

ESSAYS IN QUANTITATIVE RISK MANAGEMENT FOR
FINANCIAL REGULATION OF OPERATIONAL RISK
MODELS

PAVAN ARODA

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS
YORK UNIVERSITY
TORONTO, ONTARIO

NOVEMBER 2016

©PAVAN ARODA, 2016

Abstract

An extensive amount of evolving guidance and rules are provided to banks by financial regulators. A particular set of instructions outline requirements to calculate and set aside loss-absorbing regulatory capital to ensure the solvency of a bank. Mathematical models are typically used by banks to quantify sufficient amounts of capital. In this thesis, we explore areas that advance our knowledge in regulatory risk management.

In the first essay, we explore an aspect of operational risk loss modeling using scenario analysis. An actuarial modeling method is typically used to quantify a baseline capital value which is then layered with a judgemental component in order to account for and integrate what-if future potential losses into the model. We propose a method from digital signal processing using the convolution operator that views the problem of the blending of two signals. That is, a baseline loss distribution obtained from the modeling of frequency and severity of internal losses is combined with a probability distribution obtained from scenario responses to yield a final output that integrates both sets of information.

In the second essay, we revisit scenario analysis and the potential impact of catastrophic events to that of the enterprise level of a bank. We generalize an algorithm to account for multiple level of intensities of events together with unique loss profiles depending on the business units effected.

In the third essay, we investigate the problem of allocating aggregate capital across sub-portfolios in a fair manner when there are various forms of interdependencies. Relevant to areas of market, credit and operational risk, the multivariate shortfall allocation problem quantifies the optimal amount of capital needed to ensure that the expected loss under a convex loss penalty function remains bounded by a threshold. We first provide an application of the existing methodology to a subset of high frequency loss cells. Lastly, we provide an extension using copula models which allows for the modeling of joint fat-tailed events or asymmetries in the underlying process.

Acknowledgments

I would like to express my sincere gratitude to my supervisor Dr. Huaxiong Huang whose support had never wavered through this long process. In my opinion he embodies what a true applied mathematician should be and without his courage to keep tackling the unknown, the freedom we had to explore and discover along the way allowed this thesis to take this unique shape. I would also like to thank my committee member Dr. Aziz Guergachi for his insight and passion and use of every form of communication to connect so that there was never a time that I felt unguided or alone. Our serendipitous meeting was instrumental to the completion of this degree. I would also like to thank my committee member Dr. Melanie Cao. Her ability to get to the root of a problem and communicate a clear way forward was extremely helpful. Also, her expertise in finance and her exacting standards helped shape this thesis. I would also like to thank every member of the examining committee.

I would like to thank the Office of the Superintendent of Financial Institutions (OSFI) for providing an environment that allowed for the completion of this degree. By having such a wide exposure to the financial sector, I was able to gain insight into the many facets of banking and picked certain areas where I believed research could be advanced. There are numerous individuals that warrant special mention at OSFI. I would like to thank Romana Mizdrak for her encouragement and support through the entire process. I would like to thank the original group that was tasked to understand this thing called operational risk. The whiteboard sessions between Dr. Leopold Fossi Talom, Polly Han and myself during the early days was where we really got the

the bottom of things. Before he left OSFI to go work in The Central Bank of The Bahamas, I want to thank Abhilash Bhachech for this encouragement to finish my degree and to explore operational risk. I want to thank Dr. Jean-Paul Haddad for our whiteboard sessions and our spirited conversations on the most important risk types. I would also like to thank Jason Au and Fred Han (now at Royal Bank of Canada) for their numerous discussions and insight. I enjoyed our conversations where we tried to learn the complexity of banking but also managed to find time for light-hearted conversations. I also wanted to thank individuals spread across the world for whom I never met. Thank you to Dr. Riaan de Jongh who put me in touch with Dr. Jerry Oglesby and Dr. André de Waal from SAS and through an agreement and academic partnership, arranged the use operational risk data. Your helpfulness having never met me in person was greatly appreciated.

