4. Résultats

Nous résumons les résultats de l'estimation des paramètres pour les quatre distributions étudiées dans le tableau 2-2. Il est à remarquer que nous avons eu des solutions uniques au programme de maximisation de la fonction de vraisemblance. Les valeurs des paramètres de chacune des distributions sont raisonnables. À ce stade, il est important de tester la validité de chacun des modèles grâce aux tests graphiques et des tests formels tel que le test KS défini précédemment.

Les tests informels sont présentés dans les graphiques 2-2, 2-3, 2-4 et 2-5 pour l'ajustement respectif des distributions exponentielle, lognormale, Weibull et GB2 aux données. À partir de ces graphiques, il est clair que les distributions exponentielle et Weibull ne s'ajustent pas à la distribution empirique. En effet, les graphiques des quantiles ne sont pas linéaires et loin de la droite de 45 degrés visuellement. De plus, les distributions de répartition empirique et paramétrique ne sont pas proches l'une de l'autre. À ce stade, l'hypothèse d'un bon ajustement de ces distributions aux données semble être rejetée.

Bien que les graphiques des quantiles semblent être plus linéaires pour le cas des distributions lognormale et GB2, les distributions paramétriques ne s'ajustent pas bien à la queue de la distribution empirique visuellement. En effet, nous remarquons que la courbe s'écarte de la droite de 45 degrés lorsque les pertes sont supérieures à 50 millions \$. Il est important de valider ces résultats par des tests formels basés sur le calcul de statistique robuste. Les tests graphiques donnent donc un aperçu sur la qualité de l'ajustement qu'il faut appuyer par le test KS.

Les résultats du test KS sont présentés au tableau 2-2, à la dernière ligne. La valeur de la p-value sera un critère d'acceptabilité et de sélection du modèle. En effet, plus la p-value est élevée, dépassant le seuil de confiance de 5%, meilleur est l'ajustement. Nous remarquons que les distributions exponentielle, lognormale et Weibull ne sont pas acceptables, puisque leur p-value est nulle. De même, la distribution GB2 dont la valeur du test est trop faible (0.004) ne permet pas de décrire le comportement des montants de pertes.

D'autre part, les modèles peuvent être classés en se basant sur la valeur du logarithme de la fonction de vraisemblance. Les résultats montrent que la distribution GB2 a la valeur la plus élevée du log de vraisemblance (-3156,695) par rapport aux autres distributions. Toutefois, le modèle n'est pas accepté à un niveau de confiance de 5%.

2.5. Un modèle GB2 fractionné

Nous introduisons dans ce qui suit la notion de symétrie pondérée. En effet, comme les données sont loin d'être symétriques¹³, nous proposons un modèle pour l'ajustement des données de pertes, en introduisant le concept de symétrie pondérée. Ces résultats confirment nos attentes concernant la flexibilité de la GB2 par rapport aux autres distributions.

Nous considérons la distribution GB2 que nous combinons avec la notion de symétrie pondérée pour construire un modèle fractionné. Nous avons choisi la GB2 à cause de sa flexibilité par rapport aux autres distributions. De plus, nous rappelons que le graphique des quantiles empiriques versus ceux de la GB2 a révélé que cette dernière ne s'ajustait pas en particulier aux grosses pertes.

Nous nous attendons à ce que ce modèle fractionné que nous allons construire s'ajuste mieux aux données de pertes, dans le cas où une symétrie pondérée serait détectée. Nous allons donc, dans une première étape, définir le concept de symétrie pondérée et tester sa présence dans les données de pertes opérationnelles. Dans le cas

¹³ Voir la section sur les statistiques descriptives pour une comparaison entre la moyenne et la médiane des données d'une part et pour la valeur du coefficient d'asymétrie d'autre part.

où l'hypothèse de la présence de cette symétrie est validée, nous estimons, dans une deuxième étape, les paramètres du modèle GB2 fractionné où la distribution GB2 est divisée en deux parties.

2.5.1. Définition de la symétrie pondérée

Une variable aléatoire X ayant une fonction de distribution F présente une symétrie pondérée par rapport à θ si et seulement si il existe deux constantes $pr \in (0,1)$ et $\omega > 0$ telles que pour tout $x > 0$

$$prF(\theta - x^-) = (1 - pr)\{1 - F(\theta + \omega x)\}$$

où $F(y^-)$ désigne la limite gauche de F en y .

Ainsi, une symétrie pondérée est caractérisée par 3 paramètres à déterminer; à savoir θ le seuil de symétrie, pr la probabilité que les observations X_i dépassent le seuil θ et ω le poids de la symétrie. Il est à remarquer que la symétrie parfaite classique est un cas particulier de la symétrie pondérée avec $pr=1/2$ et $\omega=1$.

La symétrie pondérée est équivalente à un modèle fractionné en deux parties. Ce dernier peut être construit en assemblant les deux parties d'une même distribution, pondérées différemment. Il est possible de générer ces deux parties comme suit.

Soit Z une variable aléatoire positive et considérons une variable aléatoire X donnée par :

$$X = \begin{cases} X_1 = \theta - Z & \text{avec une probabilité } (1-pr) \\ X_2 = \theta + \omega Z & \text{avec une probabilité } pr \end{cases}$$

Avec $\theta \in \mathfrak{R}$, $pr \in (0,1)$ et $\omega > 0$ des paramètres déjà déterminés.

2.5.2. Estimation des paramètres de la symétrie pondérée

Nous allons estimer simultanément les paramètres inconnus de la symétrie pondérée en supposons que F satisfait la condition d'identification¹⁴. Abdous, Ghoudi et Rémillard (2003) proposent une méthode d'estimation de (θ, ω, pr) comme suit :

$$(\hat{\theta}_n, \hat{\omega}_n, \hat{pr}_n) = \arg \min_{\theta, \omega, pr} T_n(\theta, \omega, pr)$$

où :

$$T_n(\theta, \omega, pr) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \{prI(x_i \leq \theta) - (1-pr)I(x_i > \theta)\} \\ \times \{prI(x_j \leq \theta) - (1-pr)I(x_j > \theta)\} \min \left(\frac{|x_i - \theta|_{\omega}}{\omega}, \frac{|x_j - \theta|_{\omega}}{\omega} \right)$$

avec $|X - \theta|_{\omega} = |X - \theta| \{I(X > \theta) + \omega I(X \leq \theta)\}$

Ainsi, un programme de minimisation de la fonction T nous permet d'estimer simultanément les paramètres de la symétrie pondérée. Une fois les paramètres estimés, nous allons tester l'existence de la symétrie pondérée par rapport à θ .

2.5.3. Test de la symétrie pondérée

Nous allons nous baser sur l'étude de Abdous, Ghoudi et Rémillard (2003) pour mettre en place les tests adéquats permettant de valider ou non l'hypothèse de la symétrie pondérée. Il s'agit de calculer des statistiques Kolmogorov-Smirnov et Cramér-von Mises relatives à la symétrie pondérée. En effet, dans l'étude de Abdous, Ghoudi et Rémillard (2003), les auteurs proposent une approche plus robuste pour construire ces tests en utilisant des processus empiriques basés sur des pseudo-

¹⁴ Abdous, Ghoudi et Rémillard (2003) résument la condition d'identification de F comme suit : « Soit F une symétrie pondérée, continue et admet une densité avec des discontinuités finies et des minima ou des maxima locaux non nuls finis dans son domaine, alors F est identifiable.

observations tels que définis par Ghoudi et Rémillard (1998). Nous détaillons en annexe comment ces statistiques sont calculées.

Le tableau 2-3a présente les résultats de l'estimation des paramètres de la symétrie pondérée à partir du logarithme¹⁵ des pertes. Il est à remarquer que l'estimé du seuil de la symétrie θ est $\hat{\theta}=2,2892$. L'estimation de pr , la proportion des observations au-dessus du seuil est $\hat{pr}=0,2661$, alors que l'estimé du poids ω est $\hat{\omega}=0,6129$.

Pour valider l'hypothèse de la symétrie pondérée ainsi estimée, les données ont été aléatoirement réparties en deux échantillons. Un premier échantillon de pertes a servi pour estimer les paramètres de la symétrie pondérée. Le deuxième a été utilisé pour calculer les statistiques KS et CvM en traitant les paramètres $\hat{\theta}$, $\hat{\omega}$ et \hat{pr} comme étant connus. Nous répétons ce processus (génération d'échantillons aléatoires, estimation des paramètres et calcul des statistiques KS et CvM) au moins 10 000 fois. Nous calculons les probabilités que les statistiques calculées de Kolmogorov-Smirnov ainsi que de Cramér-von Mises soient inférieures aux valeurs critiques¹⁶ de ces statistiques ($KS_{critique}=2,241$ et $CvM_{critique}=1,68$, approximations respectives des 95% quantiles de KS_n et CvM_n pour un grand échantillon)

Le tableau 2-3a présente les résultats des tests. Les statistiques permettent de trancher en faveur d'une symétrie pondérée par rapport au seuil $\hat{\omega}$. En effet, nous remarquons que les valeurs des probabilités sont assez élevées montrant que les valeurs calculées des statistiques sont inférieures aux valeurs critiques dans 97% des cas pour le test KS et dans 87 % pour le test CvM. Ainsi, les deux tests mis en place permettent de valider l'hypothèse de l'existence d'une symétrie pondérée dans les données. Dans ce qui suit, nous allons retenir ce résultat et l'intégrer dans la modélisation des pertes opérationnelles avec la distribution GB2.

¹⁵ Nous prenons le logarithme des pertes, pour avoir des valeurs positives et négatives.

¹⁶ Se référer aux tableaux contenant les valeurs critiques pour différents quantiles dans l'article de Abdous, Ghoudi et Rémillard (2003)

2.5.4. Modèle GB2 fractionné

Rappelons que la distribution GB2 est la distribution la plus flexible. Son ajustement aux données a été rejeté certes, mais le graphique des quantiles montre que l'ajustement n'était pas adéquat principalement pour les grosses pertes. De plus, la GB2 présentait la fonction de vraisemblance la plus élevée parmi les distributions testées dans ce chapitre. Nous considérons donc cette distribution en introduisant éventuellement la notion de symétrie pondérée pour avoir un bon ajustement aux données. Comme les tests nous ont montré la validité de cette hypothèse, il est ainsi possible de construire un modèle combinant la distribution GB2 avec la symétrie pondérée. Nous mettons en place un modèle GB2 en deux parties ou GB2 fractionné.

Pour ce faire, nous définissons à partir des données une nouvelle variable Z telle que :

$$Z = \begin{cases} Z_1 = \hat{\theta} - X & \text{pour } X < \hat{\theta} \\ Z_2 = \frac{X - \hat{\theta}}{\hat{\omega}} & \text{pour } X \geq \hat{\theta}, \end{cases}$$

où, X est le logarithme des pertes.

La variable ainsi définie est donc positive et peut être modélisée par la distribution GB2. Nous estimons les paramètres de cette distribution avec la méthode du maximum de vraisemblance, à partir du vecteur de données $Z=[Z_1, Z_2]$, d'où apparaît le modèle en deux parties pondérées différemment ou le modèle GB2 fractionné. Le tableau 2-3b présente les résultats des estimés des paramètres de ce modèle.

Une fois les paramètres de la GB2 estimés, il y a lieu de tester le degré d'ajustement de ce modèle à la distribution empirique. Un test Kolmogorov-Smirnov avec *bootstrap* paramétrique permet de rejeter ou non l'hypothèse nulle stipulant un bon ajustement entre la distribution empirique et le modèle construit. Ainsi, au lieu de comparer la statistique calculée à la statistique tabulée comme c'est le cas pour le test

KS classique, nous proposons de comparer la statistique calculée à d'autres calculées à partir d'échantillons générés de la même distribution. Les paramètres de la symétrie pondérée ainsi que ceux de la GB2 vont être réestimés à partir de ces échantillons générés. Ce test est donc plus robuste que le test classique, et la valeur de la p-value calculée permettra de conclure quant à la qualité de l'ajustement. Nous présentons l'algorithme suivant pour accomplir le *bootstrap* paramétrique du test d'ajustement.

- 1- Définir la fonction de répartition F_Z de la variable transformée $Z=[Z_1, Z_2]$.
- 2- Calculer la statistique KS_0 , telle que définie précédemment, avec la fonction de répartition $F_Z(\hat{a}, \hat{b}, \hat{p}, \hat{q})$.
- 3- Générer un échantillon de variables W suivant la loi $GB2(\hat{a}, \hat{b}, \hat{p}, \hat{q})$, de même taille que l'échantillon de départ.
- 4- Définir la variable Y telle que :

$$Y = \begin{cases} Y_1 = \hat{\theta} - W & \text{avec une probabilité } (1 - \hat{p}) \\ Y_2 = \hat{\theta} + \hat{\omega}W & \text{avec une probabilité } \hat{p} \end{cases}$$

- 5- À partir du vecteur $Y=[Y_1, Y_2]$, estimer les paramètres de la symétrie pondérée $(\tilde{\theta}, \tilde{\omega}, \tilde{p})$.
- 6- Tester la validité de l'hypothèse de la symétrie pondérée pour l'échantillon généré, avec des tests KS et CvM tels que définis à la section précédente.
- 7- Définir la variable V telle que :

$$V = \begin{cases} V_1 = \tilde{\theta} - Y & \text{pour } Y < \tilde{\theta} \\ V_2 = \frac{Y - \tilde{\theta}}{\tilde{\omega}} & \text{pour } Y \geq \tilde{\theta} \end{cases}$$

- 8- À partir du vecteur $V=[V_1, V_2]$, estimer les paramètres $(\tilde{a}, \tilde{b}, \tilde{p}, \tilde{q})$ de la distribution GB2 avec la méthode du maximum de vraisemblance.
- 9- Déterminer la fonction de répartition F_V de la variable $V=[V_1, V_2]$.
- 10- Calculer la statistique KS_i avec la fonction de répartition $F_V(\tilde{a}, \tilde{b}, \tilde{p}, \tilde{q})$.
- 11- Comparer les statistiques KS_0 et KS_i . Si $KS_0 < KS_i$ alors un compteur j s'incrmente de 1.
- 12- Répéter les étapes 3 à 11 un grand nombre de fois N (ex. $N=10\ 000$).
- 13- Calculer la p-value comme étant $pv = \frac{j}{N}$.
- 14- Rejeter l'hypothèse nulle, si la p-value (pv) est plus petite que le niveau de confiance (5 %).

Force est de constater que nous estimons à chaque itération les paramètres de la symétrie pondérée ainsi que ceux de la GB2 d'une façon simultanée. Cette façon de faire permet de déterminer un résultat plus robuste du test KS classique. Ainsi, plus la p-value est élevée, mieux est l'ajustement. Nous allons également appliquer les tests formels d'ajustement Anderson Darling (AD) et Cramér-von Mises (CvM) avec la procédure *bootstrap* paramétrique. Ces tests sont mieux adaptés aux données appartenant à des distributions aux queues très épaisses. Les statistiques AD et CvM se calculent respectivement comme suit :

$$AD = -N - S$$

avec
$$S = \sum_{k=1}^N \frac{2k-1}{N} \left[\ln F(x_k, \hat{\theta}) + \ln(1 - F(x_{N+1-k}, \hat{\theta})) \right]$$

et

$$CvM = \frac{1}{12N} + \sum_{k=1}^N \left[\frac{2k-1}{2N} - F(x_k, \hat{\theta}) \right]^2$$

avec

F : la fonction cumulative de la distribution testée.

$\hat{\theta}$: le vecteur des paramètres estimés.

Comme pour le premier modèle, il est possible de tester au préalable l'ajustement du modèle GB2 fractionné par des tests graphiques. En effet, un graphique des quantiles de ce modèle versus la distribution empirique donnera un premier aperçu du degré d'ajustement. De même, le graphique des distributions de répartition permet d'avoir une idée sur la qualité d'ajustement. Le graphique 2-6a montre en effet, que la courbe faisant le lien entre les quantiles (empiriques versus paramétrique) est clairement linéaire et très proche de la droite de 45 degrés. De plus, les deux distributions de répartition sont très proches l'une de l'autre (graphique 2-6b). Ainsi, en se basant sur ces tests, le modèle s'ajuste bien aux données. Il faut toutefois confirmer ce constat avec les tests formels d'ajustement tels que décrits précédemment.

Les résultats présentés au tableau 2-3b montrent que les valeurs des p-value (0,65 pour le test KS, 0,81 pour le test AD et 0,77 pour le test CvM) des tests d'ajustement calculées avec *bootstrap* paramétrique permettent de ne pas rejeter l'ajustement du modèle GB2 fractionné aux données empiriques. Ainsi, grâce à la notion de symétrie pondérée, nous avons construit un modèle qui décrit bien le comportement des pertes, alors que nous n'avons pas réussi à avoir un ajustement aux données avec les distributions paramétriques, même avec celle à quatre paramètres la GB2.

2.6. Conclusion

L'objectif de ce chapitre était de proposer un modèle qui permet de mieux décrire le comportement des pertes opérationnelles et ce, dans le but de mieux calculer un capital opérationnel. En effet, plusieurs études ont montré l'importance du choix de la distribution. Une mauvaise spécification de cette dernière entraîne inévitablement une sous ou une sur évaluation du capital.

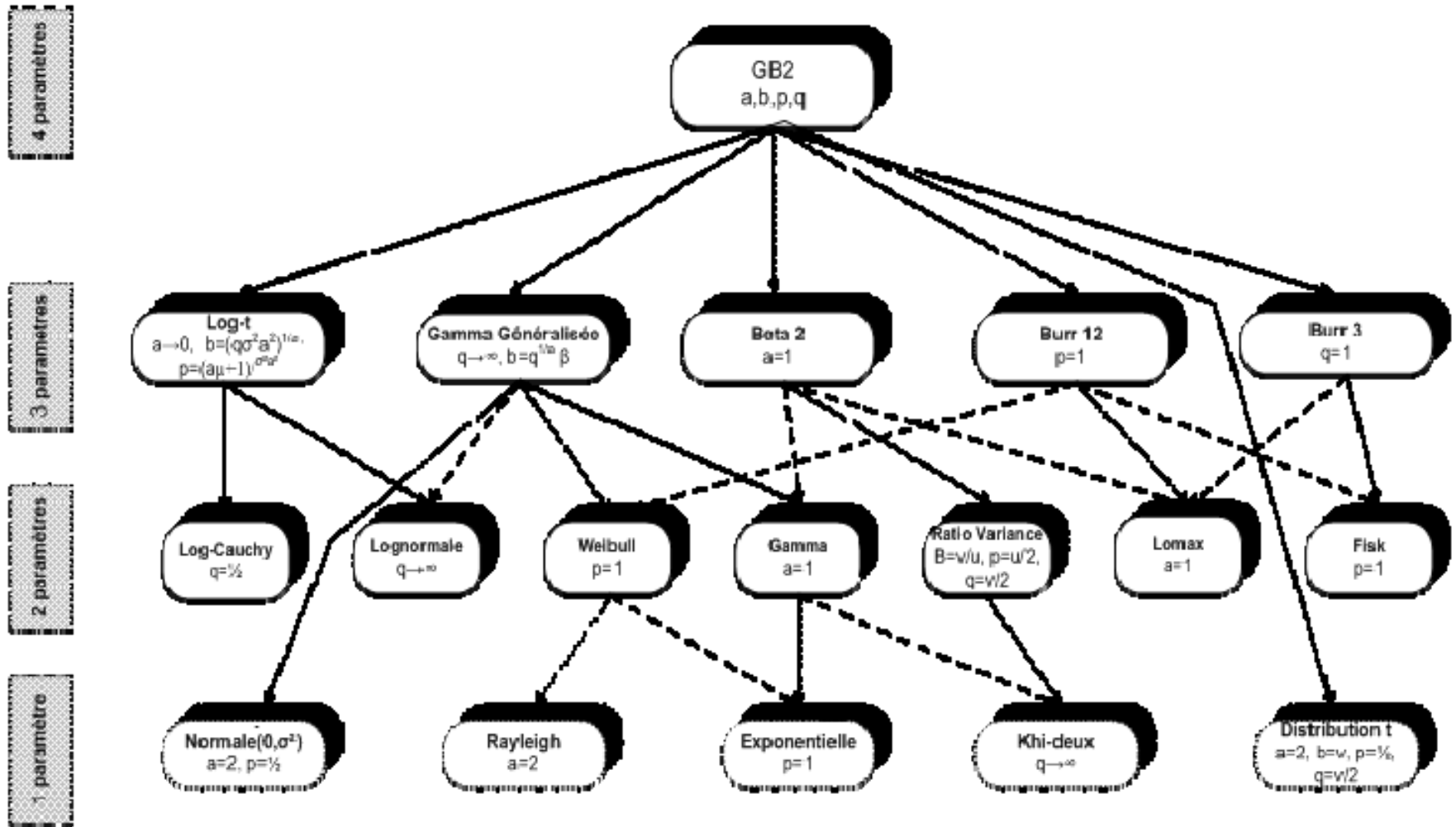
Dans un premier temps, nous avons testé l'ajustement aux données, de trois distributions usuelles (exponentielle, lognormale et Weibull) ainsi qu'une famille de distribution à quatre paramètres la GB2. Les résultats montrent qu'aucune des distributions paramétriques testées ne s'ajustaient aux données de pertes externes mises à l'échelle d'une banque canadienne. Malgré sa grande flexibilité, la GB2 n'a pu s'ajuster aux données et plus particulièrement à la queue de la distribution empirique.

Par ailleurs, l'introduction de la notion de symétrie pondérée au modèle précédent vient offrir un excellent ajustement. Étant démontrée que les données ne sont pas symétriques, nous avons testé l'existence d'une symétrie pondérée et les résultats ont effectivement montré la validité de cette hypothèse. Nous avons ainsi construit un modèle original GB2 fractionné, en combinant le modèle GB2 et la symétrie pondérée ensemble.

Nous avons également proposé une manière originale de tester le bon ajustement du modèle aux données grâce aux tests Kolmogorov-Smirnov, Anderson Darling et Cramér-von Mises avec la procédure *bootstrap* paramétrique en estimant d'une façon simultanée, lors des tests, les paramètres de la symétrie pondérée et ceux de la GB2. Les résultats ont montré que l'introduction de la symétrie pondérée a amélioré significativement le degré d'ajustement de ce modèle aux données.

Ainsi, en améliorant la qualité de l'ajustement, nous augmentons notre confiance dans le calcul du capital. En effet, plus la distribution choisie décrit la tendance et le comportement des pertes, plus la valeur à risque calculée après agrégation et simulation nous donne une meilleure idée sur l'exposition réelle d'une banque face au risque opérationnel.

Graphique 2-1: La GB2 et ses cas particuliers



Source: Cummins, J. D., G. Dionne, J. B. McDonald and B. M. Pritchett (1990)

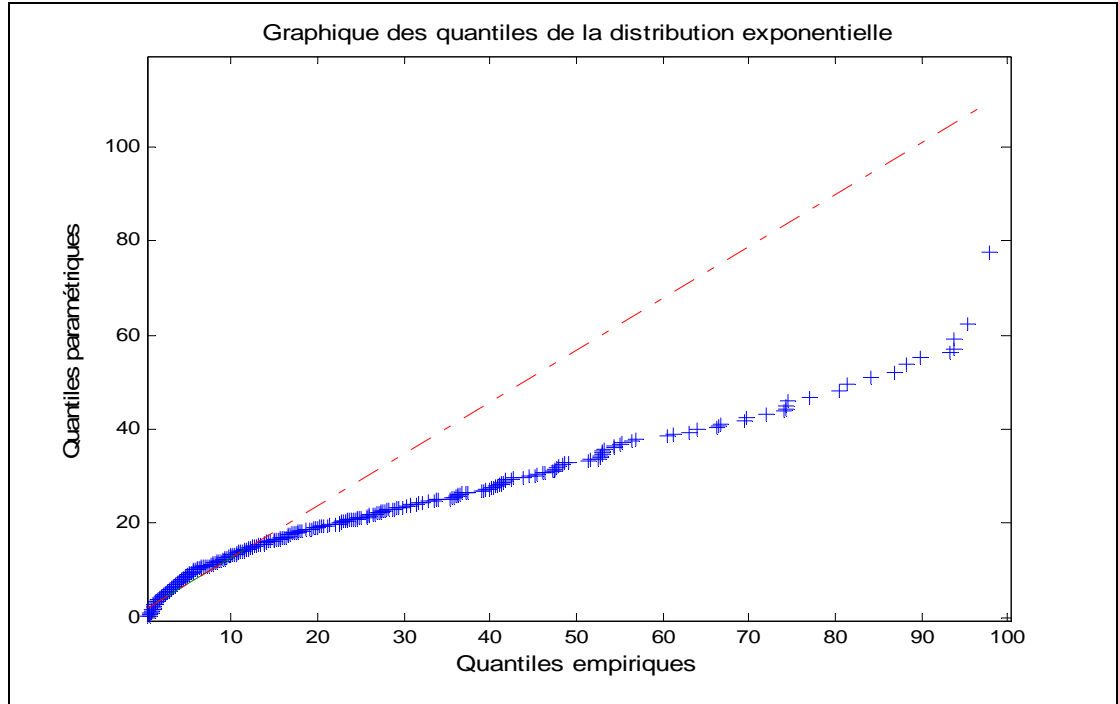
Tableau 2-2: Statistiques descriptives sur les données.

Ce tableau présente des statistiques descriptives des données de pertes mises à l'échelle d'une banque canadienne. 5% des données sont considérées comme des outliers et sont écartées de la base afin d'avoir une estimation robuste des paramètres. (Chernobai et al., 2006)

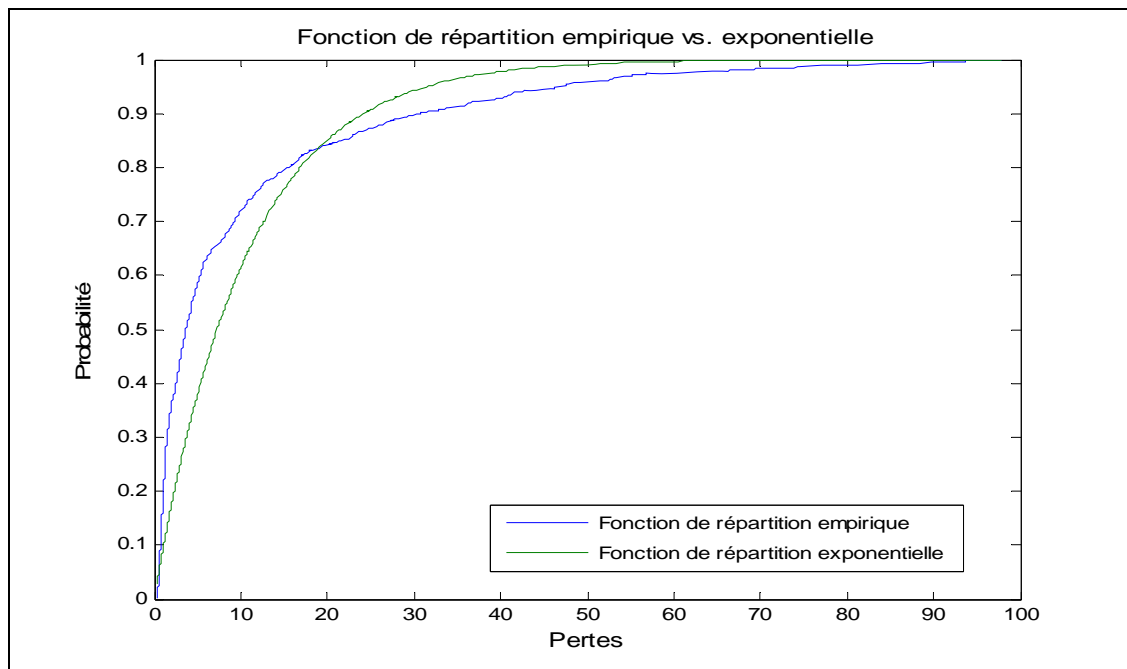
Moyenne (M\$)	10,490
Médiane (M\$)	3,670
Mode (M\$)	0,926
Écart-type (M\$)	16,217
Variance de l'échantillon	262,984
Coefficient d'aplatissement	7,727
Coefficient d'asymétrie	2,653
Minimum (M\$)	0,304
Maximum (M\$)	97,888
Somme (M\$)	10531,632
Nombre d'observations	1004

Graphique 2-2 : Tests graphiques de la distribution exponentielle

Graphique 2-2a présente le graphique des quantiles empiriques versus les quantiles de l'exponentielle.

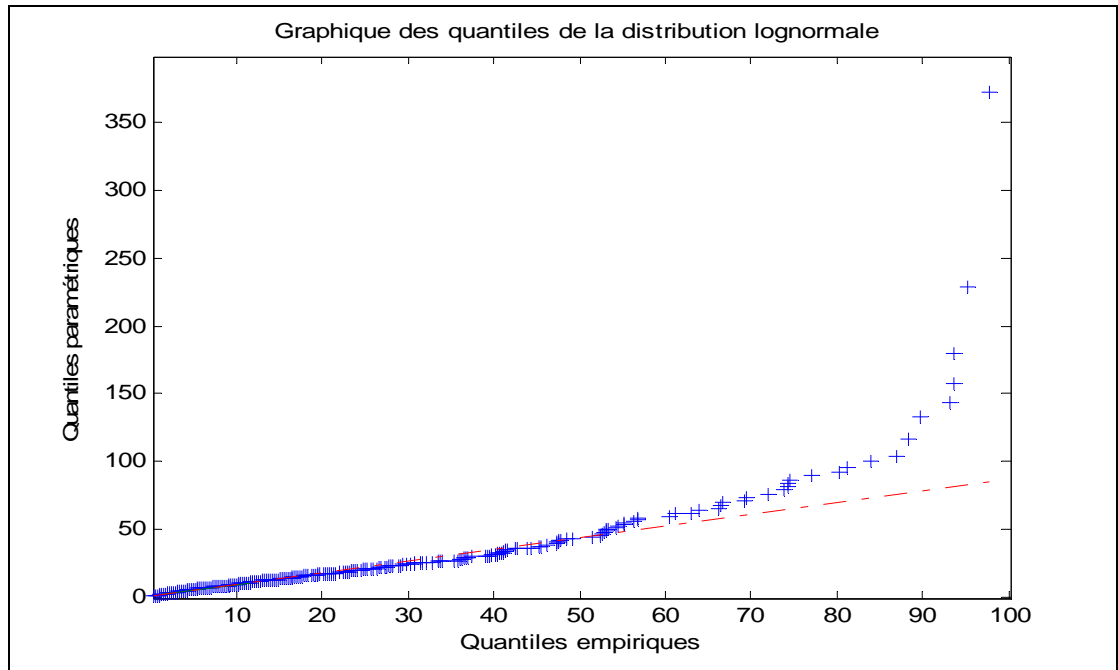


Graphique 2-2b présente le graphique des fonctions de répartition empirique versus exponentielle.

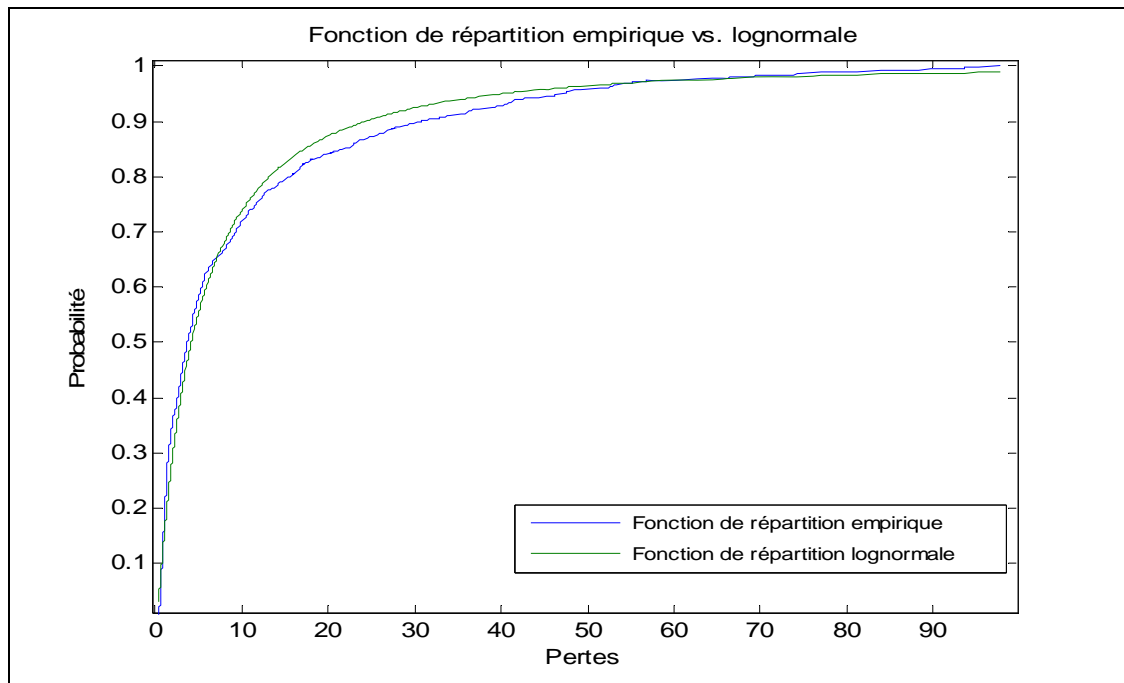


Graphique 2-3: Tests graphiques de la distribution lognormale.

Graphique 2-3a présente le graphique des quantiles empiriques versus les quantiles de la lognormale.

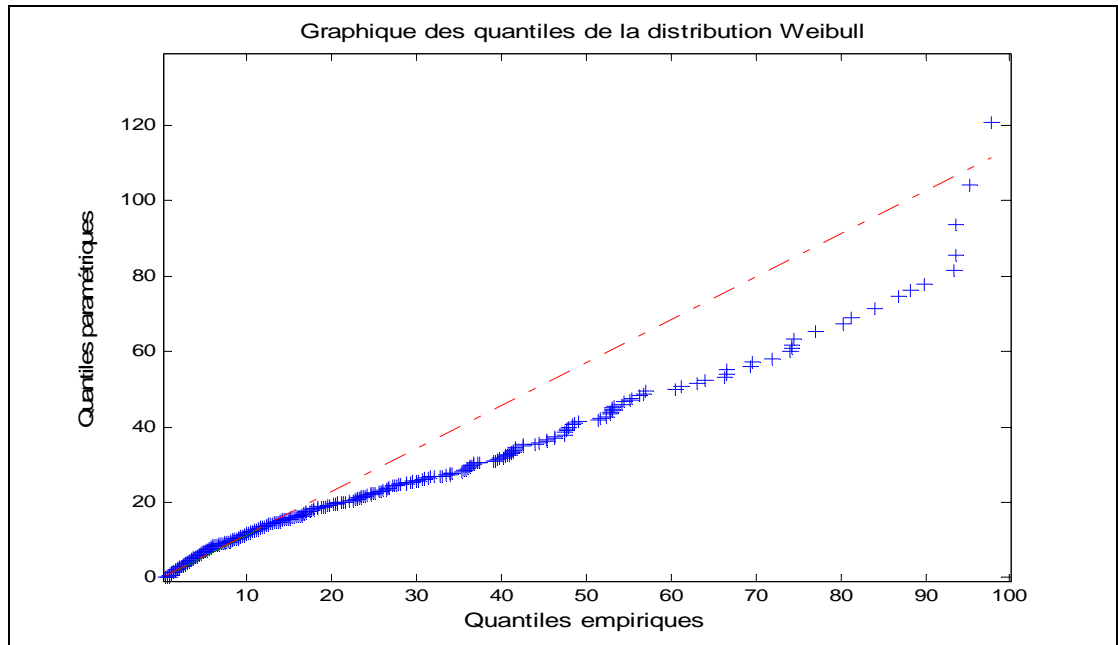


Graphique 2-3b présente le graphique des fonctions de répartition empirique versus lognormale..

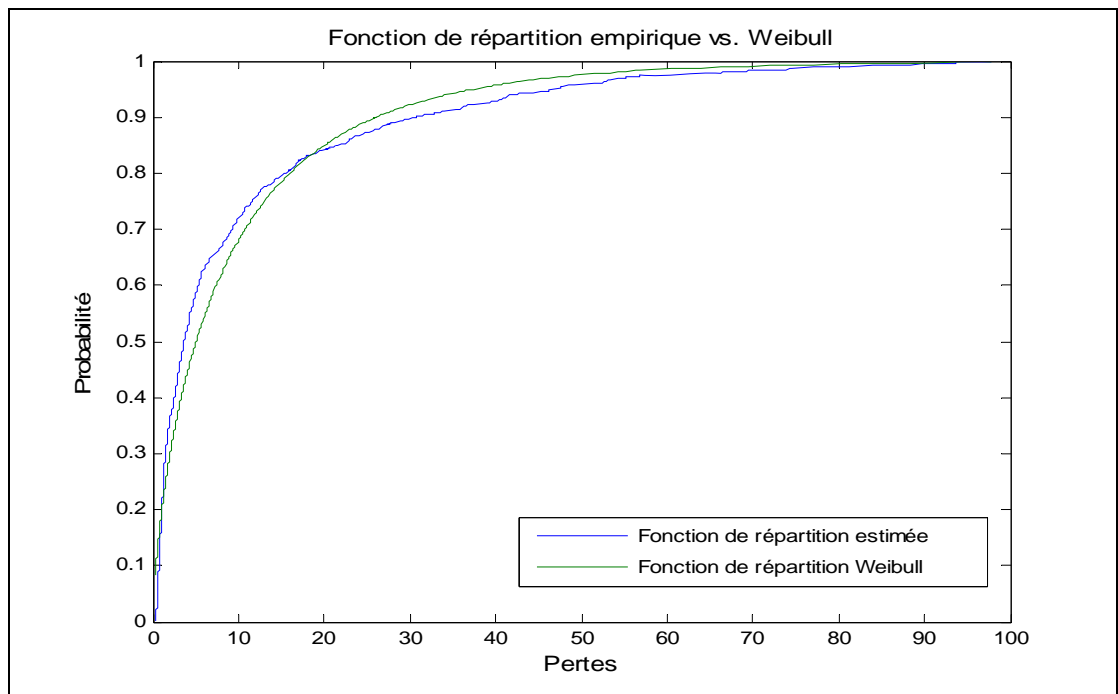


Graphique 2-4 : Tests graphiques de la distribution Weibull.

Graphique 2-4a présente le graphique des quantiles empiriques versus les quantiles de la Weibull.

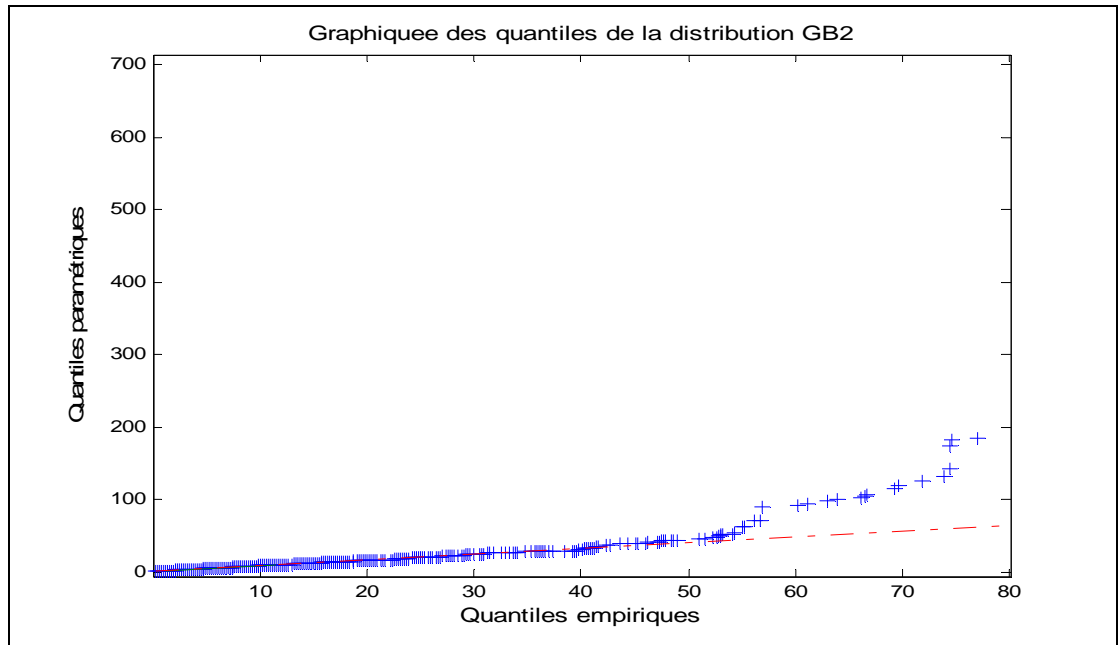


Graphique 2-4b présente le graphique des fonctions de répartition empirique versus Weibull.



Graphique 2-5 : Tests graphiques de la distribution GB2.

Graphique 2-5a présente le graphique des quantiles empiriques versus les quantiles de la GB2.



Graphique 2-5b présente le graphique des fonctions de répartition empirique versus GB2.

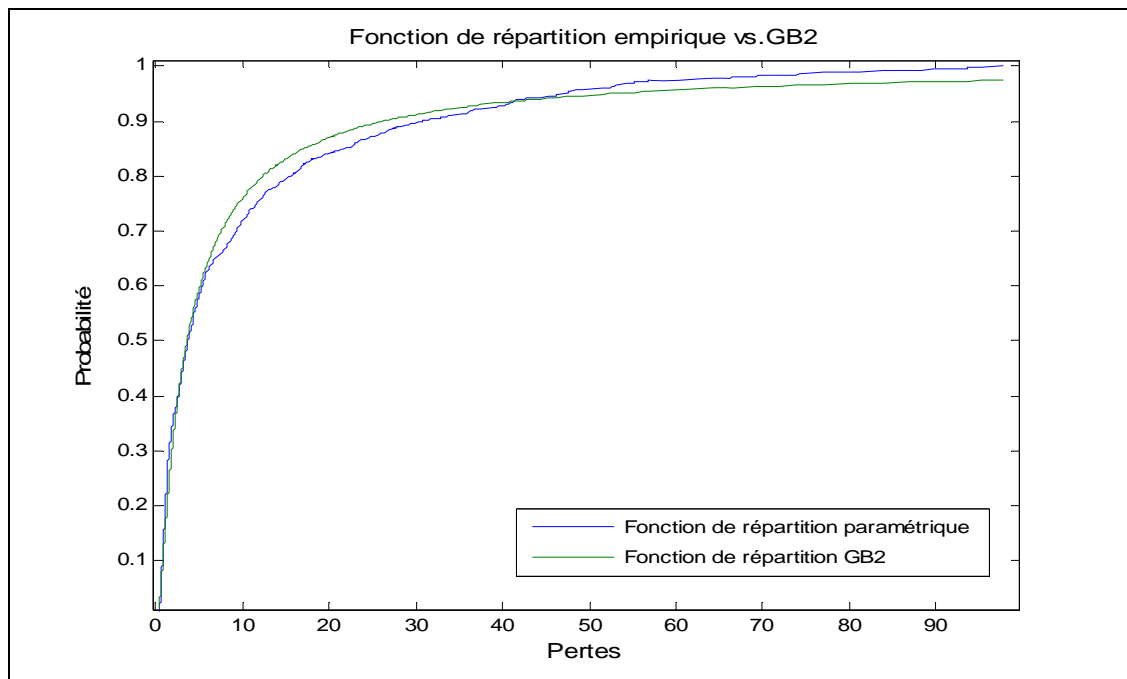


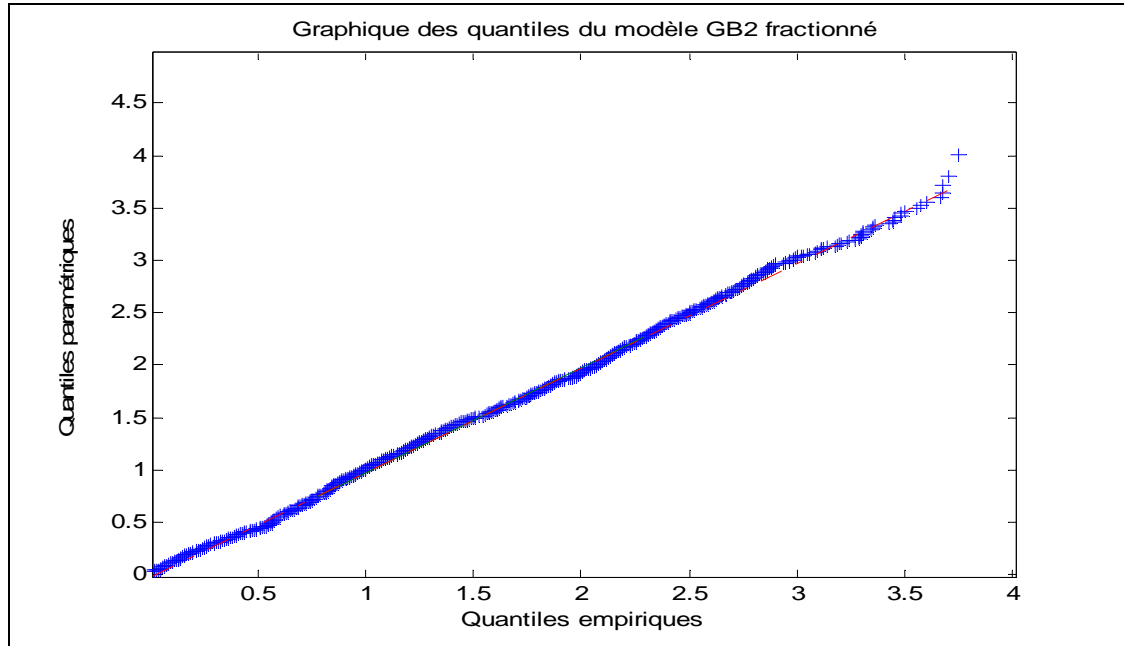
Tableau 2-3 : Résultats de l'estimation des paramètres

Les paramètres estimés par maximum de vraisemblance sont présentés dans le tableau ci-dessous. L'avant dernière ligne indique la valeur du log de vraisemblance pour les quatre distributions testées. La dernière ligne montre les valeurs du p-value du test de Kolmogorov-Smirnov avec *bootstrap* paramétrique.

	exponentielle	lognormale	Weibull	GB2
Paramètres estimés	10,4897	1,4161 1,3799	8,3506 0,7321	0,5440 0,0151 37,6630 2,2349
Log de vraisemblance	-3363,7929	-3169,123	-3257,978	-3156,695
<i>p-value KS</i>	0	0	0	0,004

Graphique 2-6 : Tests graphiques du modèle GB2 fractionné.