From my earlier years, I still attribute my desire to forge ahead in mathematics from the teachings of Dr. Herb Kunze. The time we spent together during my Masters was pivotal in giving me direction and allowed me to develop a career that will carry me through life. I still am indebted to him to this day. On the same note, I thank Luk Arbuckle and Kris Heidler for their friendship and their wisdom that has carried on beyond our Masters. I would also like thank Dr. Markiyana Sloboda from the time of my Masters. If we did not run into each other many years later and if it was not for the subsequent introduction to Dr. Guergachi, I am not sure if I would have rekindled my doctoral studies.

To my family, it is impossible to thank you enough for your continued support through this long journey. To my mother, Asha, none of this made sense without you. You gave me life and showed me where I came from. At times, you took on roles as mother, father, friend and teacher. I love you dearly.

Table of Contents

Abstract	ii
Acknowledgments	iv
Table of Contents	vi
List of Tables	ix
List of Figures	xi
Acronyms	xii
1 Introduction	1
1.1 Overview	1
1.2 Outline of Thesis	3
2 Preliminaries	5
2.1 Financial Regulation of Market, Credit and Operational Risk	5
2.2 Importance of Operational Risk	13
2.3 Basic Indicator Approach and The Standardized Approach	17
2.4 Advanced Measurement Approach	18
2.4.1 Element 1: Internal Loss Data	20
2.4.2 Element 2: External Loss Data	22
2.4.3 Loss Distribution Approach	24
2.4.4 Calibration and Selection	29
2.4.5 Loss Aggregation Approaches	34
2.4.6 Correlation, Diversification	36
2.4.7 Element 3: Scenario Analysis	40
2.4.8 Element 4: Business Environment and Internal Control Factors	44

3	Scenario Integration in AMA modeling	48
3.1	Overview	48
3.2	Scenario Formulation	49
3.3	Literature Review	52
3.4	Convolution to Integrate Scenarios	56
3.4.1	Convolution Operation	58
3.4.2	Scenario Distribution Formulation	61
3.4.3	Application of Convolution Method	66
3.5	Summary	70
4	Enterprise-Wide Scenario Analysis	73
4.1	Overview	73
4.2	Literature Review	79
4.3	Frequency Quantification	83
4.4	Severity Quantification	88
4.5	Monte Carlo Simulation	90
4.6	Example: Two Catastrophic Scenarios	92
4.7	Scenario Integration Application	99
4.8	Capital Allocation	100
4.9	Summary	103
5	Optimal Capital Allocation to Sub-portfolios with Dependency	104
5.1	Overview	104
5.2	Literature Review	107
5.3	Shortfall Risk Allocation	117
5.3.1	Risk Measures Represented as Shortfall Risk	119
5.3.2	Multivariate Shortfall Risk Allocation with Systemic Risk	120
5.4	Numerical Replication	130
5.5	Application to Operational Risk Losses	140
5.6	Extension using Copulas	158
5.7	Summary	166
6	Conclusion and Future Work	168
7	References	171
8	Appendix A: Probability Distributions	183
8.1	Frequency Distributions	183
8.2	Severity Probability Distributions	184

9	Appendix B: Loss Aggregation	186
9.1	Single Loss Approximation	186
9.2	Panjer Recursion	187
9.3	Fast Fourier Transform	189