Graphique 2-6a présente le graphique des quantiles empiriques versus les quantiles du modèle GB2 fractionné.



Graphique 2-6b présente le graphique des fonctions de répartition empirique versus le modèle GB2 fractionné.

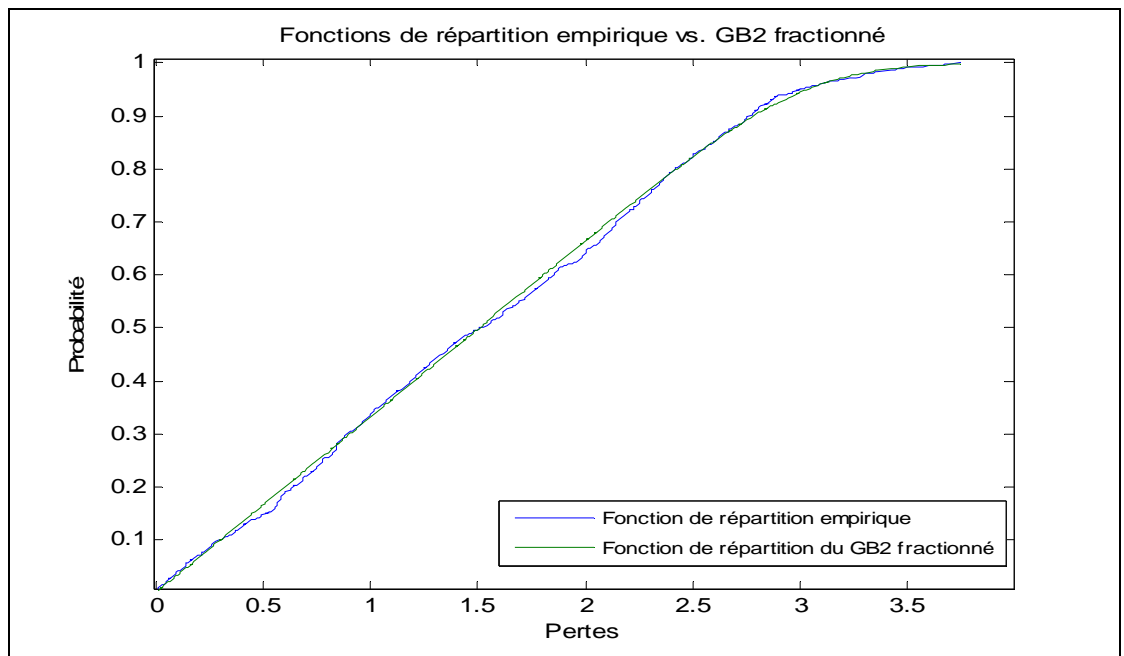


Tableau 2-4 : Résultats de l'estimation des paramètres du modèle GB2 fractionné.

Le tableau 2-4a présente les estimés des paramètres de la symétrie pondérée. Les deux dernières lignes présentent les résultats des tests d'ajustement. Les chiffres en italiques constituent la probabilité que les statistiques calculées soient inférieures aux valeurs critiques des tests (95^{ème} quantile)

Symétrie pondérée	Valeurs
Le seuil de symétrie $\hat{\theta}$	2,2892
La probabilité de dépasser le seuil \hat{p}_r	0,2661
Le poids de la symétrie $\hat{\omega}$	0,6129
Test Kolmogorov-Smirnov	1,23 <i>0,97</i>
Test Cramér-von Mises	0,26 <i>0,87</i>

Le tableau 2-4b présente les paramètres estimés du modèle GB2 en tenant compte de la symétrie pondérée, ainsi que la p-value du test de Kolmogorov-Smirnov accompli avec *bootstrap* paramétrique.

GB2	Valeurs
\hat{a}	10,2202
\hat{b}	3,3724
\hat{p}	0,0981
\hat{q}	2,3962
Log de vraisemblance	-1228,7
p-value Kolmogorov-Smirnov	0,67
p-value Anderson Darling	0,81
p-value Cramér-von Mises	0,77

Annexe 2

Abdous, Ghoudi et Rémillard (2003) ont développé des tests non paramétriques pour valider ou non l'hypothèse de la symétrie pondérée. Nous présentons dans ce qui suit comment les tests Kolmogorov-Smirnov et Cramér-von Mises sont calculés, **dans les cas où les paramètres de la symétrie pondérée sont connus.**

Ils ont défini l'inférence qui sera basée sur des fonctions du processus de la symétrie pondérée données par :

$$K_n(t, p) = \frac{\sqrt{n}[K_n(t) - E\{K_n(t)\}]}{\sqrt{p(1-p)}}, \quad t \in [-1, 1],$$

où

$$K_n(t) = \frac{1}{n} \sum_{i=1}^n I(e_{n,i} \leq t)$$

Les pseudo-observations $e_{n,i}$ sont définies par :

$$e_{n,i} = \frac{1}{n} \text{sign}(X_i - \theta) \text{rang}(X_i - \theta |_{\omega}), \quad 1 \leq i \leq n$$

Si A_1, \dots, A_n correspondent aux anti-rangs, c'est-à-dire, $A_i = k$ si et seulement si $\text{rang}(X_i - \theta |_{\omega}) = i$, alors :

$$K_n(t) = \frac{1}{n} \sum I(i \varepsilon_{n+1-i} \leq nt)$$

où $\varepsilon_{n+1-i} = \text{sign}(X_{A_i} - \theta)$.

Soit

$$\delta_i = I(\varepsilon_i = -1) = I(X_{A_{n+1-i}} < \theta)$$

Nous pouvons vérifier facilement que :

$$K_n(t, p) = \begin{cases} W_n\left(1 + \frac{1}{n} + \frac{[nt]}{n}\right), & -1 \leq t \leq 0, \\ W_n\left(1 - \frac{[nt]}{n}\right), & 0 \leq t \leq 1, \end{cases}$$

où $[x]$ est la partie entière de x .

avec

$$W_n(u) = \frac{1}{\sqrt{np(1-p)}} \sum_{i=1}^{[nu]} \{\delta_i - (1-p)\} = \frac{S_{[nu]}}{\sqrt{np(1-p)}}$$

où $S_k = \sum_{i=1}^k \{\delta_i - (1-p)\}$ est la somme partielle et $S_0 = 0$.

Soit $\theta \in \mathfrak{R}$, $p \in (0,1)$ et $\omega > 1$ sont constants, on suppose que la distribution F présente une symétrie pondérée (p, ω) par rapport à θ , alors

(i) La statistique Kolmogorov-Smirnov

$$KS_n = \sup_{-1 \leq t \leq 1} |K_n(t, p)| = \frac{1}{\sqrt{np(1-p)}} \max_{1 \leq j \leq n} |S_j|$$

converge en loi à $\sup_{0 \leq t \leq 1} |W(t)|$, quand $n \rightarrow \infty$

(ii) La statistique Cramér-von Mises

$$CvM_n = \frac{1}{2} \int_{-1}^1 \{K_n(t, p)\}^2 dt = \frac{1}{n^2 p(1-p)} \sum_{j=1}^n S_j^2$$

converge en loi à $\int_0^1 \{W(u)\}^2 du$, quand $n \rightarrow \infty$

3. DÉTERMINATION DE LA VALEUR À RISQUE OPÉRATIONNEL D'UNE BANQUE CANADIENNE

3.1. Introduction

La valeur à risque (VaR) est un concept assez récent et de plus en plus utilisé en finance (Jorion, 2000). Cette notion a été, d'ailleurs, déjà appliquée depuis des décennies dans le domaine de l'assurance sous le nom de MPY¹⁷ (Cummins et Freifelder, 1978). Bien que cette mesure ne soit pas toujours jugée cohérente (Artzner et al., 1999), nous parlons de plus en plus de VaR de marché, de VaR de crédit et récemment de VaR opérationnel. Cette mesure est effectivement facile à calculer et à implanter en pratique. Nous portons notre attention à la VaR opérationnel et nous procédons à toutes les étapes nécessaires à son calcul dans ce chapitre.

Le risque opérationnel présente plusieurs particularités par rapport aux risques de marché et de crédit. En effet, la structure des pertes, peu fréquentes et à amplitudes potentiellement très élevées, rend sa modélisation plus difficile et parfois complexe. Certaines études émettent des hypothèses assez fortes et peu réalistes qui vont simplifier la méthodologie certes, mais perdent énormément en termes de précision et de rigueur quant au calcul de la VaR ou du capital opérationnel.

L'objectif de ce chapitre est de développer une mesure précise et rigoureuse de la VaR opérationnel d'une banque. Nous comparons notre modèle au modèle standard, fortement utilisé en pratique, construit à partir des distributions lognormale et Poisson et qui ne tient pas compte des données au-dessous du seuil de troncature. Nous montrons la supériorité de notre modèle à fournir des résultats plus réalistes contrairement au modèle standard qui sous-évalue énormément l'exposition au risque d'une banque.

¹⁷ MPY : maximum probable yearly aggregate loss

Ce chapitre met l'accent sur le bon ajustement des distributions paramétriques à l'ensemble des données et vise à montrer l'importance du choix des distributions de sévérité et de fréquence des pertes. Ces éléments sont en effet négligés dans le modèle standard. De plus, notre mesure de risque fait intervenir des pertes opérationnelles externes mises à l'échelle d'une banque. Ceci permet en effet de tenir compte de certaines pertes extrêmes éventuelles n'ayant pas encore été subies. Nous analysons d'ailleurs l'impact de l'intégration des données de pertes extrêmes sur le calcul de VaR.

Ce chapitre est structuré comme suit. La section suivante présente un bref aperçu de la littérature récente relative à la modélisation du risque opérationnel. Ensuite, une description des données utilisées dans notre analyse est présentée à la section 3. Nous présentons le modèle Loss Distribution Approach (LDA) ainsi que ses hypothèses à la quatrième section. La section 5 développe le modèle de sévérité alors que le modèle de fréquence est présenté à la sixième section. Le calcul de la VaR par type de risque est détaillé à la section 7. Enfin, la dernière section résume les principaux résultats et présente les extensions et les avenues de recherche.

3.2. Revue de la littérature

Plusieurs méthodes de calcul du capital risque opérationnel ont été proposées par les autorités réglementaires. La méthode de mesure avancée est la méthode la plus représentative du niveau d'exposition de risque opérationnel d'une institution financière. De nombreuses études ont émergé développant des méthodologies quantitatives et des outils qui peuvent être appliqués dans l'approche de mesure avancée (Cruz, 2002, Alexander, 2003, King, 2001). La méthode LDA est l'approche de mesure avancée la plus populaire. Cette approche s'inspire énormément de la méthode actuarielle utilisée dans le domaine de l'assurance pour la modélisation des pertes (Cummins et Freifelder, 1978, Cummins et Wiltbank, 1983, Cummins et al, 1990, Panjer et Willmot, 1992). Klugman, Panjer et Willmot (1998) ont bien développé les différentes étapes de la méthode. Cependant, la modélisation des pertes opérationnelles présente certaines divergences par rapport à ces modèles actuariels de pertes.

Tout d'abord, les pertes opérationnelles sont, dans la plupart des cas, collectées à partir d'un certain seuil (1 000 \$, 10 000 \$, 15 000 \$, etc.). Plusieurs études ont ignoré ces pertes et n'ont donc pas tenu compte du seuil de troncature (Dutta et Perry, 2006, Böcker et Klüppelberg, 2005, Fontnouvelle, Rosengren et Jordan, 2004). Cette approche telle que définie par Chernobai et al. (2006) consiste à ajuster les distributions non conditionnelles de sévérité et de fréquence aux données de pertes au-dessus du seuil de troncature. Elle suppose en effet que seules les queues des distributions agrégées entrent en compte dans le calcul de la VaR.

Il existe des études qui ne considèrent que la modélisation de la queue de la distribution agrégée. Elles ne traitent que les pertes extrêmes en utilisant la théorie des valeurs extrêmes (TVE) (Embrechts et al., 1997, Ebnother, Vanini, McNeil et Antolinez-Fehr, 2001). En revanche, la structure des données opérationnelles ne satisfait pas nécessairement les hypothèses standards de la modélisation avec la TVE. Ceci est dû principalement au nombre restreint des pertes extrêmes peu fréquentes (Embrechts et al., 2003, Moscadelli, 2004).

Les pertes au-dessous du seuil de collecte peuvent avoir un impact significatif sur le niveau du capital surtout lorsque le seuil est assez élevé et la fréquence de ces pertes est importante (Frachot, Moudoulaud et Roncalli, 2003). Il est à remarquer que la modélisation des données tronquées rend la méthodologie d'estimation plus complexe surtout pour les distributions à plusieurs paramètres. Les travaux de Baud et al. (2002), Frachot et al. (2003), de Fontnouvelle et al. (2003), Chapelle et al. (2004) et Chernobai et al. (2005a, 2005c) ont montré l'importance de la prise en considération du seuil de troncature dans la modélisation des pertes opérationnelles. D'ailleurs, des tests d'ajustement pour des échantillons tronqués à gauche ont été développés par Chernobai, Rachev et Fabozzi (2005b). Nous allons considérer dans notre étude le seuil de collecte et modéliser les pertes opérationnelles en conséquence. De plus, nous considérons l'estimation des queues en découpant la distribution en corps et queue d'une manière semblable à celle faite par Chapelle, Crama, Hübner et Peters (2004).

Par ailleurs, le choix des distributions de sévérité est également très important. En effet, la mauvaise spécification de cette distribution peut entraîner une sous ou une sur évaluation du capital comme l'a montré l'étude de Chernobai et al. (2005c). Il est à noter que la distribution la plus utilisée pour l'estimation des montants de pertes est la lognormale vu la facilité de son implantation et l'épaisseur relative de sa queue. D'ailleurs, les travaux de Frachot et al. (2003), Chernobai et al. (2005a) et Bee (2006) se sont basés uniquement sur cette distribution pour modéliser les montants de pertes opérationnelles sans même s'assurer de la qualité de l'ajustement de cette distribution aux données.

Nous allons tester quatre distributions à savoir l'exponentielle, la lognormale, la Weibull et la distribution GB2 à quatre paramètres pour le corps de la distribution, et la distribution Pareto pour la queue. Dutta et Perry (2006) ont testé également plusieurs distributions dont la GB2 et la distribution g et h à quatre paramètres. Ils ont montré que la distribution g et h constitue le meilleur modèle pour la sévérité. Par contre, ils ont trouvé que les paramètres estimés de la GB2 n'étaient pas raisonnables tout en soulignant la complexité de l'estimation de ces paramètres et la difficulté de générer des nombres aléatoires selon cette distribution. Nous allons présenter dans cette étude les paramètres estimés de la GB2 et nous allons montrer que la GB2 est la meilleure distribution parmi celles que nous testons.

Quant à l'occurrence des pertes, peu de recherches ont été faites pour raffiner le degré d'ajustement des distributions de fréquence aux données. La plupart des travaux (Dutta et Perry, 2006, Chapelle et al., 2004, Chernobai et al., 2005a, de Fontnouvelle et al., 2003, Frachot et al., 2003) supposent que les fréquences sont modélisées selon la distribution Poisson. Böcker et Klüppelberg (2005) ont développé une approximation de la VaR où seule l'espérance du nombre des pertes est prise en compte dans la formule. Ainsi, la surdispersion des fréquences des pertes n'a pas d'impact sur la formule de la VaR. Cependant, la distribution Poisson est caractérisée par l'équidispersion c'est-à-dire l'égalité entre la moyenne et la variance, ce qui est rarement le cas en risque opérationnel. Nous allons tester dans la présente étude les modèles Poisson et binomial négatif pour un meilleur ajustement des fréquences. Le

test sera fait par un test χ^2 avec *bootstrap* paramétrique que nous allons mettre en place.

D'autre part, nous allons corriger les paramètres des distributions de fréquence pour tenir compte du fait que seul le nombre des pertes supérieures au seuil est connu. À notre connaissance, seule l'étude de Frachot, Moudoulaud et Roncalli (2003) a déterminé de façon appropriée le paramètre de la distribution Poisson. Aucun travail similaire n'a été fait pour la distribution binomiale négative. Nous ajustons donc les paramètres tout en utilisant la distribution qui s'ajuste le mieux aux données.

Chernobai, Menn, Rachev, Truck et Moscadelli (2006) ont présenté les différentes approches de traitement des données tronquées. L'approche la plus utilisée en pratique, consiste à ignorer les observations manquantes et à traiter l'échantillon tronqué comme étant complet. La deuxième approche consiste à estimer la distribution non conditionnelle de sévérité, mais corriger les paramètres de la distribution de fréquence seulement pour tenir compte du nombre de pertes au-dessous du seuil. La troisième approche consiste à tenir compte de la troncature seulement dans la distribution de sévérité. Le biais lié à cette approche est moindre, mais la méthode contient encore des lacunes puisque les paramètres de la distribution de fréquence n'ont pas été corrigés. Enfin, la dernière approche consiste à ajuster la distribution conditionnelle de sévérité et à corriger les paramètres de la distribution de fréquence. Cette méthode est statistiquement plus adéquate. Elle réduit effectivement le biais relatif au calcul du capital par rapport aux autres approches. Nous allons donc suivre cette approche dans ce chapitre.

Une fois les modèles de sévérité et de fréquence déterminés, il y a lieu de les agréger pour calculer la valeur à risque. Les modèles d'agrégation des pertes ont été déjà développés dans le domaine de l'assurance pour le calcul du montant de toutes les réclamations payées pendant une période déterminée pour le portefeuille d'un assureur bien défini. Klugman, Panjer et Willmot (1998) ont bien détaillé les méthodes d'agrégation des pertes. Pour la plupart des modèles de sévérité et de fréquence retenus, la distribution agrégée ne peut être obtenue que numériquement.

Cependant pour certaines combinaisons de distributions, telles que l'exponentielle-géométrique ou l'exponentielle-binomiale négative, un résultat analytique est disponible ; ceci réduit considérablement les problèmes liés au calcul. Il est possible d'utiliser également la méthode récursive ou les méthodes d'inversion pour évaluer rapidement la distribution agrégée (Klugman et al., 1998, Embrechts et al., 2003). Ces méthodes analytiques sont fondées sur les hypothèses d'indépendance entre la sévérité et la fréquence et de montants de pertes identiquement distribués.

D'autre part, il existe la méthode par simulation pour agréger les distributions. Contrairement aux autres méthodes, elle n'exige pas les hypothèses citées précédemment pour le calcul de la perte agrégée (Ross, 2001). Cette méthode consiste à générer des variables pseudo-aléatoires de sévérité et de fréquence selon les modèles déjà connus et à calculer la perte agrégée. La fonction cumulative empirique des pertes annuelles peut être estimée par la distribution cumulative de l'échantillon pseudo aléatoire de pertes agrégées. Cette technique est très utilisée dans le contexte du risque opérationnel pour agréger les pertes étant donné le degré de précision des résultats (Cruz, 2002, Frachot, Frachot et Roncalli, 2001). Nous allons donc utiliser cette technique d'agrégation dans notre étude.

Une fois que les pertes ont été agrégées grâce à l'une des méthodes définies précédemment, la VaR, dans le contexte opérationnel, est calculée comme étant le 99.9^{ème} centile de la distribution. Böcker et Klüppelberg (2005) développent une approximation sous forme fermée de la VaR opérationnel dans le cas où la sévérité est modélisée selon une distribution Pareto.

Nous rappelons que les données de pertes internes sont insuffisantes pour modéliser la VaR à un niveau de confiance de 99,9 %, puisqu'elles ne reflètent pas l'ampleur réelle des pertes. Les autorités réglementaires ont exigé l'utilisation des données externes pour compléter les données internes dans le cas où ces dernières ne sont pas exhaustives. Nous combinons donc les données internes et externes avant de calculer la perte non anticipée. Peu de recherches ont développé la méthode LDA en tenant compte aussi bien des données internes que des données externes (Chapelle, Crama, Hübner et Peters, 2004).

3.3. Les données

3.3.1. Description des données

Nous allons appliquer les différentes méthodes d'ajustement et d'estimation des paramètres sur les données de pertes d'une banque canadienne jumelées à des données externes. Nous modélisons les pertes selon leur type de risque. Les différents types de risque définis par les autorités réglementaires sont :

- actifs tangibles, corporels et endommagés (ATCE)
- fraude externe (FE)
- clients, produits et pratiques commerciales (CPPC)
- fraude interne (FI)
- gestion de l'exécution et de la livraison des processus (GELP)
- emploi, pratiques et sécurité environnementale (EPSE)
- perturbation des affaires et défaillance des systèmes (PADS)

Nous ne faisons pas un découpage par unité d'affaires tel que recommandé par les autorités réglementaires dans le cadre de l'accord de Bâle II, et ce, pour garder un nombre suffisant de données dans les cellules à l'étude. De plus la modélisation des pertes par type de risque présente l'avantage que les processus de pertes sont homogènes. Une exception faite pour le type de risque *perturbation des affaires et défaillance des systèmes* où la quantité de données de pertes relatives à ce risque n'est pas importante même à l'échelle de la banque. Il sera donc exclu de l'étude.

Les données de pertes couvrent une période de 3 ans du 1^{er} novembre 2001 jusqu'au 31 octobre 2004. Le seuil de collecte étant fixé à s pour tous les types de risque excepté pour les *fraude externe* et *fraude interne* pour lesquelles nous disposons de toutes les pertes ayant lieu à la banque¹⁸.

Nous allons, dans la prochaine section, chercher la distribution qui s'ajuste le mieux aux montants de pertes pour les six types de risque. Vu les différences de nombre et

¹⁸ s est un seuil de collecte connu. Pour préserver la confidentialité de la banque à l'étude nous le désignons par s tout le long de ce chapitre.

de caractéristiques entre les 6 échantillons (un échantillon par type de risque), différentes méthodes seront implantées.

Il est à mentionner que nous avons exclu toutes les pertes internes de plus d'un million de dollars américains¹⁹ de notre échantillon. En fait, ces pertes seront ajoutées à la base externe qui contient des pertes extrêmes de plus d'un million de dollars américains. Ainsi, nous avons deux échantillons indépendants, le premier est constitué de pertes de moins d'un million alors que le second contient des pertes externes mises à l'échelle de la banque canadienne en question, de plus d'un million. Il est donc plus facile de combiner ces deux échantillons lors du calcul de la VaR opérationnel par type de risque.

3.3.2. Statistiques descriptives sur les montants de pertes

Le nombre de données par type de risque est très variable. En effet, le tableau 3-1a montre que nous disposons de seulement 53 événements de type *actifs tangibles, corporels et endommagés* en dessus de s dans toute la banque sur les la période de 3 ans, alors qu'il y en a 22 178 événements de fraudes externes, dont 13 149 au dessous de s , principalement liés à des fraudes de cartes (cartes de crédit et cartes débit).

Quant à la moyenne des pertes par événement, nous notons que les moyennes les plus élevées sont liées aux types de risque *clients, produits et pratiques commerciales et emploi, pratiques et sécurité environnementale*. En effet, ces pertes sont généralement dues à des poursuites faites respectivement par des clients et des employés. Tandis que la perte moyenne en fraude externe est de l'ordre de 2 049 \$ si nous tenons compte de toutes les données. Ainsi, nous remarquons que la *fraude externe* constitue un risque à haute fréquence et basse sévérité contrairement au type de risque *emploi, pratiques et sécurité environnementale* où le nombre de pertes est seulement 97 alors que la moyenne est de 34 220 \$. Ce constat confirme le fait que les montants de pertes par type de risque présentent des caractéristiques différentes justifiant la répartition par type de risque.

¹⁹ Nous supposons un taux de change de 1USD=1,15 CAD.

Force est de constater que les données de pertes sont loin de présenter une distribution symétrique. En effet, une simple comparaison entre la moyenne et la médiane nous indique que la médiane est, dans tous les cas, beaucoup moins élevée que la moyenne, montrant ainsi des distributions asymétriques. Ce constat est confirmé par le calcul du coefficient d'asymétrie qui est positif dans tous les cas favorisant ainsi des queues plus épaisses. De plus, des coefficients d'aplatissement très élevés montrent l'existence de distribution leptocurtique à queues épaisses. Nous citons l'exemple du type de risque *fraude externe* où le coefficient d'aplatissement est de l'ordre de 2 722 et le coefficient d'asymétrie est évalué à 46. En effet, nous remarquons que 98 % des pertes sont inférieures à 10 000 \$.

Contrairement au risque de marché où il est acceptable d'assumer que les variations des prix, des taux, etc. sont normales, les pertes opérationnelles ne peuvent en aucun cas être modélisées par des distributions normales, comme le montrent ces statistiques. Il y a donc lieu de tester d'autres distributions plus flexibles.

3.3.3. Statistiques descriptives sur les fréquences des pertes

Nous présentons, dans une première étape des statistiques descriptives sur les fréquences des pertes par jour. Nous avons choisi la période jour dans le but d'avoir plus d'observations possibles sur l'historique de 3 ans. En effet, il est important d'avoir un nombre élevé de données pour avoir des estimations robustes des distributions.

Le tableau 3-1b présente les statistiques descriptives des fréquences des pertes par jour, et ce, par type de risque. Nous remarquons que la fréquence moyenne par jour est de moins d'un événement pour tous les types de risque excepté pour *fraude externe* et *gestion d'exécution et de la livraison des processus* où le nombre de pertes est assez important. Il est à noter également que les fréquences maximales par jour pour ces deux types de risque sont respectivement de 226 et 66. Par ailleurs, la comparaison entre les moyennes et les variances nous montre que la variance excède la moyenne pour tous les types de risque. Ce constat nous laisse croire qu'à priori une

distribution favorisant la surdispersion modéliserait mieux les fréquences que la distribution Poisson.

Dans une deuxième étape, nous nous intéressons aux fréquences par semaine. Une analyse détaillée du nombre de pertes par jour nous montre effectivement que les fréquences de ces pertes sur les jours de semaine ne sont pas identiquement distribuées. Le tableau 3-1c présente les fréquences moyennes par type de risque selon le jour de semaine. En effet, nous remarquons que les fréquences sont moins élevées pendant les fins de semaine. Il est possible que le nombre de pertes soit moins important les jours de fermeture des institutions bancaires. Cependant, il est difficile d'admettre qu'il y ait moins d'événements de pertes relatives à certains types de risque (tel que *fraude externe...*) les fins de semaine par rapport aux autres jours. Il est évident que ceci est dû à un biais dans la date de comptabilisation de la perte. Ainsi, une manière de pallier à ce problème est de considérer le nombre de pertes par semaine. Nous allons toutefois perdre des observations puisque nous limitons notre échantillon à 106 fréquences. Une autre façon d'aborder ce problème est de désaisonnaliser les données. Nous ne traitons pas cette méthode dans ce chapitre mais elle pourrait être envisagée comme extension.

Le tableau 3-1d présente les statistiques descriptives relatives aux fréquences des pertes par semaine. Nous remarquons que les moyennes sont inférieures aux variances favorisant le choix de distributions caractérisées par la surdispersion pour la modélisation des fréquences.

Il est à mentionner également que nous avons vérifié s'il y ait une concentration aux fins de mois ou à la fin de l'année, ce qui peut fausser toute la modélisation des fréquences. Il est difficile de déceler l'existence d'une éventuelle concentration dans les fins des années puisque nous disposons de 3 ans d'historique seulement. Par contre, nous avons remarqué une légère concentration des pertes de type *gestion d'exécution et de la livraison des processus* aux fins de mois. La répartition des pertes des autres types de risque est uniforme sur les jours ou les semaines du mois.

3.4. Le modèle *LDA*

3.4.1. Conditions d'application du modèle

Voici quelques hypothèses d'ordre général sous-jacentes à l'utilisation de la méthode LDA :

- Le passé est une bonne approximation du futur : La structure organisationnelle, les politiques et les contrôles changent rapidement dans une institution financière. Une des hypothèses de base est que quoique la cause d'un événement de risque opérationnel puisse être corrigée (par l'ajout de contrôles supplémentaires par exemple), la ligne d'affaires représente un risque inhérent tel qu'une perte de même amplitude peu se reproduire. En d'autres mots, nous supposons que les pertes passées représentent une bonne estimation des pertes futures, en termes de fréquence et de sévérité.
- L'historique des données de pertes est suffisamment long : La BRI définit qu'une « fenêtre » de données historiques suffisamment longue serait de trois ans, et préférablement, de cinq ans. La période considérée dans notre étude est de 3 ans. Pour les types de risque où nous observons une haute fréquence d'événements de risques opérationnels, cette fenêtre est acceptable. Inversement, pour les types de risque où nous observons une basse fréquence d'événements de risques opérationnels, la fenêtre d'analyse devrait être plus longue.
- Tous les types de risque à l'étude contiennent suffisamment d'événements de grande ampleur (\$) pour assurer une modélisation adéquate de la distribution de sévérité : Nous supposons que le nombre d'observations minimal requis pour effectuer la modélisation est de 50. Ainsi, seuls les types de risque contenant plus de 50 événements de perte sont couverts par cette méthodologie. Le type de risque *perturbation des affaires et défaillance des systèmes* contenant peu de données est exclu. Une autre méthode, à savoir

l'analyse par scénarios, est prévue pour le calcul du capital de ce type de risque.

- Les données collectées sont de qualité (données intègres et intégrales) : L'information collectée relative à un événement de risque opérationnel est juste et fiable de sorte qu'aucun biais de collecte (par exemple : l'utilisation de la date de comptabilisation de la perte au lieu de la date d'événement) n'est introduit dans la méthodologie. Certaines améliorations sont toujours à effectuer au niveau des données collectées.
- Les pertes opérationnelles sont indépendantes et identiquement distribuées. La distribution de sévérité est indépendante de celle des fréquences.

3.4.2. Présentation du modèle LDA

Une perte opérationnelle L_i est un montant d'argent déboursé par une unité quelconque de la banque suite à un événement opérationnel tel que défini précédemment. Seules les pertes opérationnelles directes²⁰ sont collectées et sont intégrées dans le calcul du capital opérationnel. Les pertes indirectes telles que les modifications apportées à un processus ou ajout de contrôles post événementiels, les coûts d'opportunité et les manques à gagner sont donc exclus.

Une perte agrégée S sur une période de temps est mesurée de la façon suivante :

$$S = \sum_{i=1}^N L_i$$

où N est un nombre aléatoire représentant la fréquence des pertes ayant lieu pendant une période. Comme nous l'avons indiqué dans les hypothèses, nous supposons l'indépendance entre la distribution de L_i (distribution de sévérité) et celle de N (distribution de fréquence).

²⁰ Une perte est directe dans la mesure où elle n'aurait jamais été constatée si l'événement opérationnel ne s'était pas matérialisé. Les pertes directes incluent : les radiations, les pertes de recours, les pertes réglementaires, les pertes légales et les restitutions.

Le modèle LDA consiste donc à estimer la distribution de sévérité, celle des fréquences et à déterminer un certain quantile de la distribution des S (distribution agrégée). La perte agrégée par période est obtenue soit par la transformation de Fourier telle que proposée par Klugman et al. (2004) soit par simulation Monte Carlo ou par approximation analytique. Dans cette étude, la distribution agrégée est générée par simulation Monte Carlo comme nous allons le montrer plus loin dans ce chapitre. Dans ce qui suit, nous allons estimer chacune des distributions de sévérité et de fréquence et déterminer la VaR opérationnel.

3.5. Estimation de la distribution de sévérité

La première étape de l'implantation de la méthode LDA pour le calcul du capital consiste à chercher les distributions paramétriques qui s'ajustent le mieux aux données historiques des montants de pertes. L'estimation sans biais des paramètres de ces distributions va influencer énormément les résultats du capital opérationnel. Nous allons, dans cette section, explorer plusieurs méthodes d'estimation de la distribution de telle sorte que nous arrivons à un meilleur ajustement de l'ensemble des données historiques.

La méthode, en fait, diffère selon la nature des données et leurs caractéristiques. Nous procédons au préalable à l'estimation de la distribution à partir de l'ensemble des données de pertes. Cependant, si les tests d'ajustement que nous mettons en place, rejettent ces distributions, nous devons donc adopter une autre méthode. En effet, il est possible dans ce cas de découper la distribution, surtout si la taille de l'échantillon nous le permet, et d'ajuster par conséquent une distribution à chaque partie de la distribution empirique.

Outre le comportement des pertes et le nombre de données qui entrent en considération dans le choix de la méthode d'ajustement des distributions de sévérité, le seuil de collecte des données s'avère d'une grande importance. En effet, en risque opérationnel, la plupart des institutions financières ont choisi un niveau de pertes à partir duquel les pertes opérationnelles vont être collectées. Ce niveau varie d'une banque à une autre, la plupart des banques canadiennes l'ont fixé à 10 000 \$ alors

qu'il existe dans l'industrie des seuils moins élevés (1 000 \$) ou plus grands (15 000 \$). Le fait d'avoir des données tronquées change la méthode usuelle d'estimation des paramètres. Il faut donc tenir compte de ce seuil lors de l'estimation des paramètres.

Plusieurs méthodes sont proposées dans la littérature pour l'estimation des paramètres. Nous citons le maximum de vraisemblance et la méthode des moments entre autres. Cependant, cette dernière s'avère plus difficile dans le cas de données tronquées puisque la détermination des moments d'une distribution tronquée est plus compliquée que l'estimation des paramètres d'une distribution dont les données sont non tronquées. Il est toujours possible cependant d'estimer les paramètres de la distribution des excédents des pertes par rapport au seuil de collecte (Dutta et Perry, 2006). Cette façon de faire nous ramène à l'estimation usuelle des paramètres de la distribution certes, mais elle ignore les pertes inférieures au seuil de collecte vu que la distribution n'en tient pas compte. Nous ne pouvons donc générer que les excédents des pertes grâce à la distribution choisie. Cette méthode pose un problème de sous estimation ou sur estimation du capital surtout si le seuil est assez élevé et le volume des petites pertes est non négligeable. Une correction de ce biais est donc nécessaire dans le cas d'utilisation de cette méthode.

Nous identifions dans une première étape les distributions qui seront testées. Ensuite, nous décrivons les différentes méthodes d'estimation selon la nature des données. Nous présentons enfin les résultats de l'estimation pour chaque type de risque, et ce, pour les quatre distributions testées.

3.5.1. Distributions testées

Dans le but d'avoir un bon ajustement des distributions empiriques par des distributions paramétriques, nous testons plusieurs distributions pour pouvoir en choisir la meilleure. Il ne faut pas se limiter à une seule distribution comme le font la plupart des praticiens dans l'industrie. Il vaut mieux commencer par tester des distributions simples (à un seul paramètre) comme l'exponentielle, des distributions usuelles comme la lognormale ou la Weibull (à deux paramètres) et des distributions

plus complexes, mais plus flexibles comme la GB2 (à quatre paramètres). Nous estimerons les paramètres de ces quatre distributions. Les caractéristiques ainsi que les fonctions de densité et de répartition de ces fonctions ont été déjà présentées au tableau 2-1 du chapitre 2.

D'autre part, nous proposons la distribution Pareto pour l'estimation des queues, dans des cas particuliers. Sa fonction de densité f de paramètre $\alpha > 0$ est donnée par l'expression suivante :

$$f(x, \alpha) = \frac{\alpha}{x^{\alpha+1}} \quad \text{pour } x \geq 1$$

Sa fonction cumulative est donnée par :

$$F(x, \alpha) = 1 - \frac{1}{x^\alpha} \quad \text{pour } x \geq 1$$

Il est à remarquer que la fonction de densité est décroissante pour tout $x \geq 1$. De plus, elle décroît plus rapidement lorsque α augmente. L'utilisation de cette fonction est appropriée pour la modélisation des distributions asymétriques à queue épaisse. D'ailleurs, la moyenne, la variance et les autres moments ne sont finis que pour des valeurs suffisamment élevées du paramètre de forme α .

La Pareto peut être généralisée en introduisant un paramètre d'échelle β . Supposons que Z suit une distribution Pareto de paramètre α , la variable aléatoire $X = \beta Z$ suit une distribution Pareto de paramètre de forme α et de paramètre d'échelle β . Les fonctions de densité f et de répartition F sont respectivement les suivantes :

$$f(x, \alpha, \beta) = \frac{\alpha \beta^\alpha}{x^{\alpha+1}} \quad \text{pour } x \geq \beta$$

$$F(z, \alpha, \beta) = 1 - \left(\frac{\beta}{x}\right)^\alpha \quad \text{pour } x \geq \beta$$

3.5.2. Modélisation de l'ensemble des données

La modélisation des montants de pertes diffère selon la nature des données et leurs caractéristiques. Nous allons dans une première étape modéliser l'ensemble des données de pertes. Cependant, vu que certaines pertes relatives à certains types de

risque sont tronquées alors que d'autres ne le sont pas, il est important de distinguer les deux méthodes d'estimation. Nous allons ainsi présenter d'abord la méthode d'estimation des paramètres dans le cas où les données sont complètes. Nous développons ensuite la méthode d'estimation dans le cas de données tronquées.

3.5.2.1. Cas de données complètes

Cette méthode consiste à utiliser les méthodes usuelles d'estimation des paramètres telles que la méthode par maximum de vraisemblance et la méthode des moments. Comme les données sont complètes, il s'agit donc de faire une estimation non conditionnelle des paramètres.

Soient :

x : les montants de pertes.

\mathcal{D} : une distribution continue de sévérité dont les paramètres sont représentés par le vecteur θ .

$f(x, \theta)$: la fonction de densité de la distribution \mathcal{D} de paramètres θ .

$F(x, \theta)$: la fonction de répartition de la distribution \mathcal{D} de paramètres θ .

Pour le maximum de vraisemblance, il s'agit de trouver les paramètres θ qui maximisent la fonction de vraisemblance. La solution optimale θ_{MLE} solutionne :

$$\theta_{MLE} = \arg \max_{\theta} L(x, \theta)$$

avec :

$$L(x, \theta) = \sum_{i=1}^n \text{Log}[f(x_i, \theta)] \quad (1)$$

Il s'agit des types de risque *fraude interne* et *fraude externe* pour lesquels toutes les données de pertes opérationnelles sont collectées. Les résultats de l'estimation seront présentés seulement pour le type de risque *fraude interne*. Nous n'allons pas présenter les résultats de l'estimation des paramètres relative aux pertes de type *fraude externe* puisque les tests d'ajustement n'ont permis de retenir aucune des distributions testées.

Nous allons donc proposer une autre méthode de modélisation de ces pertes à la section 5.3.

3.5.2.2. Cas de données tronquées

Comme nous l'avons décrit précédemment, la collecte des pertes opérationnelles se fait dans la plupart des cas au-dessus d'un certain seuil. Pour tenir compte des pertes non collectées, nous procédons à une estimation des paramètres conditionnelle au fait que nous n'observons que les pertes supérieures au seuil de collecte.

Soient :

s : seuil de collecte,

y : les montants de pertes observés, avec $y \geq s$,

n : le nombre de pertes collectées,

\mathcal{D} : une distribution continue de sévérité dont les paramètres sont représentés par le vecteur θ .

La fonction de densité de la distribution \mathcal{D} conditionnelle au fait que nous n'observons que y excédant le seuil s s'écrit :

$$f(y, \theta | y \geq s) = \frac{f(y, \theta)}{1 - F(s, \theta)} \text{ pour } y \geq s$$

La fonction de répartition de la distribution \mathcal{D} conditionnelle au fait que nous n'observons que y excédant le seuil s est la suivante :

$$F(y, \theta | y \geq s) = \frac{F(y, \theta) - F(s, \theta)}{1 - F(s, \theta)} \text{ pour } y \geq s$$

La méthode d'estimation des paramètres que nous retenons est par optimisation du maximum de vraisemblance qui est plus facile à implanter, dans ce cas, que les autres méthodes, vu que les distributions sont tronquées. La solution optimale θ_{MLE} solutionne :

$$\theta_{MLE} = \arg \max_{\theta} L(y_i, \theta)$$

avec

$$\begin{aligned} L(y_i, \theta) &= \sum_{i=1}^n \text{Log}[f(y_i, \theta | y_i \geq s)] \\ &= \sum_{i=1}^n \text{Log}\left[\frac{f(y_i, \theta)}{1 - F(s, \theta)}\right] \\ &= \sum_{i=1}^n \text{Log}(f(y_i, \theta)) - n\text{Log}(1 - F(s, \theta)) \end{aligned} \quad (2)$$

Parmi les pertes dont nous disposons dans notre base, nous avons celles relatives aux types de risque *actifs tangibles, corporels et endommagés, emploi, pratiques et sécurité environnementale, gestion d'exécution et de la livraison des processus et clients, produits et pratiques commerciales* qui sont tronquées. Les tests d'ajustement que nous avons mis en place ont rejeté l'ajustement de toutes les distributions proposées pour les deux derniers types de risque cités précédemment. Nous n'allons pas présenter ces résultats. En revanche, nous allons les modéliser par une autre méthode que nous développons à la prochaine section.

3.5.3. Découpage de la distribution

Dans certains cas, les distributions paramétriques testées ne s'ajustent pas à l'ensemble des données. Ceci se présente généralement lorsque les pertes sont très nombreuses ou lorsque leur distribution empirique possède deux sommets, etc. Nous proposons dans ce cas de découper la distribution empirique en parties et de faire l'ajustement des distributions par parties.

Plusieurs articles dans la littérature traitent l'ajustement des queues des distributions avec principalement la théorie des valeurs extrêmes. Cependant, ajuster seulement la queue de la distribution suppose la négligence du corps de la distribution. Alors que ceci devrait entrer en compte dans le calcul du capital et pourrait avoir un impact non négligeable.

Il n'y a pas un seul et unique découpage d'une distribution, l'important étant de trouver un bon ajustement. Le découpage devrait aller au pair avec l'objectif de l'ajustement. En d'autres termes, si nous faisons l'ajustement dans le but d'estimer un

capital à un niveau de 99.9 %, les queues des distributions vont donc entrer en jeu. La qualité d'ajustement est essentielle pour éviter une sous ou une sur évaluation du capital. Ainsi dans notre cas, il est important de bien estimer la queue ou l'aile droite de la distribution des pertes.

Une fois la distribution des queues estimée, il y a lieu d'estimer le corps de la distribution. Si nous n'arrivons pas à avoir un bon ajustement avec les distributions testées, nous allons découper le corps de la distribution en deux ou plusieurs parties. Nous allons dans ce qui suit présenter une méthode d'estimation de la partie gauche, du centre et de la queue de la distribution.

C'est le cas des pertes de type *clients, produits et pratiques commerciales, gestion d'exécution et de la livraison des processus et fraude externe* où la modélisation par une seule distribution paramétrique n'a pas donné des résultats concluants. Nous recourons par conséquent au découpage de la distribution. Il est à signaler qu'il est important d'avoir un nombre suffisamment élevé d'observations pour accomplir cette méthode. En effet, nous devons avoir assez de données²¹ dans chaque partie pour pouvoir estimer les paramètres des distributions. Cette condition est bien satisfaite pour les types de risque que nous allons modéliser dans cette section.

3.5.3.1. Estimation de la queue de la distribution :

Dans ce qui suit, nous proposons de modéliser les queues des distributions empiriques. Nous choisissons, au préalable, plusieurs seuils s_q de pertes correspondants à différents quantiles élevés (0,75; 0,8 ; 0,85...). Ensuite, nous estimons le paramètre de la distribution Pareto pour chacun des échantillons de pertes supérieures aux différents seuils s_q . Enfin, nous retenons le seuil pour lequel la Pareto donne un meilleur ajustement (la distribution pour laquelle la p-value du test d'ajustement est la plus élevée). Cette méthode est semblable à celle développée par Peters et al. (2004).