List of Tables

2.1	ASRF inputs	10
2.2	Common Equity Tier 1 qualifying instruments	16
2.3	Components of capital ratio	16
2.4	TSA weights across business lines	18
2.5	Gross income by business line weighted by β	19
2.6	Matrix for recording operational risk losses	21
2.7	Business Line/Event Type mapping	22
2.8	Retail banking/external fraud simulation	22
2.9	Retail banking/execution, delivery and process management simulation	23
2.10	Vendor databases	24
2.11	Consortia databases	24
2.12	Business Environment factors	45
2.13	Internal Control factors	45
3.14	Elements needed to perform baseline LDA and scenario convolution	67
3.15	Champion fit	69
3.16	VaR from scenario model and VaR from baseline and scenario model	70
4.17	Large operational risk losses	75
4.18	Contrast between workshop and new contribution.	78
4.19	Summary frequencies and probabilities of earthquake predictions	84
4.20	Summary Poisson parameters for corresponding earthquake magnitudes	85
4.21	Summary by magnitude ranges	87
4.22	New York flood parameters	94
4.23	California earthquake parameters	94
4.24	New York flood full information for single UoM	95
4.25	California earthquake full information for single UoM	95
4.26	New York flood UoM loss estimates	96
4.27	California earthquake UoM loss estimates	97
4.28	New York flood UoM parameter estimates	97

4.29	California earthquake UoM parameter estimates	98
4.30	Individual UoM VaR	98
4.31	Allocated capital per UoM	102
5.32	Bivariate case with systemic risk $\alpha = 0$	132
5.33	Bivariate case with systemic risk $\alpha = 0$ averaged over $N = 30$ runs	135
5.34	Bivariate case with systemic risk $\alpha = 0$ with varying bounds c	136
5.35	Bivariate case with $\alpha = 1$ using Monte Carlo with $M = 2$ million	138
5.36	Trivariate case with $\alpha = 0$ using Monte Carlo with $M = 2$ million	139
5.37	Trivariate case with $\alpha = 1$ using Monte Carlo with $M = 2$ million	140
5.38	Number of losses by UoM	142
5.39	Aggregate loss amounts by UoM	142
5.40	North American loss count data	144
5.41	High frequency operational loss count data	145
5.42	Mapping BL/ET to UoM sub-portfolios	145
5.43	Systemic risk $\alpha = 0$, bound $c = \$16$ billion	148
5.44	Systemic risk $\alpha = 0$, bound $c = \$16$ billion, non-negative allocation	150
5.45	Systemic risk $\alpha = 0$, varying bound c	150
5.46	Systemic risk $\alpha = 0$, varying bound c , non-negative allocation	151
5.47	Monthly frequency correlation matrix	153
5.48	Matrix of p -values	153
5.49	Monthly frequency correlation matrix filtering for statistical significance	153
5.50	Annual frequency correlation matrix	154
5.51	Matrix of p -values corresponding to annual correlation matrix	154
5.52	Annual frequency correlation matrix filtering for statistical significance	154
5.53	Systemic risk $\alpha = 1$, bound $c = \$16$ billion	156
5.54	Systemic risk $\alpha = 1$, varying bound c	156
5.55	Systemic risk $\alpha = 1$, bound $c = \$16$ billion, calibrations at quarter end	157
5.56	Bivariate case with $\alpha = 1$ and various copula dependencies	166

List of Figures

2.1	Loss regions	9
2.2	Identical marginal distributions with different dependence structures .	40
2.3	Quantiles for which AMA elements impact	47
4.4	Single vs. multiple UoM scenario	76
4.5	Cumulative probabilities for occurrence of earthquake	86
5.6	Linear relationship for m^* vs. c	136

Acronyms

ABA	American Bankers Association
AD	Anderson-Darling
AIC	Akaike Information Criterion
AM	Asset Management
AMA	Advanced Measurement Approach
AS	Agency Services
ASRF	Asymptotic Single Risk Factor
BCBS	Basel Committee on Banking Supervision
BCP	Business Continuity Planning
BEICF	Business Environment and Internal Control Factors
BIA	Basic Indicator Approach
BIC	Bayesian Information Criterion
BL	Business Line
CB	Commercial Banking
CCP	Central Clearing Counterparty
CDF	Cumulative Distribution Function
CDO	Collateralized Debt Obligations
CET1	Common Equity Tier 1
CF	Corporate Finance
CFTC	Commodity Futures Trading Commission
CPBP	Clients Products and Business Practices
CPI	Consumer Price Index

CSA	Control Self Assessment
CV	Coefficient of Variation
CVM	Cramér-von Mises
DFT	Discrete Fourier Transform
DPA	Damage to Physical Assets
EAD	Exposure At Default
EDPM	Execution Delivery Process Management
EF	External Fraud
EL	Expected Loss
ELD	External Loss Data
EPWS	Employment Practices and Workplace Safety
ES	Expected Shortfall
ET	Event Type
EVT	Extreme Value Theory
FFT	Fast Fourier Transform
GPD	Generalized Pareto Distribution
i.i.d.	Independent and Identically Distributed
IDFT	Inverse Discrete Fourier Transform
IF	Internal Fraud
ILD	Internal Loss Data
KRI	Key Risk Indicator
KS	Kolmogorov-Smirnov
LDA	Loss Distribution Approach
LGD	Loss Given Default
LGM	Loss Generating Mechanism
LIBOR	London Interbank Offered Rate
LN	Lognormal
MLE	Maximum Likelihood Estimation
OpCaR	Operational Risk Capital-at-Risk

ORX	Operational Risk Exchange
OTC	Over The Counter
P&L	Profit and Loss
PD	Probability of Default
PDF	Probability Density Function
PMF	Probability Mass Function
PO	Poisson
PS	Payment and Settlement
RAROC	Risk Adjusted Return On Capital
RB	Retail Banking
RBR	Retail Brokerage
RWA	Risk Weighted Asset
SME	Subject Matter Expert
TS	Trading and Sales
TSA	The Standardized Approach
UCERF	Uniform California Earthquake Rupture Forecast
UL	Unexpected Loss
UoM	Unit of Measure
USD	United States Dollar
USGS	United States Geological Survey
VaR	Value-at-Risk

1 Introduction

1.1 Overview

Mathematical modeling in financial institutions has come under increasingly amounts of scrutiny, especially since the 2007/2008 global financial crisis (BCBS (2009) [21], BCBS (2011) [24], BCBS (2013) [27], BCBS (2014) [29]). Slowly as the world evolves, there are different camps of thinking. Some believe that with better technology and advanced theory and research, that models used within the financial sector can be fine-tuned to better predict future events (Andersen *et al* (2007) [5], Spronk and Hallerbach (1997) [107]). Others are starting to accept that models, by their nature, are descriptions or proposals of some underlying process and have to be used with caution. These individuals acknowledge there are limitations inherent in modeling that do not accurately address the nonlinear behaviour of the financial markets nor do one set of models work during times of financial stress (Danielsson (2002) [48], Derman (2011) [54]). With these polarized views, one may even want to choose not to believe in the benefit of any models and believe everything should be simplified. However, digging deeper we see that financial models are everywhere and even the simplest set of calculations can be defined as a model.

Financial institutions, specifically banks, use numerous mathematical models in every facet of their organization. For example, traders rely on a series of models to help price and value numerous financial instruments in order to trade within the market place. Whether it is the underlying security such as foreign exchange rate, equity,

commodity, interest rate, credit or a derivative security such as an option referencing an underlier, pricing alone is a complex task. Other parts of a bank may require models to extend credit or loans to other institutions, individuals or to any type of obligor. The ability to ascertain and assimilate relevant information on an entity is of paramount importance in order to price in the appropriate margin or profit that is to be expected.

Then there are the vast number of models used to risk-manage the activities that a bank undertakes. Whether used internally to inform the business side of the bank or used to inform the risk management side of a bank, the reliance on models does not diminish. There also are other purposes for which risk management models fall under much scrutiny. This is where the notion of financial regulation puts the issue under a microscope and does so from an arm's-length. In order for a bank to operate within a particular country, they are subject to regular reporting of risk metrics, data, and detailed information to an independent financial regulator. One such area of interest is the branch of models and risk measures used to quantify regulatory capital. At a very high level, money/assets/capital or something of value should be kept aside in order to absorb losses that are to be expected during the normal course of business. For those losses that are unexpected, a best-guess with the use of models are used to quantify the amount of capital to set aside to absorb those potential losses. This idea of solvency is of utmost importance to the health of the financial system as well as to the economy in general as stated by Berger *et al* (1995) [30].