²¹ Nous ne nous intéressons pas dans cette étude à déterminer le nombre minimal d'observations que nous devons avoir pour pouvoir modéliser le paramètre de la distribution. Ceci pourrait faire l'objet d'autres recherches.

Nous avons choisi des quantiles élevés afin d'avoir un ajustement uniquement pour la queue. Cependant, il est important de garder un nombre suffisant d'observations pour pouvoir faire l'estimation. Nous avons jugé qu'il faut garder un minimum de 50 observations pour l'estimation du seul paramètre de forme α de la loi Pareto pour chacun des types de risque.

Soient :

s_q : seuil de perte correspondant au quantile q , ce seuil correspondra au paramètre d'échelle de la Pareto généralisée. En effet, nous allons supposer à chaque fois le paramètre d'échelle égal au seuil s_q et nous estimons le seul paramètre α .

z : les montants de pertes supérieures au seuil s_q

n_q : le nombre d'observations z dans l'échantillon défini par $(z_i | z_i \geq s_q, i = 1, 2, \dots, n_q)$

\mathcal{P} : la distribution Pareto à tester

La fonction de densité de la distribution \mathcal{P} est donc :

$$f(z, \alpha, s_q) = \frac{\alpha s_q^\alpha}{z^{\alpha+1}} \quad \text{avec } z \geq s_q$$

Pour le maximum de vraisemblance, il s'agit de trouver le paramètre α qui maximise la fonction suivante pour les différents seuils s_q

$$\alpha_{MLE} = \arg \max_{\alpha} \sum_{i=1}^{n_q} \text{Log}[f(x_i, \alpha | s_q)]$$

$$\alpha_{MLE} = \frac{n_q}{\sum_{i=1}^{n_q} \text{Log}\left(\frac{x_i}{s_q}\right)} \quad (3)$$

Nous avons donc estimé le paramètre α correspondant à chaque seuil s_q , et ce, pour chacun des trois types de risque cités auparavant. Les tests d'ajustement nous permettent de choisir le seuil qui permet d'avoir le meilleur ajustement de la queue.

3.5.3.2. Estimation du corps de la distribution :

Une fois que nous avons modélisé la queue de la distribution, il faut estimer la distribution qui s'ajuste le mieux au corps de la distribution. Sachant que ce dernier est représenté par toutes ou une partie des pertes au dessous du seuil de quantile retenu à la section précédente, nous allons commencer par différentes distributions : exponentielle, lognormale, Weibull et GB2 pour l'ensemble de l'échantillon de pertes.

Soient :

s_q : seuil de queue de perte retenu correspondant au quantile q

z' : les montants de pertes inférieures au seuil s_q

n' : le nombre d'observations z' , c'est le nombre des $z'_i | z'_i < s_q, i = 1, 2, \dots, n_q$

\mathcal{D} : la distribution à tester

La fonction de densité f de la distribution \mathcal{D} des pertes inférieures au seuil de queue retenu est la suivante :

$$f(z', \theta | z' < s_q) = \frac{f(z', \theta)}{F(s_q, \theta)}$$

La fonction cumulative F de la distribution \mathcal{D} des pertes inférieures au seuil de queue retenu est donc :

$$F(z', \theta | z' < s_q) = \frac{F(z', \theta)}{F(s_q, \theta)}$$

Pour le maximum de vraisemblance, il s'agit de trouver les paramètres θ qui maximisent la fonction suivante :

$$\theta_{MLE} = \arg \max_{\theta} L(z'_i, \theta)$$

avec, après substitution et réécriture,

$$\begin{aligned}
L(z'_i, \theta) &= \sum_{i=1}^{n'} \text{Log}[f(z'_i, \theta | z'_i < s_q)] \\
&= \sum_{i=1}^{n'} \text{Log} \left[\frac{f(z'_i, \theta)}{F(s_q, \theta)} \right] \\
&= \sum_{i=1}^{n'} \text{Log}(f(z'_i, \theta)) - n' \text{Log}(F(s_q, \theta)) \tag{4}
\end{aligned}$$

Cette formule est utilisée également lorsque les données sont censurées, ou inférieures à n'importe quel seuil.

Toutefois, lorsque les données de pertes sont tronquées comme nous l'avons vu précédemment, il faut en tenir compte dans l'estimation du corps de la distribution.

Soient :

s : seuil de collecte

s_q : seuil de queue retenu correspondant au quantile q

z'' : les montants de pertes inférieures au seuil de queue s_q et supérieures au seuil de collecte s .

n'' : le nombre d'observations z'' , c'est le nombre des $z''_i | s < z''_i < s_q, i = 1, 2, \dots, n_q$

\mathcal{D} : la distribution à tester

La fonction de densité de la distribution \mathcal{D} des pertes comprises entre le seuil s et le seuil de queue s_q retenu est la suivante :

$$f(z'', \theta | s \leq z'' < s_q) = \frac{f(z'', \theta)}{F(s_q, \theta) - F(s, \theta)}$$

La fonction de répartition de la distribution \mathcal{D} des pertes comprises entre le seuil s et le seuil de queue s_q retenu s'écrit :

$$F(z'', \theta | s \leq z'' < s_q) = \frac{F(z'', \theta) - F(s, \theta)}{F(s_q, \theta) - F(s, \theta)}$$

Pour le maximum de vraisemblance, il s'agit de trouver les paramètres θ qui maximisent la fonction suivante :

$$\theta_{MLE} = \arg \max_{\theta} L(z''_i, \theta)$$

où:

$$\begin{aligned}
 L(z''_i, \theta) &= \sum_{i=1}^{n''} \text{Log}[f(z''_i, \theta | s \leq z''_i < s_q)] \\
 &= \sum_{i=1}^{n''} \text{Log}\left[\frac{f(z''_i, \theta)}{F(s_q, \theta) - F(s, \theta)}\right] \\
 &= \sum_{i=1}^{n''} \text{Log}[f(z''_i, \theta)] - n'' \text{Log}[F(s_q, \theta) - F(s, \theta)] \quad (5)
 \end{aligned}$$

Cette formule peut être utilisée lorsque les données sont comprises entre deux seuils ou bornes.

Pour le type de risque *clients, produits et pratiques commerciales*, nous avons modélisé les pertes comprises entre le seuil de collecte et le seuil de queue retenu. Les résultats de l'estimation seront présentés à la section 5.5.

Quant aux types de risque *fraude externe et gestion d'exécution et de la livraison des processus*, la modélisation du corps de la distribution avec les distributions paramétriques proposées s'est avérée non concluante, puisque les tests d'ajustement n'ont permis d'en retenir aucune²². Dans ce cas, nous proposons un découpage différent de la distribution. Nous citons l'exemple du type de risque *fraude externe* où aucune des distributions testées ne s'ajustait aux données. Une analyse minutieuse des pertes de fraudes externes montre effectivement que la distribution empirique des données présente deux sommets, dont le deuxième au niveau de la perte 5000 \$. Nous découpons ainsi la distribution au niveau de ce point. Nous estimons l'aile gauche de la distribution avec la méthode du maximum de vraisemblance selon l'équation (4). Par ailleurs, le centre de la distribution empirique est estimé par une autre distribution paramétrique selon la fonction (5) du maximum de vraisemblance, vu que les pertes sont comprises entre deux seuils.

Quant au type de risque *gestion d'exécution et de la livraison des processus*, l'estimation de l'aile gauche de la distribution empirique est faite avec la fonction de

²² Nous n'avons pas présenté les résultats de cette estimation, mais ils sont disponibles à la demande.

vraisemblance (5) puisque les données sont tronquées au seuil de collecte, de même pour le centre de la distribution.

3.5.4. Tests de bon ajustement

Il est possible d'estimer n'importe quelle distribution paramétrique de domaine de définition positive. Cependant, seules quelques distributions offrent un bon ajustement. Il est possible d'évaluer le degré d'ajustement des distributions grâce aux méthodes graphiques et aux tests formels. Il existe plusieurs tests formels développés dans la littérature tels que le test de Khi deux, le test d'Anderson Darling, de Cramér-von-Mises, de Kolmogorov-Smirnov, etc. ainsi que des tests graphiques tels que le graphique des quantiles (Q-Q plot), ou le graphique de la fonction cumulative versus celle empirique.

Nous avons tout d'abord appliqué ces tests graphiques pour avoir une idée préalable sur le degré d'ajustement. Les résultats des tests graphiques doivent être confirmés par des tests formels basés sur le calcul d'une statistique. Cependant, si les graphiques montrent que la distribution empirique est loin d'être dérivée de la distribution paramétrique testée, alors il est invraisemblable que les tests formels donnent des résultats différents.

Ensuite, nous mettons en place les tests Kolmogorov-Smirnov, Anderson Darling et Cramér-von Mises avec la procédure *bootstrap* paramétrique tel que défini dans le chapitre 2. Il est à noter que nous devons tenir compte des caractéristiques de l'échantillon dans l'implantation du test. En d'autres termes, si les données sont tronquées ou comprises entre deux seuils, il faut générer un échantillon respectant la même condition dans la procédure *bootstrap* paramétrique avant de calculer la statistique.

3.5.5. Résultats de l'estimation des paramètres

Les tests graphiques pour chacun des types de risque montrent en général que seule la distribution GB2 offre un bon ajustement aux données. En effet, les graphiques des

quantiles sont non linéaires pour les distributions lognormale, exponentielle et Weibull. Ce constat devrait être approuvé par les tests statistiques.

Le tableau 3-2 présente les résultats de l'estimation des distributions pour les types de risque *fraude interne, actifs, tangibles, corporels et endommagés* et *emploi, pratiques et sécurité environnementale*. Contrairement aux autres types de risque dont les pertes sont tronquées, les pertes relatives au type de risque *fraude interne* sont intégrales. Les résultats de ce type de risque synthétisés dans le tableau 3-2a montrent que les distributions lognormale et GB2 s'ajustent bien aux données de pertes alors que nous rejetons l'ajustement des distributions exponentielle et Weibull. Cependant, la distribution lognormale a une valeur de p-value plus élevée de celle de la GB2. Nous retenons ainsi la distribution lognormale comme la distribution de sévérité qui décrit le mieux le comportement des fraudes internes dans la banque étudiée.

Les tableaux 3-2b et 3-2c présentent les résultats des types de risque *actifs, tangibles, corporels et endommagés* et *emploi, pratiques et sécurité environnementale*. Rappelons que ces pertes sont tronquées au niveau de s . Seules les distributions lognormale et GB2 offrent un bon ajustement aux données de pertes tronquées. Nous n'avons pas pu accomplir les tests de Kolmogorov-Smirnov, Anderson Darling et Cramér-von Mises pour la distribution Weibull pour les deux types de risque. En effet, la modélisation des pertes en tenant compte de la troncature complique beaucoup l'estimation des distributions et les tests standards ne pourraient être appliqués. Pour les types de risques *actifs, tangibles, corporels et endommagés* et *emploi, pratiques et sécurité environnemental*, nous remarquons que les résultats des tests d'ajustement sont similaires pour les distributions lognormale et GB2. Nous retenons dans ce qui suit la lognormale vue sa simplicité par rapport à la GB2.

Quant aux autres types de risque, nous procédons au découpage de la distribution empirique vu que la première méthode n'a permis de retenir aucune distribution. De plus, le nombre d'observations est assez élevé pour les trois échantillons, le découpage est donc possible, tel que décrit à la section précédente. Pour le type de risque *clients, produits et pratiques commerciales*, nous estimons tout d'abord la queue de la distribution en testant la distribution Pareto sur plusieurs échantillons de

queue correspondants à des quantiles élevés. Les résultats du tableau 3-3a nous permettent de retenir le seuil de 62 000 \$. En d'autres termes, les pertes supérieures à ce seuil suivent une Pareto de paramètre 1,18. Les pertes au dessous de ce seuil ont été modélisées par les quatre distributions citées à la section 5.1. Les résultats des tests Kolmogorov-Smirnov, Anderson Darling et Cramér-von Mises rejettent toutes les distributions excepté la GB2. Nous retenons donc cette distribution pour la modélisation des pertes situées au-dessous du seuil 62 000 \$.

De même pour le type de risque *gestion d'exécution et de la livraison des processus*, nous commençons par la modélisation de la queue. Plusieurs seuils correspondant à différents quantiles ont été choisis. Le meilleur ajustement de la Pareto, en se basant sur le test de Kolmogorov-Smirnov, Anderson Darling et Cramér-von Mises est obtenu pour le seuil de 26 000 \$ correspondant au quantile de 96 % (tableau 3-5a). Le paramètre de la Pareto est estimé 1,11. Ensuite, nous avons estimé le corps de la distribution avec les distributions proposées. Aucune des distributions ne s'ajustait aux données empiriques au-dessous du seuil de queue. Nous avons par conséquent découpé le corps de la distribution en deux parties qui constituent l'aile gauche et le centre de la distribution. Les résultats du tableau 3-5b, portant sur l'estimation de l'aile gauche (données entre s et s') de la distribution, montrent que seule la GB2 offre un bon ajustement aux données historiques selon la p-value du test de Kolmogorov-Smirnov. Il est à mentionner que l'estimation des distributions lognormale et Weibull ne donnent pas de paramètres cohérents. Les tests de Kolmogorov-Smirnov, Anderson Darling et Cramér-von Mises n'ont pu donc être établi. Quant au centre de la distribution, nous avons modélisé les pertes par les mêmes distributions. Les résultats du test d'ajustement montrent que les distributions lognormale et GB2 offrent un bon ajustement. La comparaison des p-values des deux distributions favorise le modèle GB2.

La même méthode a été appliquée pour le type de risque *fraude externe*. Pour l'estimation de la queue, nous avons testé plusieurs seuils correspondant à de nombreux quantiles (entre 0,85 et 0,997), puisque la taille de l'échantillon nous le permet (22 179). Nous avons retenu le seuil de 38 000 \$ et le paramètre de la Pareto

est estimé à 1,33. Nous avons ensuite découpé le corps de la distribution empirique puisque l'ajustement de l'ensemble des données au-dessous du seuil de queue n'a pas donné de résultats concluants. Nous avons ainsi estimé l'aile gauche de la distribution constituée par les pertes au-dessous de s'' . Les résultats présentés dans le tableau 3-6b permettent de retenir uniquement la GB2. En effet, les tests Kolmogorov-Smirnov, Anderson Darling et Cramér-von Mises rejettent l'ajustement des autres distributions. Enfin, le centre de la distribution empirique est modélisé également par une distribution GB2.

Il est à remarquer que dans aucun cas, les distributions exponentielle et Weibull n'ont offert un bon ajustement aux données. La distribution GB2 était, dans la plupart des cas, la meilleure candidate pour la modélisation des montants de pertes opérationnelles. Ceci confirme notre hypothèse de départ stipulant que la GB2 offre un bon ajustement aux données vu sa grande flexibilité. Nous notons, par ailleurs, que l'estimation de cette distribution, surtout dans le cas tronqué, est assez ardue.

3.6. Estimation de la distribution de fréquence

L'objectif de cette section est de modéliser le nombre de pertes opérationnelles sur une période déterminée. En effet, comme nous l'avons déjà montré précédemment, la VaR opérationnel dépend de la sévérité des pertes et de leur fréquence. La bonne estimation de la distribution des fréquences conduit à une meilleure estimation du capital opérationnel. La plupart des études précédentes n'ont pas donné une grande importance au choix de la distribution de fréquence, et ce, en prenant la distribution Poisson comme étant la bonne candidate sans faire de tests. Force est de constater que le modèle Poisson présente la caractéristique de l'équidispersion ce qui n'est pas le cas dans les données de fréquences opérationnelles. Nous allons dans ce qui suit proposer deux distributions et tester le degré d'ajustement de chacune. Comme pour certains types de risque la collecte de pertes se fait à partir d'un certain seuil, il y a lieu de corriger le biais qui en résulte sur les données de fréquences.

3.6.1. Distributions testées

Nous proposons les distributions Poisson et binomiale négative pour estimer le nombre de pertes opérationnelles. Plusieurs études ont déjà modélisé les fréquences des pertes opérationnelles avec la Poisson alors que peu de recherches l'ont fait avec le modèle binomial négatif.

Soit Y une variable aléatoire positive discrète :

$$Y \sim \text{Poisson}(\lambda) \quad \text{alors} \quad P(Y = y_i) = \frac{e^{-\lambda} \lambda^{y_i}}{y_i!}$$

$$Y \sim \text{Binomiale négative}(\alpha, \beta) \quad \text{alors} \quad P(Y = y_i) = \frac{\Gamma(\alpha + y_i)}{\Gamma(\alpha) y_i!} \times \left(\frac{\beta}{1 + \beta} \right)^{y_i} \times \left(\frac{1}{1 + \beta} \right)^\alpha$$

3.6.2. Estimation des paramètres

Nous estimons les paramètres de chacune des deux distributions par maximum de vraisemblance, et ce, pour chacun des six types de risque. Comme nous l'avons déjà signalé, nous modélisons les fréquences par jour et les fréquences par semaine afin de corriger un éventuel biais de collecte.

Rappelons que pour certains types de risque (*actifs tangibles, corporels et endommagés, emploi, pratiques et sécurité environnementale, clients, produits et pratiques commerciales et gestion d'exécution et de la livraison des processus*), les pertes ne sont collectées qu'à partir d'un certain seuil. Nous avons tenu compte de ce seuil de troncature lors de l'estimation de la distribution de sévérité. Il faut faire de même pour la modélisation des fréquences. Il est à préciser qu'il ne s'agit pas de distributions de fréquences tronquées. En effet, nous disposons du nombre de pertes supérieures au seuil de collecte. Il faut donc corriger les paramètres estimés pour tenir compte du nombre des pertes au-dessous du seuil de troncature.

3.6.3. Correction des paramètres

Pour la distribution Poisson, la correction du paramètre λ a été déjà proposée par Frachot et al. (2003); et ce, pour tenir compte du nombre de pertes opérationnelles non collectées. Nous explorons le cas où le paramètre λ est aléatoire.

Soient :

y_{obs} : le nombre de pertes collectées,

$y_{\text{réel}}$: le nombre de pertes réelles,

X : montant des pertes,

s : seuil de collecte,

F : fonction cumulative de la distribution de sévérité, avec :

$$F(s) = \Pr[X \leq s]$$

$$\bar{F}(s) = 1 - F(s)$$

Supposons que la distribution de fréquence de toutes les pertes est Poisson de paramètre aléatoire $\lambda_{\text{réel}}$. Nous allons dans ce qui suit déterminer la distribution des fréquences observées. La probabilité d'observer i pertes opérationnelles dont le montant est supérieur à s est :

$$\begin{aligned}
P(y_{\text{obs}} = i) &= E \left\{ \sum_{k=i}^{\infty} P[y_{\text{réel}} = k \text{ et nombre } (X \geq s) = i] \right\} \\
&= E \left\{ \sum_{k=i}^{\infty} P[y_{\text{réel}} = k] \times \binom{k}{i} \bar{F}(s)^i F(s)^{k-i} \right\} \\
&= E \left\{ \sum_{k=i}^{\infty} \frac{e^{-\lambda_{\text{réel}}} \lambda_{\text{réel}}^k}{k!} \times \binom{k}{i} \bar{F}(s)^i F(s)^{k-i} \right\} \\
&= \sum_{k=i}^{\infty} \frac{k!}{i!(k-i)!} \bar{F}(s)^i F(s)^{k-i} E \left(\frac{e^{-\lambda_{\text{réel}}} \lambda_{\text{réel}}^k}{k!} \right) \\
&= \sum_{j=0}^{\infty} \frac{F(s)^j \bar{F}(s)^i}{i! j!} E(e^{-\lambda_{\text{réel}}} \lambda_{\text{réel}}^{i+j}) \quad \text{avec } j = k - i \\
&= E \left\{ \frac{\bar{F}(s)^i e^{-\lambda_{\text{réel}}}}{i!} \lambda_{\text{réel}}^i \sum_{j=0}^{\infty} \frac{F(s)^j \lambda_{\text{réel}}^j}{j!} \right\} \\
&= E \left\{ e^{-\lambda_{\text{estimé}}} \frac{\lambda_{\text{estimé}}^i}{i!} \right\} \quad \text{avec } \sum_{j=0}^{\infty} \frac{F(s)^j \lambda_{\text{réel}}^j}{j!} = e^{\lambda_{\text{réel}} F(s)}
\end{aligned}$$

où $\lambda_{\text{estimé}} = \lambda \bar{F}(s)$.

Pour le cas particulier où λ est non aléatoire, nous trouvons que la distribution des fréquences des pertes au-dessus du seuil est une Poisson de paramètre $\lambda_{\text{estimé}}$. Nous pouvons donc estimer $\lambda_{\text{estimé}}$ à partir des données de fréquence observées et déterminer par la suite le vrai paramètre $\lambda_{\text{réel}}$ de la distribution, tel que :

$$\lambda_{\text{réel}} = \frac{\lambda_{\text{estimé}}}{\bar{F}(s)} \quad (6)$$

Cette équation sert à corriger le paramètre estimé $\lambda_{\text{estimé}}$ pour tenir compte du nombre de pertes au-dessus du seuil de collecte.

Dans le cas où $\lambda \sim \Gamma(\alpha, \beta)$ qui correspond au cas où $Y \sim \text{binomiale négative} \left(\alpha, \frac{1}{1+\beta} \right)$, les paramètres estimés, à partir des données observées, est $\lambda_{\text{estimé}} \sim \Gamma(\alpha, \beta \bar{F}(s))$. Le nombre de pertes collectées suit donc une

distribution binomiale négative de paramètres $\left(\alpha, \frac{1}{1+\beta\bar{F}(s)}\right)$. Ainsi, une fois les paramètres du modèle binomial négatif estimés, nous corrigeons ces paramètres pour tenir compte des fréquences non observées des pertes au-dessous du seuil de collecte. Les paramètres réels $(\alpha_{\text{réel}}, \delta_{\text{réel}})^{23}$ sont donc :

$$\begin{cases} \alpha_{\text{réel}} = \alpha_{\text{estimé}} \\ \delta_{\text{réel}} = \frac{\delta_{\text{estimé}} \bar{F}(s)}{1 - \delta_{\text{estimé}} F(s)} \end{cases} \quad (7)$$

3.6.4. Tests de bon ajustement

Il est important à ce stade de tester le degré d'ajustement des distributions estimées, pour les différentes périodes choisies (jour et semaine) et pour chacun des types de risque. Comme nous sommes en présence de données discrètes, il nous est impossible d'appliquer les mêmes tests utilisés auparavant pour tester les distributions de sévérité. Nous mettons donc en place un test χ^2 avec *bootstrap* paramétrique, de la même façon que le test de Kolmogorov-Smirnov est construit.