Banks are allowed, through regulatory guidance, to set capital requirements either through standardized formulas or internally-derived models. Risk management models built internally used to set capital requirements are formally approved by a regulator. As regulatory guidance changes and matures, so do the models in lock-step. Whereas much attention in the previous years has been placed on market and credit risk, little emphasis has been placed on operational risk. In our work, much of our

attention will be focused on the advanced method to set capital requirements for operational risk. In this thesis, our contribution starts by highlighting facets of operational risk modeling that are subject to debate or have had very little attention and propose new ways of tackling the problems. Our goal is to advance the theory and application of mathematical tools used in operational risk modeling. The implication of investigating such risks is that it imparts an enterprise-wide effect and so we also address the impact to the organization as a whole. Moreover, the formulation of certain problems are also applicable for wider risk management purposes and hence have importance not only for regulatory reasons, but also influence internal business judgement.

1.2 Outline of Thesis

The remainder of this thesis is organized as follows. In Chapter 2, we define the main risk types a bank is exposed to and focus the discussion on operational risk. We approach the problem from a risk management perspective and introduce the methods used to set regulatory capital requirements. We cover the current state of the modeled method used to quantify operational risk called the Advanced Measurement Approach (AMA), dissect its elements, and motivate areas for research. In Chapter 3, we postulate a new method to integrate expert opinion via scenario analysis into a baseline model calibrated on loss data. We present a literature review covering possible integration techniques and introduce a procedure using the convolution operator. After the necessary theory is introduced, an example is provided. In Chapter 4, we investigate an area of scenario quantification that has attracted minimal attention - that of the enterprise-wide/catastrophic scenarios. We present a literature review showing the sparse attention received by this type of scenario and how the industry has concerned itself with a more granular approach. We generalize an algorithm to quantify the aggregate risk conditioned on a particular type and intensity of scenario and then show the mechanics of the method in an example. We also draw the parallel

that the output of the algorithm could also be used in the context of the previous chapter and hence provide an extension over the original framework. We complete the problem by allocating enterprise-wide calculated capital to granular levels of businesses. In Chapter 5, we study the topic of shortfall risk allocation to sub-portfolios with various forms of interdependencies. The problem is formulated as an optimization problem where the goal is to determine the minimum amount of capital to offset the expected loss of each sub-portfolio in a financial network under a convex penalty function. The framework is applied to a set of high frequency loss cells in an operational risk context. In addition, we extend the problem to include the case where various forms of dependencies are found to be inherent in the system and incorporate this modeling behaviour with copulas. In Chapter 6, we conclude and discuss further areas of research. In the Appendices, we provide formulas for commonly used probability distributions relevant for our work and also provide alternate methods to compute the aggregate loss distribution.

2 Preliminaries

2.1 Financial Regulation of Market, Credit and Operational Risk

The regulation of financial institutions is a process that appears to be ever-green and requires collaboration among experts to set standards and rules. The Bank for International Settlements (BIS) is the world's oldest international financial organization whose mission is to serve central banks in their pursuit of monetary and financial stability and to foster international cooperation. A specialized committee belonging to the BIS called the Basel Committee on Banking Supervision (BCBS) is responsible for setting standards for the prudential regulation of banks. Its mandate is to strengthen the regulation, supervision and practices of banks worldwide with the purpose of enhancing financial stability as stated in a charter from BCBS (2013) [26]. The regulatory rules and guidance prescribed by independent regulators are subject to interpretation and best practice measures among the financial community.

The BCBS (2006) [19] defines market risk as the risk of losses in on and off-balance-sheet positions arising from movements in market prices. The risks subject to this requirement are those pertaining to interest rate related instruments and equities in the trading book and foreign exchange and commodities risk throughout the bank. Credit risk is most simply defined as the potential that a bank borrower or counter-

party will fail to meet its obligations in accordance with agreed terms (e.g. mortgage term loan, revolving credit card). Operational risk is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk. As a more generic definition, operational risk could be seen as everything that is neither market nor credit risk.