Ce test exige que les données soient groupées en classes ou groupes. La statistique du test est calculée comme suit (Klugman, Panjer et Willmot, 1998) :

$$Q = \sum_{i=1}^k \frac{(n_j - E_j)^2}{E_j}$$

avec :

n_j : le nombre d'observations dans le groupe j , avec $j=1, \dots, k$,

E_j : le nombre d'observations espéré dans chaque groupe étant donné que le modèle est correct et les paramètres ont leurs valeurs estimées. Il est calculé :

$$E_j = n \Pr(X \in j^{\text{ème}} \text{ groupe}; \hat{\theta}) \text{ pour } j=1 \dots k$$

où n est la taille de l'échantillon.

²³ Le paramètre δ est défini tel que $\delta = \frac{1}{1+\beta}$

Cette statistique est comparée à une valeur critique tabulée. En effet, si Q excède $\chi^2_{d,\alpha}$ (où $d=k-r-1$ est le nombre de degrés de liberté et α est le seuil de significativité) alors l'hypothèse nulle est rejetée. Cependant pour des résultats plus fiables du test, nous proposons la méthode avec *bootstrap* paramétrique pour le calcul de la p-value du test.

Cette méthode, contrairement au test classique, consiste à calculer une p-value et les valeurs critiques par simulation Monte Carlo. La valeur de la p-value nous permet de décider si nous rejetons ou non l'hypothèse nulle stipulant le bon ajustement de la distribution aux données. Ainsi, les valeurs critiques vont être calculées à partir d'échantillons générés. Il s'agit d'appliquer donc l'algorithme suivant pour le calcul de la p-value :

- 9- Construire les groupes (se référer à Klugman et al., 1998).
- 10- Calculer le nombre de pertes n_j dans chaque groupe.
- 11- Calculer $E_j(\hat{\theta})$ pour chaque groupe, où $\hat{\theta}$ sont les paramètres estimés de la distribution à tester (Poisson ou binomiale négative).
- 12- Calculer la statistique Q_0 , telle que définie précédemment.
- 13- Générer, à partir des paramètres estimés de la distribution à tester, des fréquences de pertes de même échantillon que notre échantillon de départ.
- 14- À partir de l'échantillon généré, estimer les paramètres $\tilde{\theta}$ de la même distribution avec la méthode du maximum de vraisemblance.
- 15- Calculer n_j , $E_j(\tilde{\theta})$ et la statistique Q_i .
- 16- Comparer les statistiques Q_0 et Q_i . Si $Q_0 < Q_i$ alors un compteur h s'incrémente de 1.
- 17- Répéter les étapes 5 à 8 un grand nombre de fois N (ex. $N=10\ 000$).
- 18- Calculer la p-value comme étant $pv = \frac{h}{N}$.
- 19- Rejeter l'hypothèse nulle, si la p-value (pv) est plus petite que le niveau de confiance (5 %).

D'autre part, nous pouvons appliquer le test de ratio de vraisemblance. Ce test permet de comparer les résultats des deux modèles (Poisson versus binomial négatif). Il nous permet également de valider le résultat trouvé précédemment avec la p-value du test de χ^2 . Le ratio de vraisemblance est :

$$LR = -2 * (LL_{\text{Poisson}} - LL_{\text{Bin Neg}})$$

avec :

LL : la valeur du log de vraisemblance du modèle.

La valeur du ratio est comparée à $\chi^2(5\%,1) = 3,8414$. Ainsi, nous rejetons l'hypothèse nulle stipulant que le coefficient de sur dispersion est nul lorsque le ratio de vraisemblance excède la statistique de $\chi^2(5\%,1)$.

3.6.5. Résultat de l'estimation des paramètres

Le tableau 3-6 présente les résultats de l'estimation des distributions Poisson et binomiale négative pour chacun des six types de risque. Il est à remarquer que dans tous les cas, la distribution Poisson est rejetée par le test χ^2 . Ceci confirme nos attentes puisque les données présentent une surdispersion qu'il faut modéliser par une autre distribution. Ce constat met en cause toutes les études sur le calcul du capital opérationnel qui se basent sur le modèle Poisson pour l'estimation des fréquences des pertes. La distribution binomiale négative offre un bon ajustement aux nombres de pertes par jour ou par semaine.

Concernant la période de fréquence, nous remarquons que pour les types de risque de *fraude externe*, *fraude interne* et *gestion d'exécution et de la livraison des processus* (tableaux 3-6a, 3-6b, 3-6f) la modélisation est meilleure en choisissant les fréquences par semaine. Ceci est conforme à notre analyse de départ montrant un biais potentiel dans la comptabilisation de certaines pertes. Pour les autres types de risque, nous retenons les fréquences par jour en nous basant sur les valeurs des p-values du test de χ^2 (tableaux 3-6c, 3-6d, 3-6e).

D'autre part, le test du ratio de vraisemblance est supérieur à $\chi^2(5\%,1)$ dans tous les cas. Ce résultat montre que nous rejetons l'hypothèse nulle stipulant que le coefficient de surdispersion est nul. Ainsi, ce constat confirme les résultats trouvés précédemment affirmant le rejet du modèle Poisson. Nous retenons ainsi le modèle binomial négatif comme le modèle qui s'ajuste le mieux aux données de fréquences.

Nous avons procédé également à la correction des paramètres pour prendre en considération le nombre de pertes au-dessous du seuil, et ce, pour tous les types de risque excepté pour *fraude externe* et *fraude interne*. Il est à mentionner que nous avons fixé $F(s)$ d'une façon exogène en nous basant sur l'expertise et le jugement des dirigeants de la banque étudiée.

3.7. Calcul de la VaR opérationnel par type de risque

3.7.1. Agrégation des distributions avec les données internes

Nous rappelons que notre objectif est d'estimer la valeur à risque annuelle à un niveau de confiance de 99,9 % tel que recommandé par les autorités réglementaires. Nous avons déterminé précédemment la distribution qui modélise la sévérité des pertes opérationnelles par événement, et ce, pour chacun des types de risque. Dans un deuxième temps, nous avons déterminé la distribution qui s'ajuste le mieux aux données de fréquences. Il y a lieu, à ce stade d'analyse, d'agréger les deux distributions pour déterminer la distribution agrégée des pertes annuelles. Plusieurs méthodes d'agrégation, que nous avons déjà décrites auparavant, existent dans la littérature. Nous choisissons la méthode de simulation Monte Carlo pour dériver la distribution non paramétrique des pertes annuelles. Cette méthode n'est pas la plus rapide mais elle a l'avantage d'être assez précise.

La méthode a déjà été détaillée par Cruz (2002) dans le contexte du risque opérationnel. En effet, il s'agit de :

1. Générer un nombre n de pertes par jour ou par semaine selon la distribution de fréquence estimée;

2. Générer n montants de pertes X_i selon la distribution de sévérité estimée, et ce, par jour ou par semaine;
3. Répéter les étapes 1 et 2 pour $N=365$ (pertes par jour) ou $N=52$ (pertes par semaine). Faire la sommation de tous les montants X_i générés pour avoir le montant de perte annuelle;
4. Répéter les étapes 1, 2 et 3 un grand nombre de fois (1 000 000 par exemple) pour avoir la distribution des pertes annuelles;
5. La VaR est déterminée comme étant le quantile d'ordre 99,9 % de la distribution agrégée empirique.

La VaR est en effet définie comme étant :

$$\text{VaR} = F_{\text{agrégé}}^{-1}(0,999)$$

où :

$F_{\text{agrégé}}^{-1}$ est la distribution cumulative agrégée inverse

3.7.2. Comparaison de notre modèle avec le modèle standard

Nous définissons le modèle standard comme étant le modèle simple répandu dans la littérature et obtenu à partir de la distribution lognormale et la distribution Poisson. De plus aucune prise en considération des pertes manquantes n'est faite. En fait, nous avons choisi ce modèle, car il est fréquemment utilisé en pratique et ayant fait l'objet de plusieurs recherches. Nous allons donc tester le degré de robustesse de ce modèle en le comparant au modèle que nous avons développé. Nous nous attendons à ce que la mauvaise spécification des distributions ainsi que le non ajustement des distributions à l'ensemble des données font biaiser le calcul du capital requis.

Nous avons choisi uniquement trois types de risque pour mener cette analyse à savoir *fraude interne, emploi, pratiques et sécurité environnementale* et *fraude externe*. Ces types de risque vont nous permettre de faire des comparaisons entre notre modèle et le modèle standard. Nous présentons dans le tableau 3-7a les résultats de l'estimation non conditionnelle de la distribution lognormale. De plus, nous présentons les résultats de la modélisation du nombre de pertes par semaine par la distribution Poisson. Aucun test d'ajustement n'est fait puisqu'il s'agit du modèle standard.

Le tableau 3-7b présente les pourcentages de variation des résultats de la perte annuelle moyenne ainsi que la VaR (niveaux de confiance à 90; 95; 99 et 99,9%) du modèle standard par rapport au modèle développé lors de cette étude. Nous remarquons que les résultats sont très différents dans la plupart des cas. Le modèle standard sous estime énormément la moyenne annuelle des pertes et la VaR.

Pour le type de risque *fraude interne*, les données ne sont pas tronquées et la distribution de sévérité est la même dans les deux modèles. Ainsi, l'écart dans les résultats des deux modèles mesure l'impact de la mauvaise spécification de la distribution de fréquence, pour ce type de risque. Les pertes annuelles moyennes sont presque égales alors qu'un écart allant jusqu'à 2% a été constaté entre les VaR des deux modèles.

Pour le type de risque *emploi, pratiques et sécurité environnementale*, les données sont tronquées et la distribution de sévérité est la même dans les deux modèles. Ainsi, l'écart dans les résultats des deux modèles mesure l'impact de la mauvaise spécification de la distribution de fréquence et la non considération du seuil de troncature, pour ce type de risque. Ainsi, la sous-estimation de la perte annuelle moyenne (-5%) et de la VaR (-27%) montre l'importance de bien choisir la distribution de fréquence et de considérer les données au dessous du seuil.

Quant au type de risque *fraude externe*, les distributions de sévérité et de fréquences ne sont pas les mêmes dans les deux modèles. Les résultats vont donc mesurer l'impact du mauvais choix des distributions sur la moyenne annuelle des pertes et sur la VaR. En effet, la VaR à 99,9% calculée avec le modèle standard est sous-estimée de (82%) par rapport à celle calculée avec notre modèle. Ceci confirme donc nos attentes. Il est important de bien choisir les distributions qui s'ajustent aux données de pertes pour avoir un niveau de capital reflétant la vraie exposition au risque opérationnel d'une banque.

De plus, nous remarquons que la sous estimation est plus importante lorsque le seuil de confiance de la VaR est élevé. En effet, pour le type de risque *fraude externe*, la différence entre les moyennes des pertes annuelles des deux modèles est minime,

alors qu'elle est de l'ordre de 6% entre les VaR à 90% et atteint 82% entre les VaR à 99.9%. Ceci montre donc que le modèle standard sous estime l'épaisseur de la queue de la distribution agrégée.

Ainsi, cette analyse montre que la mauvaise spécification des distributions de sévérité et de fréquence entraîne des biais importants dans le calcul du capital. La VaR calculée avec le modèle standard ne reflète aucunement l'exposition réelle de la banque. De plus, la non considération des pertes non collectées conduit à une estimation incorrecte et biaisée du capital. Notre méthode est donc une bonne alternative au modèle standard puisqu'elle permet de bien décrire les queues des distributions de sévérité sans ignorer le corps.

3.7.3. Combinaison des données internes et données externes

Le contexte actuel du risque opérationnel impose l'utilisation des données externes pour le calcul de la valeur à risque. En effet, les pertes collectées ne reflètent pas l'exposition réelle de la banque puisque certaines pertes peu fréquentes, mais potentiellement lourdes ne sont pas nécessairement captées dans la base de données interne. Elles peuvent, par contre, influencer beaucoup le capital risque opérationnel. Ce constat est principalement expliqué du fait que l'historique des pertes est court et la collecte n'est pas encore exhaustive. Tous ces facteurs sont en faveur de la combinaison des pertes internes et externes. Comme nous l'avons bien expliqué au chapitre 1, nous devons procéder à une mise à l'échelle des pertes externes avant leur utilisation.

Nous avons déterminé les pertes externes mises à l'échelle de la banque canadienne étudiée ainsi que la distribution qui modélise leur sévérité dans le chapitre 2. En effet nous avons montré que le meilleur modèle décrivant le comportement de ces pertes est le GB2 fractionné. D'autre part, nous avons montré dans le premier chapitre comment il est possible de déterminer les paramètres de la distribution de fréquences qui permettent la mise à l'échelle et l'ajustement des données. Nous allons ainsi retenir le modèle binomial négatif avec composante de régression. Ce modèle présente une caractéristique avantageuse à savoir :

Théorème (Larsen et Marx, 2001)

Si $X_i \sim$ binomiale négative (r_i, p) indépendants pour $i=1,2,\dots, n$, alors $\sum_{i=1}^n X_i \sim$ binomiale négative $\left(\sum_{i=1}^n r_i, p\right)$.

Il en découle de ce théorème que si les fréquences des pertes de plus d'un million de dollars sont indépendantes et identiquement distribuées selon un modèle binomial négatif, alors les paramètres de la distribution binomiale négative annuelle sont $\left(\frac{r_i}{11}, p\right)$ avec (r_i, p) les paramètres de la binomiale négative déterminés dans le premier chapitre pour la modélisation du nombre de pertes sur un horizon de 11 années.

Connaissant la distribution de sévérité et de fréquence, nous allons les agréger de la même façon que nous avons fait pour les données internes. Nous avons mis en place l'algorithme suivant qui permet d'agréger les données internes et externes dans le but de déterminer la distribution empirique des pertes annuelles et d'en calculer le 99,9^{ème} centile. Comme nous désirons calculer la VaR par type de risque, nous répétons donc cet algorithme pour chacun d'eux.

1. Générer un nombre n de pertes par jour ou par semaine selon la distribution de fréquence des données internes;
2. Générer n montants de pertes X_i ($i=1 \dots n$) selon la distribution de sévérité estimée des données internes, et ce, par jour ou par semaine. Nous contraignons les pertes générées d'être au-dessous de l'équivalent d'un million de dollars américains;
3. Répéter les étapes 1 et 2 pour $N=365$ (pertes par jour) ou $N=52$ (pertes par semaine). Faire la sommation de tous les montants X_i générés pour avoir S_1 la partie de la perte annuelle provenant des montants de moins de 1M \$US;
4. Générer un nombre n^* de pertes par année selon la distribution Poisson modélisant les données externes;
5. Générer n^* montants de pertes X_i^* ($i=1 \dots n^*$) selon la distribution de sévérité des données externes. Nous contraignons les pertes générées d'être au-dessus de l'équivalent d'un million de dollars américains. La somme des montants

générés donne S_2 la partie de la perte annuelle provenant des montants de plus de 1M \$US;

6. La perte annuelle totale S est égale à $S_1 + S_2$;
7. Répéter les étapes 1 à 6 un grand nombre de fois (1 000 000 par exemple) pour avoir la distribution des pertes annuelles;
8. La VaR est calculée en prenant le 99.9^{ème} centile de la distribution empirique des pertes annuelles.

3.7.4. Détermination de la VaR

Le tableau 3-8 présente les résultats des pertes annuelles moyennes et la VaR calculées à des niveaux de confiance de 95%, 99% et 99,9%. Ces résultats tiennent compte des pertes internes collectées au sein de la banque ainsi que les pertes externes mises à l'échelle de l'institution financière à l'étude. Dans le but de conserver la confidentialité des informations de la banque étudiée, les résultats sont présentés comme un pourcentage du total des actifs de la banque étudiée à l'année 2004. Nous présentons les résultats par type de risque et nous montrons que le type de risque *clients, produits et pratiques commerciales* présente la valeur à risque opérationnel la plus importante. Ce constat confirme le résultat trouvé par Dutta et Perry (2006) lors d'une étude basée sur des données de 7 banques. Les types de risque *fraude externe* et *gestion d'exécution et de la livraison des processus* viennent respectivement en deuxième et troisième position en terme d'ampleur de la VaR. Le type de risque qui présente le moins de risque est *actifs tangibles, corporels et endommagés*. Il est à remarquer que ce classement des types de risque n'est pas respecté lorsque nous observons les valeurs des pertes annuelles moyennes.

Nous avons également construit un autre modèle en faisant l'ajustement et la mise à l'échelle des fréquences des pertes externes avec le modèle Poisson. De la même manière que nous avons généré des fréquences de pertes externes avec le modèle binomial négatif, nous allons le faire avec la distribution Poisson. Cette dernière présente une propriété avantageuse comme c'est le cas de la binomiale négative. En effet, lorsque les nombres annuels de pertes de plus de 1 million \$ sont indépendants

et suivent la même loi Poisson alors son paramètre annuel est égal à $\frac{\lambda}{11}$, avec λ le paramètre déterminé dans le premier chapitre pour la modélisation du nombre de pertes sur un horizon de 11 années.

Nous mesurons ainsi la variation de la perte annuelle moyenne et la VaR entre les deux modèles. Nous nous attendons à ce que le modèle construit à partir de la distribution Poisson pour la mise à l'échelle sous estime la VaR à un niveau de confiance élevé car le modèle binomial négatif présente une variance plus importante. En effet, la probabilité de générer des fréquences élevées avec la distribution binomiale négative est plus importante qu'avec le modèle Poisson.

Le tableau 3-9 présente les résultats de la comparaison des deux modèles. Ces résultats sont conformes à nos attentes. En effet, nous remarquons que la différence entre les pertes moyennes annuelles est faible comparativement aux VaR, ceci s'explique principalement par le fait que les deux distributions ont des moyennes égales. Toutefois, les VaR aux niveaux de 90% et 95% sont plus élevées lorsqu'elles sont calculées avec le modèle construit avec la distribution Poisson. En effet, comme la Poisson est caractérisée par l'équidispersion alors les fréquences sont générées autour de la moyenne, contrairement à la distribution binomiale négative. Les fréquences générées avec cette dernière sont effectivement plus dispersées ce qui explique les valeurs plus élevées des VaR au niveau de confiance de 99,9%.

3.7.5. Impact de la combinaison des données internes et externes sur les résultats

Nous considérons trois types de risque à savoir *actifs tangibles, corporels et endommagés, emploi, pratiques et environnement de travail et fraude interne* pour mettre en relief l'impact de l'intégration des données externes dans le calcul de la VaR. Nous avons choisi ces types de risque parce que nous n'avons pas de pertes extrêmes de plus d'un million de dollars de ces types dans la base interne. Ainsi, une comparaison entre les VaR calculées avec seulement les données internes et celles

calculées à partir de la combinaison des données internes et externes nous montrera l'importance des données externes pour compléter les queues des distributions.

Nous calculons la perte annuelle moyenne ainsi que la VaR à différents quantiles à partir d'un premier échantillon constitué de données de pertes internes seulement alors que le deuxième échantillon fait intégrer les données de pertes externes, et ce, pour les trois types de risque cités précédemment. Les résultats présentés au tableau 3-10 montrent que la VaR est largement sous estimée lorsque nous nous basons seulement sur les données internes. La sous estimation atteint 99% pour le type de risque *fraude interne* (VaR à 99,9%). En effet, les pertes reliées à ce type de risque peuvent être trop élevées telles que les pertes subies à Allied Irish Bank et Barings suite à des transactions non autorisées. Rappelons que l'historique des pertes pour ces trois types de risque ne contient pas d'événements extrêmes. L'inclusion d'événements de plus d'un million de dollars pouvant être subis permettra de mieux décrire les queues des distributions agrégées et reflétera ainsi le niveau des pertes non anticipées.

3.8. Conclusion

La méthode LDA vise à identifier les risques opérationnels et à modéliser la distribution des pertes de la banque. Le but est de mesurer l'exigence de fonds propres pour la couverture des pertes opérationnelles. Le régulateur, dans un souci d'assurer la stabilité et la sécurité du système financier, incite les grandes banques à opter progressivement pour l'approche AMA.

L'objectif de ce chapitre était de mettre en place une méthode robuste de mesure de capital opérationnel. Nous avons remis en question certaines pratiques assez fréquentes à savoir :

- l'utilisation des distributions lognormale et Poisson sans s'assurer de leur bon ajustement aux données.
- la non considération des pertes au-dessous du seuil de collecte.
- la non inclusion des pertes externes dans le calcul de la VaR

Nous avons donc présenté une méthode qui consiste à bien choisir la distribution de sévérité en s'assurant de tenir compte des pertes non collectées. Les résultats ont montré que dans la majorité des cas la distribution GB2 s'ajuste bien à l'ensemble des données ou au corps des distributions empiriques. Ainsi, la GB2 s'avère une bonne candidate à prendre en considération lors de la détermination de la distribution de sévérité des pertes opérationnelles. Après avoir été appliquée récemment dans plusieurs domaines de la finance, ce chapitre met en valeur son application pertinente dans la modélisation du risque opérationnel. Quant aux queues des distributions, nous avons choisi la distribution Pareto.

De plus, nous avons testé les distributions Poisson et binomiale négative pour modéliser les fréquences des pertes. Nous avons montré que dans aucun cas le modèle Poisson n'est retenu, contrairement au modèle binomial négatif. Afin de prendre en considération le nombre de pertes au-dessous du seuil de collecte, une correction des paramètres estimés a été faite.

Dans le but de montrer la robustesse de notre modèle, nous l'avons comparé au modèle standard construit à partir des distributions les plus utilisées (lognormale et Poisson) tout en ignorant le seuil de troncature. Les résultats ont montré que le modèle standard sous-estime énormément la valeur à risque opérationnel.

D'autre part, nous avons mis en place un algorithme afin de combiner les pertes internes et externes mises à l'échelle de la banque canadienne étudiée. Les résultats de la VaR reflètent mieux la réalité puisque les données internes n'incluent pas des pertes à ampleur très élevée. D'ailleurs, une analyse de l'impact de l'intégration des données externes sur la VaR a révélé que cette dernière est largement sous estimée lorsque les calculs sont basés uniquement sur les données internes. D'où l'importance de la combinaison des données internes et externes dans le calcul de la perte non anticipée d'une banque.

Tableau 3-1 : Statistiques descriptives

Nous présentons des statistiques descriptives des montants et les fréquences des pertes opérationnelles d'une banque canadienne sur la période de 1/11/2001 jusqu'au 31/10/2004

Tableau 3-1a : Ce tableau présente les statistiques descriptives des montants de pertes par événement et par type de risque sur une période de trois ans.

	ATCE*	CPPC*	EPSE*	FE*	FI*	GELP*
Moyenne	20 797\$	25 582\$	34 220\$	2 049\$	13 882\$	7 479\$
Médiane	7 728\$	3 169\$	12 520\$	677\$	2 975\$	2 000\$
Écart-type	30 110\$	76 304\$	54 868\$	13 272\$	71 602\$	33 012 \$
Coefficient d'aplatissement	9,25	47,66	12,54	2 722,29	78,57	341,00
Coefficient d'asymétrie	2,80	6,25	3,22	46,12	8,80	15,94
Maximum	157 138\$	819 717\$	334 034\$	1 003 045\$	645 700\$	863 876\$
Nombre d'observations	53	509	97	22 178	81	1 547

*ATCE :Actifs tangibles Corporels endommagés / CPPC : Clients, produits et pratiques commerciales / EPSE : Emploi, pratiques et sécurité environnementale / FE : Fraude externe / FI : Fraude interne / GELP : Gestion d'exécution et livraison des processus.

Tableau 3-1b : Ce tableau présente les statistiques descriptives des fréquences des pertes par jour et par type de risque sur une période de trois ans.

	ATCE*	CPPC*	EPSE*	FE*	FI*	GELP*
Moyenne	0,0484	0,4644	0,0885	20,2354	0,0739	1,4115
Médiane	0	0	0	18	0	1
Mode	0	0	0	0	0	0
Variance	0,0643	1,0617	0,1209	369,7491	0,1069	9,3693
Minimum	0	0	0	0	0	0
Maximum	3	14	4	226	4	66

*ATCE :Actifs tangibles Corporels endommagés / CPPC : Clients, produits et pratiques commerciales / EPSE : Emploi, pratiques et sécurité environnementale / FE : Fraude externe / FI : Fraude interne / GELP : Gestion d'exécution et livraison des processus.

Tableau 3-1c : Ce tableau présente les fréquences moyennes par type de risque selon le jour de la semaine.

	ATCE*	CPPC*	EPSE*	FE*	FI*	GELP*
Lundi	0,06	0,56	0,15	25,20	0,13	1,68
Mardi	0,11	0,52	0,12	28,49	0,08	1,79
Mercredi	0,03	0,63	0,08	29,19	0,10	1,64
Jeudi	0,04	0,72	0,14	28,44	0,10	1,82
Vendredi	0,07	0,72	0,09	26,41	0,10	2,56
Samedi	0,01	0,03	0,04	2,62	0,00	0,20
Dimanche	0,02	0,06	0,01	1,43	0,00	0,20

*ATCE :Actifs tangibles Corporels endommagés / CPPC : Clients, produits et pratiques commerciales / EPSE : Emploi, pratiques et sécurité environnementale / FE : Fraude externe / FI : Fraude interne / GELP : Gestion d'exécution et livraison des processus.

Tableau 3-1d : Ce tableau présente les statistiques descriptives sur les fréquences des pertes par semaine et par type de risque sur une période de trois ans.

	ATCE*	CPPC*	EPSE*	FE*	FI*	GELP*
Moyenne	0,3376	3,2420	0,6178	141,2611	0,5159	9,8535
Médiane	0	2	0	148	0	8
Mode	0	2	0	125	0	4
Variance	0,4302	10,3513	0,8530	2689,7070	0,7513	87,8566
Minimum	0	0	0	50	0	0
Maximum	3	17	4	447	4	80

*ATCE :Actifs tangibles Corporels endommagés / CPPC : Clients, produits et pratiques commerciales / EPSE : Emploi, pratiques et sécurité environnementale / FE : Fraude externe / FI : Fraude interne / GELP : Gestion d'exécution et livraison des processus.

Tableau 3-2 : Estimation des distributions de sévérité pour les types de risque FI, ATCE, EPSE.

Tableau 3-2a : Résultats de l'estimation des 4 distributions paramétriques testées pour le type de risque *fraude interne*. La première ligne présente les paramètres estimés, ensuite nous présentons les valeurs du log de vraisemblance. La dernière ligne présente les p-values des tests Kolmogorov-Smirnov avec *bootstrap* paramétrique. Nous disposons de 84 observations non tronquées.

	Exponentielle	Lognormale	Weibull	GB2
Paramètres	13881,70	8,03 1,36	6308,72 0,59	1,06 1001,21 2,35 1
Log de vraisemblance	-853,60	-790,48	-808,33	-788,48
p-value KS	0	0,37	0	0,25
p-value AD	0	0,44	0	0,27
p-value CvM	0	0,54	0	0,30

Tableau 3-2b : Résultats de l'estimation des distributions paramétriques pour le type de risque *actifs tangibles, corporels et endommagés*. La première ligne présente les paramètres estimés, ensuite nous présentons les valeurs du log de vraisemblance. La dernière ligne présente les p-values des tests Kolmogorov-Smirnov avec *bootstrap* paramétrique. Nous disposons de 53 observations de plus de s.

	Exponentielle	Lognormale	Weibull	GB2
Paramètres	19797,22	8,75 1,59	8316,94 0,52	0,04 113,55 446,36 373,93
Log de vraisemblance	-577,34	-569,18	-568,21	-569,20
p-value KS	0	0,46	NaN	0,46
p-value AD	0	0,22	NaN	0,25
p-value CvM	0	0,29	NaN	0,31

Tableau 3-2c : Résultats de l'estimation des distributions paramétriques pour le type de risque *emploi, pratiques et sécurité environnementale*. La première ligne présente les paramètres estimés, ensuite nous présentons les valeurs du log de vraisemblance. La dernière ligne présente les p-values des tests Kolmogorov-Smirnov avec *bootstrap* paramétrique. Nous disposons de 97 observations de plus de s.

	Exponentielle	Lognormale	Weibull	GB2
Paramètres	33220,14	9,40 1,48	17214,52 0,57	0,05 7,56 418,31 285,33
Log de vraisemblance	-1106,86	-1089,61	-1089,61	-1089,64
p-value KS	0	0,69	NaN	0,69
p-value AD	0	0,51	NaN	0,37
p-value CvM	0	0,45	NaN	0,48

Tableau 3-3: Estimation des distributions de sévérité pour le type de risque CPPC

Tableau 3-3a : Estimation de la queue de la distribution pour le type de risque *clients, produits et pratiques commerciales*. Nous prenons différents quantiles de la distribution empirique (première colonne) dont les seuils correspondants ainsi que la taille de l'échantillon sont indiqués respectivement dans la deuxième et la troisième colonne. Ensuite, nous estimons le paramètre de la distribution Pareto pour chacun des échantillons selon la formule 3. La dernière colonne du tableau donne la p-value du test du Kolmogorov-Smirnov avec *bootstrap* paramétrique. La ligne en gras et italique représente le quantile, le seuil et le nombre d'observations correspondants ainsi que le paramètre de la Pareto. Nous avons choisi cet échantillon parce qu'il nous donne la p-value du test d'ajustement la plus élevée.

Quantiles	Seuils	Taille de l'échantillon	Pareto			
			Coefficient	p-value KS	p-value AD	p-value CvM
0,65	6 100\$	179	0,6191	0	0	0
0,70	8 900\$	153	0,6760	0	0	0
0,75	11 500\$	128	0,6700	0	0	0
0,80	16 800\$	102	0,6902	0	0	0
0,85	36 900\$	77	1,010	0,11	0,05	0,04
<i>0,90</i>	<i>62 000\$</i>	<i>51</i>	<i>1,1810</i>	<i>0,37</i>	<i>0,14</i>	<i>0,14</i>

Tableau 3-3b : Résultats de l'estimation du corps de la distribution pour le type de risque *clients, produits et pratiques commerciales*. Il s'agit de pertes entre s et 62 000\$. L'échantillon comprend 461 observations. L'estimation des paramètres pour chacune des distributions est faite avec la méthode du maximum de vraisemblance selon la formule 5.

	Exponentielle	Lognormale	Weibull	GB2
Paramètres	6615,32	-70,05 11,42	11698,11 $1,3 \times 10^{-6}$	4,52 62,65 1 0,13
Log de vraisemblance	-4486,63	-4297,83	-4400,51	-4297,50
p-value KS	0	0	NaN	0,05
p-value AD	0	0	NaN	0,24
p-value CvM	0	0	NaN	0,01

Tableau 3-4 : Estimation des distributions de sévérité pour le type de risque GELP

Tableau 3-4a : Estimation de la queue de la distribution pour le type de risque *gestion d'exécution et de la livraison des processus*. Nous prenons différents quantiles de la distribution empirique (première colonne) dont les seuils correspondants ainsi que la taille de l'échantillon sont indiqués respectivement dans la deuxième et la troisième colonne. Ensuite, nous estimons le paramètre de la distribution Pareto pour chacun des échantillons selon la formule 3. La dernière colonne du tableau donne la p-value du test du Kolmogorov-Smirnov avec *bootstrap* paramétrique. La ligne en gras et italique représente le quantile, le seuil et le nombre d'observations correspondants ainsi que le paramètre de la Pareto. Nous avons choisi cet échantillon parce qu'il nous donne la p-value du test d'ajustement la plus élevée.

Quantiles	Seuils	Taille de l'échantillon	Pareto			
			Coefficient	p-value KS	p-value AD	p-value CvM
0,75	4 295\$	387	0,9916	0,37	0	0,19
0,80	5 536\$	309	1,0243	0,37	0,6	0,46
0,85	7 480\$	233	1,0537	0,55	0,4	0,48
0,90	10 600\$	155	1,0168	0,06	0,14	0,09
0,95	22 700\$	78	1,1901	0,48	0,09	0,14
<i>0,96</i>	<i>26 000\$</i>	<i>62</i>	<i>1,1065</i>	<i>0,74</i>	<i>0,21</i>	<i>0,49</i>

Tableau 3-4b : Résultats de l'estimation de l'aile gauche de la distribution pour le type de risque *gestion d'exécution et de la livraison des processus*. Il s'agit des pertes entre s et s' . L'échantillon comprend 800 observations. L'estimation des paramètres pour chacune des distributions est faite avec la méthode du maximum de vraisemblance selon la formule 5.

	Exponentielle	Lognormale	Weibull	GB2
Paramètres	801,62	-46,11 7,67	17505,54 $3,47 \times 10^{-7}$	15,50 301,38 4,53 0,06
Log de vraisemblance	-5469,42	-5459,73	-5473,60	-5459,70
p-value KS	0	NaN	NaN	0,08
p-value AD	0	NaN	NaN	0,24
p-value CvM	0	NaN	NaN	0,01

Tableau 3-4c : Résultats de l'estimation du corps de la distribution pour le type de risque *gestion d'exécution et de la livraison des processus*. Il s'agit de pertes entre s' et 26 000\$. L'échantillon comprend 685 observations. L'estimation des paramètres pour chacune des distributions est faite avec la méthode du maximum de vraisemblance selon la formule 5.

	Exponentielle	Lognormale	Weibull	GB2
Paramètres	4226,01	1,26 3,058	17258,5 $4,25 \times 10^{-6}$	1 9,18 87,50 0,95
Log de vraisemblance	-6388,39	-6342,14	-6450,22	-6341,99
p-value KS	0	0,27	NaN	0,38
p-value AD	0		NaN	0,38
p-value CvM	0		NaN	0,49

Tableau 3-5 : Estimation des distributions de sévérité pour le type de risque FE

Tableau 3-5a : Estimation de la queue de la distribution pour le type de risque *fraude externe*. Nous prenons différents quantiles de la distribution empirique (première colonne) dont les seuils correspondants ainsi que la taille de l'échantillon sont indiqués respectivement dans la deuxième et la troisième colonne. Ensuite, nous estimons le paramètre de la distribution Pareto pour chacun des échantillons selon la formule 3. La dernière colonne du tableau donne la p-value du test du Kolmogorov-Smirnov avec *bootstrap* paramétrique. La ligne en gras et italique représente le quantile, le seuil et le nombre d'observations correspondants ainsi que le paramètre de la Pareto. Nous avons choisi cet échantillon parce qu'il nous donne la p-value du test d'ajustement la plus élevée.

Quantiles	Seuils	Taille de l'échantillon	Pareto			
			Coefficient	p-value KS	p-value AD	p-value CvM
0,85	2 873\$	3326	1,4705	0,001	0	0
0,90	3 881\$	2218	1,5436	0	0	0
0,95	5 840\$	1109	1,4310	0,002	0,10	0
0,96	6 805\$	888	1,4249	0	0,03	0
0,97	8 275\$	665	1,4081	0	0	0
0,98	10 325\$	444	1,2711	0,001	0	0
0,99	18 300\$	222	1,3805	0	0	0
0,991	19 890\$	196	1,2815	0	0	0,01
0,992	19 890\$	178	1,2349	0,003	0	0,04
0,993	20 360\$	156	1,1114	0,14	0,14	0,14
0,994	23 900\$	134	1,1450	0,159	0,27	0,17
0,995	28 350\$	111	1,1550	0,107	0,13	0,10
<i>0,996</i>	<i>38 000\$</i>	<i>89</i>	<i>1,3298</i>	<i>0,57</i>	<i>0</i>	<i>0,34</i>
0,997	46 500\$	67	1,3017	0,567	0,07	0,23
0,998	59 000\$	45	1,1665	0,500	0,84	0,73

Tableau 3-5b : Résultats de l'estimation de l'aile gauche de la distribution pour le type de risque *fraude externe*. Il s'agit de pertes inférieures à 5 000\$. L'échantillon comprend 20725 observations. L'estimation des paramètres pour chacune des distributions est faite avec la méthode du maximum de vraisemblance selon la formule 5.

	Exponentielle	Lognormale	Weibull	GB2
Paramètres	1082,66	6,84 1,88	1078,48 0,82	0,68 13723,58 1,45 8,80
Log de vraisemblance	-164374,54	-164076,19	-163874,14	-163610,43
p-value KS	0	0	NaN	0,56
p-value AD	0	0	NaN	0,18
p-value CvM	0	0	NaN	0,61

Tableau 3-5c : Résultats de l'estimation du corps de la distribution pour le type de risque *fraude externe*. Il s'agit de pertes entre 5 000\$ et 38 000\$. L'échantillon comprend 1364 observations. L'estimation des paramètres pour chacune des distributions est faite avec la méthode du maximum de vraisemblance selon la formule 4

	Exponentielle	Lognormale	Weibull	GB2
Paramètres	4634,40	-20,16 4,22	171906,58 $2,98 \times 10^{-9}$	6,73 606,68 24,69 0,25
Log de vraisemblance	-12868,92	-12804,80	-13306,21	-12804,74
p-value KS	0	NaN	NaN	0,18
p-value AD	0	NaN	NaN	0,03
p-value CvM	0	NaN	NaN	0,11

Tableau 3-6 : Estimation des distributions de fréquences par type de risque

Nous présentons les paramètres estimés par maximum de vraisemblance pour chacun des types de risque, ainsi que la valeur de la fonction du log de vraisemblance et la p-value du test χ^2 . Nous présentons pour les quatre derniers types de risque la valeur des paramètres corrigés afin de tenir compte de la troncature des données.

Tableau 3-6a : Estimation de la distribution de fréquence pour le type de risque *fraude interne*

	Fréquences par jour		Fréquences par semaine	
	Poisson	Binomiale négative	Poisson	Binomiale négative
Paramètres	0,0739	0,1810 0,7100	0,5159	1,0034 0,6604
Log de vraisemblance	-304,7151	-287,7930	-158,2382	-152,6294
Ratio de vraisemblance	33,8441		11,2376	
p-value (test χ^2)	0	0,177	0,013	0,846

Tableau 3-6b : Estimation de la distribution de fréquence pour le type de risque *fraude externe*

	Fréquences par jour		Fréquences par semaine	
	Poisson	Binomiale négative	Poisson	Binomiale négative
Paramètres	20,2350	0,7760 0,0369	141,2611	8,7775 0,0585
Log de vraisemblance	-11808,4500	-4401,9390	-1893,8787	-827,9793
Ratio de vraisemblance	14813,0220		2131,7987	
p-value (test χ^2)	0	0	0	0,269

Tableau 3-6c : Estimation de la distribution de fréquence pour le type de risque *actifs tangibles, corporels et endommagés*

	Fréquences par jour		Fréquences par semaine	
	Poisson	Binomiale négative	Poisson	Binomiale négative
Paramètres	0,0484	0,1280 0,7260	0,3376	1,1366 0,7710
Paramètres corrigés	0,0569	0,1280 0,6925	0,3972	1,1366 0,7411
Log de vraisemblance	-220,1877	-208,9131	-120,7829	-118,6158
Ratio de vraisemblance	22,5492		4,3341	
p-value (test χ^2)	0,001	0,589	0,06	0,493

Tableau 3-6d : Estimation de la distribution de fréquence pour le type de risque *emploi, pratiques et sécurité environnementale*

	Fréquences par jour		Fréquences par semaine	
	Poisson	Binomiale négative	Poisson	Binomiale négative
Paramètres	0,0885	0,2304 0,7225	0,6170	1,6894 0,7322
Paramètres corrigés	0,0983	0,2304 0,7009	0,6856	1,6894 0,7110
Log de vraisemblance	-346,1778	-330,4865	-171,9059	-168,1748
Ratio de vraisemblance	31,3825		7,4622	
p-value (test χ^2)	0	0,459	0,026	0,416

Tableau 3-6e : Estimation de la distribution de fréquence pour le type de risque *clients, produits et pratiques commerciales*

	Fréquences par jour		Fréquences par semaine	
	Poisson	Binomiale négative	Poisson	Binomiale négative
Paramètres	0,4644	0,4282 0,4797	3,2420	1,5917 0,3293
Paramètres corrigés	0,5464	0,4282 0,4394	3,8141	1,5917 0,2944
Log de vraisemblance	-1114,0496	-984,0448	-426,8773	-360,0661
Ratio de vraisemblance	260,0097		133,6224	
p-value (test χ^2)	0	0,935	0	0,166

Tableau 3-6f : Estimation de la distribution de fréquence pour le type de risque *gestion d'exécution et de la livraison des processus*

	Fréquences par jour		Fréquences par semaine	
	Poisson	Binomiale négative	Poisson	Binomiale négative
Paramètres	1,4115	0,6131 0,3028	9,8535	2,0069 0,1692
Paramètres corrigés	2,3525	0,6131 0,2067	16,4225	2,0069 0,1089
Log de vraisemblance	-2363,4134	-1773,0805	-791,8502	-510,3689
Ratio de vraisemblance	1180,6659		562,9626	
p-value (test χ^2)	0	0,002	0	0,382

Tableau 3-7 : Comparaison entre notre modèle et le modèle standard.

Nous présentons le modèle standard pour lequel les modèles de sévérité et de fréquence sont pris par hypothèse comme étant lognormal et Poisson, respectivement. De plus aucune considération des pertes au-dessous du seuil de collecte n'est faite. Les résultats sont présentés pour les types de risque *actifs tangibles, corporels et endommagés, emploi, pratiques et sécurité environnementale et fraude interne*.

Tableau 3-7a : Nous présentons les paramètres estimés par maximum de vraisemblance de la distribution lognormale et la distribution Poisson; et ce, pour les types de risque *actifs tangibles, corporels et endommagés, emploi, pratiques et sécurité environnementale et fraude interne*.

Types de risque	Lognormale	Poisson
FI*	8,03 1,36	0,52
EPSE*	9,55 1,35	0,62
FE*	6,41 1,60	141,26

* FI : Fraude interne / EPSE : Emploi, pratiques et sécurité environnementale / FE : Fraude externe

Tableau 3-7b : Nous comparons les pertes annuelles moyennes ainsi que les VaR annuelles, aux niveaux de confiance de 90; 95; 99 et 99,9%, calculées à partir du modèle standard et celui que nous avons mis en place dans ce chapitre, sans les données externes; et ce, pour les types de risque *fraude interne, emploi, pratiques et sécurité environnementale et fraude externe*. Nous présentons la variation en pourcentage des résultats du modèle standard par rapport à ceux de notre modèle.

Types de risque	Variation du modèle standard par rapport à notre modèle (%)				
	Variation des pertes /an	Variation des VaR annuelles à 90%	Variation des VaR annuelles à 95%	Variation des VaR annuelles à 99%	Variation des VaR annuelles à 99.9%
FI*	-0,06	-2,06	-2,18	-1,68	-1,34
EPSE*	-4,67	-9,40	-11,89	-17,07	-27,32
FE*	0,95	-5,62	-13,50	-40,28	-81,59

* FI : Fraude interne / EPSE : Emploi, pratiques et sécurité environnementale / FE : Fraude externe

Tableau 3-8 : La VaR opérationnel par type de risque

Nous présentons les résultats de l'estimation des pertes annuelles moyennes et les valeurs à risque annuelles aux niveaux de confiance de 90%, 95%, 99% et 99,9%, et ce par type de risque. Ces résultats résultent de la combinaison des données internes et des données externes mises à l'échelle. Les résultats sont présentés comme un pourcentage du total des actifs de la banque étudiée, à l'année 2004.

Types de risque	Perte annuelle moyenne (%)	VaR annuelle à 90% (%)	VaR annuelle à 95% (%)	VaR annuelle à 99% (%)	VaR annuelle à 99,9% (%)
ATCE*	0,0005	0,0007	0,0009	0,0015	0,0331
CPPC*	0,0079	0,0064	0,0076	0,1123	0,3887
EPSE*	0,0016	0,0021	0,0024	0,0038	0,0774
FE*	0,0178	0,0182	0,0191	0,0727	0,2159
FI*	0,0005	0,0004	0,0005	0,0023	0,0760
GELP*	0,0064	0,0058	0,0066	0,0753	0,2614

*ATCE :Actifs tangibles Corporels endommagés / CPPC : Clients, produits et pratiques commerciales / EPSE : Emploi, pratiques et sécurité environnementale / FE : Fraude externe / FI : Fraude interne / GELP : Gestion d'exécution et livraison des processus

Tableau 3-9 : Comparaison entre la VaR opérationnel par type de risque calculée avec le modèle Poisson et celle calculée avec le modèle binomial négatif

Nous comparons dans ce tableau la perte annuelle moyenne et les VaR à plusieurs niveaux de confiance calculées en utilisant le modèle Poisson pour l'ajustement et la mise à l'échelle des fréquences des pertes externes avec celles calculées en utilisant le modèle binomial négatif pour l'ajustement et la mise à l'échelle des fréquences des pertes externes. Nous présentons les variations des statistiques calculées avec le modèle Poisson par rapport à celles calculées avec la distribution binomiale négative.