In 1988, the BCBS compiled a set of international banking regulations, now known as the Basel I Accord (Basel I (1988) [15]), with the intention of introducing a set of credit risk capital requirements for financial institutions. This was introduced because during the 1980's, banks were lending extensively, countries were indebted, and the risk that a loss would occur because a borrower would not fulfill its obligations was high. The capital rules were simplistic in their assumptions and sophistication required to calculate risk metrics. In 1996, there was a growing concern for market risk and the concern for the level of exposure of financial institutions to this risk type. The BCBS then amended the Basel I Accord to include market risk capital requirements as seen in Basel (1996) [16]. In subsequent years, there was a realization of an overestimation of credit risk through the application of the capital rules and banks began to lobby the BCBS for permission to use more sophisticated methods to measure credit risk. Moreover, the international community began to believe that the Accord needed substantial revisions and that the new revision should include operational risk as part of the financial risks reflected in the requirements. The change came in the form of Basel (2006) [19] with an initial publication in 2004 with slight revisions up until 2006. The new Accord was based on a three-pillar framework. Pillar I referred to the requirement that banks needed to calculate a minimum capital charge for market, credit and operational risk. Pillar II referred to the internal supervisory review process that ensured that a bank's capital level was sufficient to cover its exposure to overall risk not captured under Pillar I. Pillar III aimed to provide supervisory expectations and minimum requirements for both the qualitative and quantitative

disclosure by banks to the public. An expeditious update was introduced via BCBS (2009) [21] with a slightly different naming convention termed Basel 2.5. The update saw revisions to the market risk framework to rectify deficiencies highlighted in the 2007/2008 global financial crisis. A more comprehensive change came in the form of Basel III via Basel (2011) [24]. Among some of the strengthening measures introduced in order to promote more resilient banks and banking systems were (i) raising the quality, consistency and transparency of the capital base, (ii) enhancing risk coverage, (iii) supplementing the risk-based capital requirement with a leverage ratio, (iv) reducing procyclicality and promoting countercyclical buffers, (v) addressing systemic risk and interconnectedness. With such exhaustive changes, Basel III phase-in arrangements were made to allow banks to get up to speed. The timeline for phase-in started in 2013 and stretches out to 2019. However with all the changes, the three pillar framework is still adopted and for our discussion on the banking side, the standards that must be followed for Pillar I translate to minimum regulatory capital that must be set aside by banks in order to adequately reflect the risk-taking activities that are taken. It is here that the concept of holding the “right” amount of capital is imperative from a bank’s perspective.

Banks are usually publicly-traded companies. That is among the many other reasons, they list their institution on a stock market exchange to raise money from various individuals. If an individual chooses to purchase a set of shares, the shares allow for partial ownership into the organization. In turn, the bank should be able to create value for that individual and make a return on their money (sometimes referred to as shareholder value). Now if a bank, for regulatory purposes, is told to hold a certain amount of capital (assets) to buffer against unaccounted losses, they are not purely putting that money to work for their shareholders. A bank creates value by increasing assets - for example by opening branches, making investments, or acquiring other banks to name a few. From the view point of the shareholder, they may want the bank to hold less capital in order to fully realize the potential of

their investment. However, capital is not a bad thing. Market, credit or operational loss events do happen and hence shareholders should support the concept of enough capital to be kept on reserve in order to cover those unexpected losses. For example, in 2011 Wells Fargo was fined by the US Federal Reserve board for \$85 million due to mortgage fraud dealings as seen in Operational Risk and Regulation (2011) [97]. Also, in 2013, the Royal Bank of Scotland Group was fined by the US Commodity Futures Trading Commission (CFTC) for \$325 million due to manipulation of the London Interbank Offered Rate (LIBOR) as seen in Operational Risk and Regulation (2013) [98]. Hence, a middle ground of capital requirements must be sought out as stated by Berger *et al* (1995) [30]. Typically under the Pillar I framework, banks have a choice to set capital requirements using a standardized or advanced method for market, credit or operational risk. Standardized methods are typically risk-insensitive and are capital-intensive. The BCBS sets rules and formulas for each risk type and differentiates on a more granular basis the nuances of each asset classes. However, as banks grow and become more sophisticated, they are incentivized to proceed along a path to advanced or sometimes referred to as “modeled methods”. The aim of capital requirements depending on risk type is to seek assurance for loss-absorbing capabilities for unexpected losses. In some instances, even the general provisions in the expected loss region should be capitalized for and added to the unexpected loss value. We depict this in Figure 2.1.

The most popular method to quantify market risk and hence market risk capital requirements is the Value-at-Risk (VaR) method. VaR aims to satisfy the statement, “I am X percent certain there will not be a loss of more than V dollars in the next N days”. For market risk capital requirements, X = 99 %, N = 10 days (1 day VaR usually scaled to 10 days by the square root of time rule). The underlying modeling approach used to characterize the random process will yield a unique value of V in dollars. For example, a trading desk may have a portfolio of financial instruments whose value fluctuates with market conditions aside from the trading strategy and its payoff structure. To compute the 1-day VaR, there are three popular methods:

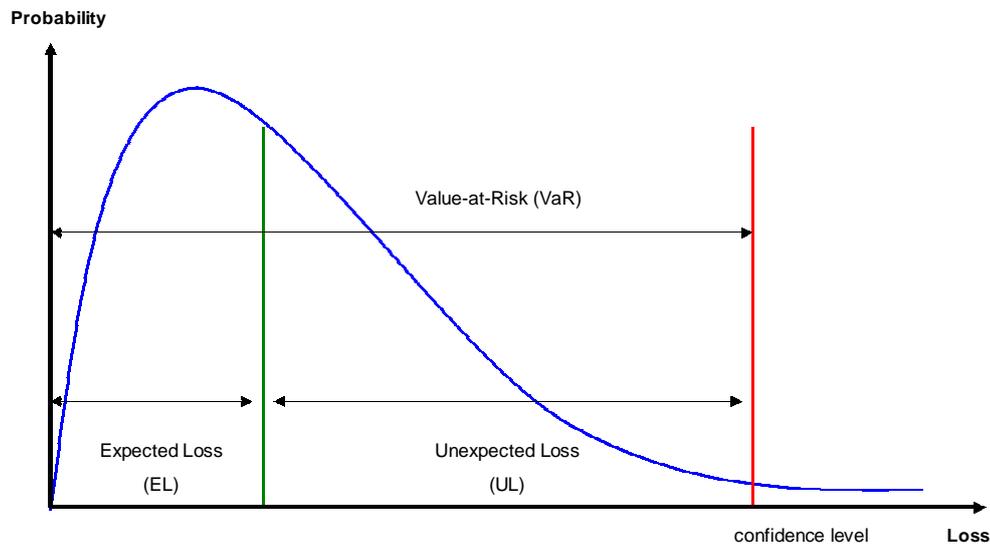


Figure 2.1: Loss regions.

Historical Simulation, Monte Carlo methods and Variance-Covariance methods.

In the first case, a long time series of data is collected on the returns of the assets. A percentage change of each asset is computed to generate scenarios. The results are reordered by the magnitude of the change in the value of the portfolio and a desired percentile is chosen to quantify VaR. For example, if there are 1,000 scenarios and a VaR of 95 % is required, then the actual loss exceeding 5% (50 cases) is the 51st worst loss. For Monte Carlo methods, an underlying model (e.g. Geometric Brownian Motion) and hence distribution is assumed to generate scenarios. Parameters for the model are calibrated on market data and random draws from the Standard Normal are used to simulate paths for the process. The value of the portfolio for each scenario is computed and a distribution is again formed for which the desired percentile is taken. For the least common of the three methods, the Variance-Covariance method (used in Risk Metrics and originally developed by JP Morgan) uses volatility and correlation data and constructs a variance -covariance matrix to calculate VaR. The assumptions underlying the model is that the likely distribution of returns on