Types de risque	Perte annuelle moyenne (%)	VaR annuelle à 90% (%)	VaR annuelle à 95% (%)	VaR annuelle à 99% (%)	VaR annuelle à 99.9% (%)
ATCE*	0,23	0,08	0,16	0,56	-1,72
CPPC*	-2,95	20,27	327,43	-34,31	-70,19
EPSE*	-1,68	0,78	1,25	4,81	-11,92
FE*	-3,67	1,58	3,75	20,18	-11,51
FI*	-0,52	1,71	3,22	-6,14	-47,88
GELP*	-1,99	9,75	22,65	-17,53	-58,99

*ATCE :Actifs tangibles Corporels endommagés / CPPC : Clients, produits et pratiques commerciales / EPSE : Emploi, pratiques et sécurité environnementale / FE : Fraude externe / FI : Fraude interne / GELP : Gestion d'exécution et livraison des processus

Tableau 3-10 : Impact de l'intégration des données de pertes externes sur la VaR

Nous présentons les résultats de la comparaison entre les pertes annuelles moyennes et les valeurs à risque annuelles aux niveaux de confiance de 90%, 95%, 99% et 99,9%, calculées à partir des données internes et celles calculées à partir des données de pertes internes et externes .

	Variation du modèle avec données internes par rapport au modèle avec données internes et externes (%)				
Types de risque	Perte annuelle moyenne (%)	VaR annuelle à 90% (%)	VaR annuelle à 95% (%)	VaR annuelle à 99% (%)	VaR annuelle à 99.9% (%)
ATCE*	-11,83%	1,46%	3,11%	15,24%	-88,73%
EPSE*	-18,42%	0,67%	3,68%	-1,08%	-91,26%
FI*	-57,30%	-5,31%	-8,98%	-72,29%	-98,63%

*ATCE :Actifs tangibles Corporels endommagés / EPSE : Emploi, pratiques et sécurité environnementale / FI : Fraude interne.

CONCLUSION GÉNÉRALE

La gestion du risque opérationnel constitue une des préoccupations majeures des dirigeants des institutions bancaires pendant ces dernières années. En effet, plusieurs champs de sa gestion sont encore à explorer et à découvrir alors que les autorités réglementaires mettent de la pression pour une quantification rigoureuse du capital risque opérationnel. Dans cette thèse, nous avons répondu à plusieurs questions reliées à la mesure de la perte non anticipée. Cette recherche sera effectivement d'une grande utilité aux institutions bancaires qui désirent adopter une approche avancée de mesure de capital, puisqu'elle fournit tous les outils pour le calcul d'une mesure précise de la valeur à risque. Son application empirique pour le cas d'une banque canadienne montre quels sont les défis à relever lors de son implantation.

L'objectif de cette thèse était ainsi la quantification du risque opérationnel des institutions bancaires. Nous avons développé une méthode robuste pour mesurer la valeur à risque opérationnel qui reflétera l'exposition réelle au risque d'une banque. D'ailleurs, nous avons combiné des données de pertes internes et externes de plus d'un million de dollars afin de prendre en considération les pertes extrêmes qui ne sont pas nécessairement contenues dans la base interne, et ce, dans le but d'éviter une sous estimation du capital risque opérationnel.

Dans une première partie, nous avons corrigé le biais d'échelle dû à l'utilisation des données externes de pertes opérationnelles dans la mesure du capital avec une méthode avancée. Nous avons développé un modèle permettant de mettre les montants de pertes observées dans le secteur bancaire à l'échelle d'une banque donnée. Nous avons montré que la taille, le lieu, la ligne d'affaires où la perte a eu lieu ainsi que le type de risque sont des facteurs importants à prendre en considération dans le modèle de mise à l'échelle de la sévérité. Nous avons prouvé également que l'effet de la taille est peu important comparativement aux autres facteurs de la mise à échelle. Notre modèle vient donc améliorer les modèles existants dans la littérature qui se basent uniquement sur la taille. Par ailleurs, nous avons validé notre méthode en prenant le cas de la banque Merrill Lynch et nous avons constaté que les

statistiques des pertes observées de cette banque sont sensiblement proches de celles trouvées après mise à échelle.

Par ailleurs, nous avons mis en place un modèle de mise à l'échelle des fréquences de pertes externes de plus d'un million de dollars, et ce, dans le but de déterminer le nombre de pertes extrêmes mises à l'échelle qu'une banque pourrait avoir. Nous avons considéré des modèles de comptage tronqués avec composante de régression dont les variables sont la taille et la répartition géographique des activités d'une banque. Les résultats ont montré que le modèle binomial négatif domine le Poisson. Il sera donc retenu pour faire l'ajustement et la mise à l'échelle des fréquences des pertes externes.

Il est à mentionner que nous avons supposé que nous disposons de tous les événements de pertes de plus d'un million de dollars des banques citées dans la base externe. Ce qui n'est pas nécessairement le cas vu que seules les pertes qui ont été médiatisées et publiées sont rapportées dans la base. Il serait ainsi intéressant de voir si les principaux résultats de cette étude changent lorsqu'une base plus fiable serait disponible sur le marché.

Nous notons également que nous n'avons pas inclus des variables estimant la qualité de l'environnement de contrôle des banques dans le modèle de mise à l'échelle de la sévérité. Cette variable est très importante certes, mais aucune information caractérisant le niveau de contrôle n'est incluse dans la base externe.

La seconde partie constitue une exploration des méthodes d'ajustement des distributions paramétriques aux montants de pertes mises à l'échelle d'une banque canadienne. Nous avons effectivement testé des distributions usuelles à savoir la lognormale, l'exponentielle et la Weibull et une distribution à quatre paramètres la GB2. D'autre part, en validant l'existence d'une symétrie pondérée dans les données de pertes, nous avons construit un modèle GB2 fractionné. Les résultats ont montré que ce dernier offre un excellent ajustement aux données de montants de pertes. Un tel modèle permet ainsi de bien décrire le comportement des pertes opérationnelles et

d'avoir plus de précision et de rigueur à la mesure de la VaR que nous explorons dans la dernière partie.

Il serait intéressant de tester le degré d'ajustement de ce modèle aux données internes d'une banque. Toutefois, la grande particularité des données internes c'est qu'elles ne sont collectées qu'à partir d'un certain seuil de matérialité. La modélisation des données tronquées avec ce modèle pourrait être une extension intéressante.

La dernière partie de cette thèse porte sur le développement d'une valeur à risque opérationnel en nous basant sur la méthodologie actuarielle. Nous avons effectivement estimé les distributions de sévérité et de fréquence, tout en choisissant la distribution qui offre le meilleur ajustement. Nous avons également considéré les montants des pertes au-dessous du seuil de collecte ainsi que leur fréquence, et ce, pour une mesure plus correcte, non biaisée de la VaR. D'ailleurs, nous avons comparé notre modèle à un modèle standard, construit à partir des distributions fortement utilisées en pratique (lognormale et Poisson) et sans tenir compte du seuil de troncature. Les résultats ont été conformes à nos attentes puisque le modèle standard sous estime énormément la VaR opérationnel. Par ailleurs, nous avons proposé un algorithme permettant la combinaison des données de pertes internes et externes. De plus, Nous avons analysé l'impact de l'intégration des données externes sur la VaR. Les résultats ont effectivement montré l'importance de l'utilisation des données externes afin d'éviter une sous estimation substantielle de la perte non anticipée.

Ainsi, ce modèle, tel que développé et appliqué sur des données réelles d'une banque canadienne, peut être implanté à n'importe quelle institution financière. Cependant, il est à signaler qu'il est nécessaire d'avoir un nombre suffisant de données pour la modélisation paramétrique des distributions. Nous avons fixé ce nombre d'une façon arbitraire puisqu'il ne fait pas l'objet de la thèse. Une étude statistique permettant la détermination du nombre minimal requis pour la modélisation des distributions permettrait donc de justifier notre choix.

Nous rappelons que nous avons fait le calcul de la perte non anticipée pour seulement six types de risque. Pour le septième type de risque *perturbation des affaires et défaillances des systèmes*, le nombre d'observations est limité, ne nous permettant pas d'estimer la perte avec la méthode LDA. Nous pensons ainsi que d'autres méthodes basées sur des facteurs qualitatifs peuvent être appliquées telle que l'analyse de scénarios (Scandizzo, 2006). Il serait donc intéressant de voir comment cette méthode pourrait être implantée en pratique.

De plus, il serait également intéressant de calculer le capital agrégé de toute la banque. La prise en considération de la dépendance entre les types de risque permet de trouver la valeur exacte du capital agrégé. L'utilisation de la théorie des copules peut être utilisée à cette fin et peut donc constituer une avenue de recherche intéressante (Frachot, Roncalli et Salomon, 2004 et Chavez-Demoulin, Embrechts et Nešlehová, 2006).

Par ailleurs, l'étude de cas faite dans le troisième chapitre peut faire l'objet d'un business case. En effet, si nous disposons du capital risque opérationnel alloué à tous les types de risque incluant le type de risque *perturbation des affaires et défaillances des systèmes*, il sera possible de calculer le capital agrégé au niveau de la banque en utilisant la théorie des copules. Nous pouvons, par conséquent, comparer le capital trouvé avec la méthode avancée par rapport au capital calculé avec la méthode standardisée ou de base. Il est attendu que le capital estimé avec la méthode avancée soit moins élevé que celui trouvé avec les autres méthodes, et ce, lorsque la vraie dépendance entre les types de risque est bien évaluée.

BIBLIOGRAPHIE

- Abdous, B, K. Ghoudi et B. Rémillard (2003), “Nonparametric Weighted Symmetry Tests”, *The Canadian Journal of Statistics*, 31, 4, 357-381.
- Alexander, C. (2003), *Operational Risk: Regulation, Analysis and Management*, FT Prentice Hall, London.
- Artzner, P., F. Delbaen, J. M. Eber et D. Heath (1999), “Coherent Measures of Risk”, *Mathematical Finance*, 9, 203-228.
- Basel Committee on Banking Supervision (2001), *Operational Risk-Consultative Document*, Supporting Document to the New Basel Capital Accord.
- Basel Committee on Banking Supervision (2001), *Working Paper on the Regulatory Treatment of Operational Risk*.
- Basel Committee on Banking Supervision (2003), *Third Consultative Paper*, The New Basel Capital Accord.
- Baud, N., A. Frachot et T. Roncalli (2002), “How to Avoid Over-estimating Capital Charge for Operational Risk ?” Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais, France.
- Bee, M. (2006), “Estimating the Parameters in the Loss Distribution Approach : How Can We Deal with Truncated Data”, *The Advanced Measurement Approach to Operational Risk*, edited by E. Davis, Risk Books, London.
- Böcker et Klüppelberg (2005), “Operational VaR : a Closed-Form Approximation”, *Risk Magazine*, Décembre 2005, 90-93.
- Bookstaber, R. M. et J. B. McDonald (1987), “A General Distribution for Describing Security Price Returns.” *Journal of Business* 60, 401-424.

- Boyer, M., G. Dionne et C. Vanasse (1991), “Econometric Models of Accident Distributions”, in G. Dionne (ed.): *Contributions to Insurance Economics*, Norwell, MA: Kluwer Academic Publishers.
- Cameron, A. C. et P. K. Trivedi (1986) “Econometric Models Based on Count Data: Comparison and Applications of Some Estimators and Tests”, *Journal of Applied Econometrics* 1, 29-53.
- Cameron, A. C. et P. K. Trivedi (1998), *Regression Analysis of Count Data. Econometric Society Monographs*, 30.
- Cameron, A. C., P. K. Trivedi, F. Milne et J. Piggott (1988), “A Microeconomic Model of the Demand for Health Care and Health Insurance in Australia,” *Review of Economic Studies* 55, 85-106.
- Chapelle, A., Y. Crama, G. Hübner et JP. Peters (2004), “Basel II and Operational Risk: Implications for Risk Measurement and Management in the Financial Sector”, Working Paper 51, National Bank of Belgium.
- Chavez-Demoulin, V., P. Embrechts et J. Nešlehová (2006), “Quantitative Models for Operational Risk: Extremes, Dependence and Aggregation”, *Journal of Banking and Finance*, 30, 10, 2635-2658.
- Chernobai, A., C. Menn, S. T. Rachev et C. Trûck (2005c), “Estimation of Operational Value-at-Risk in the Presence of Minimum Collection Thresholds.” Technical Report, University of California Santa Barbara.
- Chernobai, A., C. Menn, S. T. Rachev, C. Trûck et M. Moscadelli (2006), “Treatment of Incomplete Data in the Field of Operational Risk : The Effects on Parameter Estimates, EL and UL Figures”, *The Advanced Measurement Approach to Operational Risk*, edited by E. Davis, Risk Books, London.
- Chernobai, A., C. Menn, S. Trûck et S. T. Rachev (2005a), “A Note on the Estimation of the Frequency and Severity Distribution of Operational Losses.” *Mathematical Scientist* 30, 2, 87-97.

- Chernobai, A. et S. Rachev (2006), “Applying Robust Methods to Operational risk Modeling.” *Journal of Operational Risk* 1, 1, 27-41.
- Chernobai, A., S. Rachev et F. Fabozzi (2005b). “Composite Goodness-of-Fit Tests for Left-Truncated Loss Samples.” Technical Report, University of California Santa Barbara .
- Cox, D. R. et P. A. W. Lewis (1966), *The Statistical Analysis of Series of Events*, Chapman and Hall, London.
- Cruz, M. G. (2002). *Modeling, Measuring and Hedging Operational Risk*, John Wiley & Sons, LTD, Chichester.
- Cummins, J. D., G. Dionne, J. B. McDonald et B. M. Pritchett (1990), “Applications of the GB2 Family of Distributions in Modelling Insurance Loss Processes”, *Insurance: Mathematics and Economics* 9, pp: 257-272.
- Cummins, J. D. et L. R. Freifelder (1978), “A Comparative Analysis of Alternative Maximum Probable Yearly Aggregate Loss Estimators”, *The Journal of Risk and Insurance*, 45, 1, 27-52.
- Cummins, J. D., C. M. Lewis et R. D. Phillips (1999), “Pricing Excess of Loss Reinsurance Contracts Against Catastrophic Loss.” In Kenneth Froot, ed., *The Financing of Catastrophe Risk* (Chicago : University of Chicago Press).
- Cummins, J. D. et L. J. Wiltbank (1983), “Estimating the Total Claims Distribution Using Multivariate Frequency and Severity Distribution”, *The Journal of Risk and Insurance*, 50, 3, 377-403.
- de Fontnouvelle, P., V. DeJesus-Rueff, J. Jordan et E. Rosengren (2003), “Capital and Risk : New Evidence on Implications of Large Operational Losses”, Working Paper, Federal Reserve Bank of Boston.

- de Fontnouvelle, P, J. Jordan et E. Rosengren (2004), “Implications of Alternative Operational Risk Modelling Techniques”, Working Paper, Federal Reserve Bank of Boston.
- Diebold, F. X., T. Schuermann et J. D. Stroughair (1998), “Pitfalls and Opportunities in The Use of Extreme Value Theory in Risk Management.” Working paper, The Wharton School.
- Dionne, G., Gagné, R., Gagnon, F., Vanasse, C. (1997), “Debt, Moral Hazard and Airline Safety: An Empirical Evidence”, *Journal of Econometrics* 79, 379-402.
- Dionne, G. et C. Vanasse (1989), “A Generalization of Actuarial Automobile Insurance Rating Models: the Negative Binomial Distribution with a Regression Component”, *Astin Bulletin* 19, 199-212.
- Dionne, G. et C. Vanasse (1992), “Automobile Insurance Ratemaking in the Presence of Asymmetrical Information”, *Journal of Applied Econometrics* 7, 2, 149-165.
- Dutta, K. K. et Babbel, D.F (2002a), “On Measuring Skewness and Kurtosis in Short Rate Distributions: The Case of the US Dollar London Inter Bank Offer Rate.” Working Paper, NERA.
- Dutta, K. et J. Perry (2006), “A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital”, Working Paper 06-13, Federal Reserve of Boston.
- Ebnother, S., P. Vanini, A. McNeil et P. Antolinez-Fehr (2001), “Modelling Operational Risk”. Working Paper.
- El Sayyad, G. M. (1973), “Bayesian and Classical analysis of Poisson Regression”, *Journal of the Royal Statistical Society, Series B*, 35, 445-451.
- Embrechts, P., H. Furrer et R. Kaufmann (2003), “Quantifying Regulatory Capital for Operational Risk”, *Derivatives Use, Trading and Regulation*, 9, 3, 217-233.

- Embrechts, R. Kaufmann et G. Samorodnitsky (2004) “Ruin Theory Revisited: Stochastic Models for Operational Risk.” In *Risk Management for Central Bank Foreign Reserves* (Eds. C. Bernadell et al.) European Central Bank, Frankfurt a.M., 243-261.
- Embrechts, P., C. Klüppelberg et T. Mikosch (1997), *Modelling Extremal Events for Insurance and Finance*, Springer Verlag, Berlin.
- Federal Reserve System, Office of the Comptroller of the Currency, Office of Thrift Supervision and Federal Deposit Insurance Corporation (2005), *Results of the 2004 Loss Data Collection Exercise for Operational Risk*.
- Frachot, A., O. Moudoulaud et T. Roncalli (2003). “Loss Distribution Approach in Practice.”, in *The Basel Handbook : A Guide for Financial Practitioners*, edited by Micheal Ong, Risk Books, 2004.
- Frachot, A et T. Roncalli (2002), “Mixing Internal and External Data for Managing Operational Risk”, Groupe de Recherche Opérationnelle, Crédit Lyonnais.
- Frachot, A., T. Roncalli et E. Salomon (2004). “The Correlation Problem in Operational Risk.”, Groupe de Recherche Opérationnelle, Crédit Lyonnais, France.
- Frome, E., M. Kutner et J. Beauchamp. (1973), “Regression Analysis of Poisson Distributed Data”, *Journal of the American Statistical Association*, 68, 935-940.
- Ghoudi, K. et B. Rémillard (1998). “Empirical Processes Based on Pseudo-Observations”. In *Asymptotic Methods in Probability and Statistics, A Volume in Honour of Miklos Csorgo: Proceedings of the International Conference (ICAMPS’97) held at Carleton University, Ottawa, ON, July 8-13, 1997* (B. Szyszkowicz,ed.), North-Holland, Amsterdam, 171-197.
- Gouriéroux, C. et A. Monfort (2002). “Pricing With Splines.” Série de Documents de Travail du CREST (Centre de Recherche en Économie et Statistique) 2002-50,

- Institut National des Statistiques et des Études Économiques (INSEE),
Malakoff, France.
- Gouriéroux, C., A. Monfort et A. Trognon. (1984), “Pseudo Maximum Likelihood Methods: Applications to Poisson Models”, *Econometrica* 52, 3, 701-720.
- Gurmu, S. (1991), “Tests for Detecting Overdispersion in the Positive Poisson Regression Model”, *Journal of Business and Economic Statistics* 9, 2, 215-222.
- Gurmu, S., et P. K. Trivedi. (1992), “Overdispersion Tests for Truncated Poisson Regression Models”, *Journal of Econometrics* 54, 347-370.
- Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw et W. A. Stahel. (1986), *Robust Statistics: The Approach based on Influence Functions*, Wiley, New York.
- Hartung, T. (2004), “Operational Risks: Modelling and Quantifying the Impact of Insurance Solutions”, Working Paper, Institute of Risk Management and Insurance Industry, Ludwig-Maximilians-University Munich, Germany.
- Hausman, J., B. Hall et Z. Griliches. (1984), “Econometrics Models for count Data with an Application to the Patents R. & D. Relationship”, *Econometrica* 52, 4, 909-938.
- Huber, P. J. (1981), *Robust Statistics*, Wiley, New York.
- Jorion, P. (2000), *Value-at-Risk: The New Benchmark for Managing Financial Risk*, Second Edition, McGraw-Hill, New York.
- King, J. L. (2001), *Operational Risk, Measurement and Modelling*, Wiley Finance, New York.
- Klugman, S. A, H. H. Panjer et G. E. Willmot (1998), *Loss Models, From Data to Decisions*, Wiley Series in Probability and Statistics, New York.
- Larsen, R. J. et M. L. Marx (2001), *An introduction to Mathematical Statistics and its Applications*, 3ème édition, Prentice Hill.

- Lefrançois, P. (1989). "Allowing for Asymmetry in Forecast Errors: Results from a Monte Carlo Study." *International Journal of Forecasting*, 5, 99-110
- Maddala, G. S. (1983), *Limited-dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.
- McDonald, J. B. (1984). "Some Generalized Functions of the Size Distribution of Income." *Econometrica* 52, 647-663.
- McDonald, J. B. et Y. J. Xu (1995). "A Generalization of the Beta Distribution with Applications." *Journal of Econometrics*, 66, 1, 133-152.
- McLachlan, G. et K. Basford (1988). *Mixture Models. Inference and Applications to Clustering*. Statistics: Textbooks and Monographs, 84. Marcel Dekker, Inc, New York.
- Medova, E. A. (2001). "Operational Risk Capital Allocation and Integration of Risks." Working Paper.
- Moscadelli, M. (2004), "The Modelling of Operational Risk: Experience With the Analysis of the Data Collected by the Basel Committee." Working Paper, Banca d'Italia, Italie.
- Na, H. S. (2004), *Analysing and Scaling Operational Risk*, Master Thesis, Erasmus University Rotterdam, Netherlands.
- Na, H. S., J. Van Den Berg, L. C. Miranda et M. Leipoldt (2006), "An Economic Model to Scale Operational Losses", *The journal of Operational Risk* 1, 2, 11-31.
- Nešlehová, J., P. Embrechts et V. Chavez-Demoulin (2006) "Infinite mean models and the LDA for operational risk" *Journal of Operational Risk*, 1, 1, 3-25
- Panjer, H. H. et G. E. Willmot (1992), *Insurance Risk Models*, Society of Actuaries, Schaumburg, Illinois.

Peters, J. P., Y. Crama et G. Hübner (2004), “An Algorithmic Approach for the Identification of Extreme Operational Losses Threshold”, Working Paper, Université de Liège, Belgique.

Ross, S. (2001), *Simulation*, Third Edition, MA: Academic Press, Boston.

Scandizzo, S. (2006), “Scenario Analysis in Operational Risk Management”, ”, *The Advanced Measurement Approach to Operational Risk*, edited by E. Davis, Risk Books, London.

Shih. J, A. Samad-Khan et P. Medapa (2000), “Is the Size of an Operational Loss Related to Firm Size”, *Operational Risk Magazine* 2, 1.

Tunaru, R., A. Kadam et G. Albota (2005) “ An Investigation of Parametric Risk Neutral Density Estimation”. Working paper. Cass Business School, City University London