Parameter	Definition	Explanation
LGD	Loss Given Default	Loss realized during an economic downturn from an obligor defaulting
N	Standard normal distribution function	Applied to threshold and conservative value of systematic factor
R	Asset Correlation	Quantifies how asset value of one borrower depends on asset of another borrower determined by asset class
PD	Probability of Default	Expected default rates usually over 1 year time horizon of an obligor
G	Inverse standard normal distribution function	Applied to PD to derive default threshold
M	Maturity adjustment	Captures effect that long-term credits are riskier than short-term credits
b	Smooth (regression) maturity adjustment (smoothed over PD)	$b(PD) = (0.11852 - 0.05478 * \log(PD))^2$

Table 2.1: ASRF inputs.

financial assets in the future can be measured based on the observed volatility of the returns on those assets in the past and the distribution of those returns in the future will be Normal.

For modeled methods for credit risk, the BCBS explanatory note on the risk weight function (2005) [18] explains one of the most popularized model for credit risk. The formula used to determine the risk weight (K) is shown in (2.1). The value K is a modeled risk-weight function which once applied to a notional or exposure dollar value would produce a capital requirement. The motivation and derivation is outlined in Gordy (2003) [72] and the model is referred to as the Asymptotic Single Risk Factor (ASRF) model. The risk weight is computed as

$$K = \left[LGD * \mathbf{N}[(1 - R)^{-0.5} * \mathbf{G}(PD) + \left(\frac{R}{1 - R}\right)^{0.5} * \mathbf{G}(0.999)] - PD * LGD \right] * \frac{(1 + (M - 2.5) * b(PD))}{(1 - 1.5 * b(PD))}, \quad (2.1)$$

where the inputs are defined in Table 2.1.

To provide some more context, banks are allowed through this formula to determine the borrower's probabilities of default (PD), loss given default (LGD) and exposure at default (EAD) on an exposure-by-exposure basis. These risk parameters are converted into risk weights and regulatory capital requirements by means of the risk weight formula. An artifact that is used by banks is that in a particular year, the average level of credit losses can be reasonably forecasted. This expected loss (EL) is managed by various means such as pricing of credit exposures and through accounting provisioning. Losses above expected levels are referred to as unexpected losses (UL) and one use of bank capital is to provide a buffer for this region of losses. Hence capital is set to ensure that UL exceeds the capital buffer with low, fixed probability.

The derivation of capital charges for UL is based on the ASRF model and has a few inherent properties. The model is portfolio invariant in the sense that the capital required for any given loan depends on the risk of the loan and not on the portfolio that it is added too. Portfolio invariance schemes are also referred to as rating-based schemes which imply that obligor-specific characteristics like PD, LGD, and EAD are enough to determine capital for credit loans. Also associated with the ASRF scheme is the fact that when a portfolio consists of a large number of small exposures, idiosyncratic risks tend to cancel out with one another and only systematic risks that affect many exposures have a dominant effect on portfolio losses. Most simply, a single risk factor is believed to drive the systemic risk component which is reflective of the state of the global economy, industry or even regional risks.

The ASRF model shows that it is possible to calculate the sum of EL and UL associated with each credit exposure. The conditional EL for an exposure is expressed as a product PD and LGD which describes the loss rate on the exposure in the event of default. The method by which this is achieved is that average PDs estimated by banks are turned into conditional PDs by using a supervisory mapping function. This essentially pushes out the PD to a high quantile given a conservative value of the sys-

tematic risk factor. Note that LGD is not transformed in the same manner and scales linearly in the formula. Moreover, EAD does not appear until the end calculation which applies a notional value. The mapping function used to produce the conditional PD is derived from an adaption of the work by Merton (1974) [89] and Vasicek (2002) [111]. Therein, borrower's default if they cannot meet their obligations at a fixed assessment horizon (e.g. one year) because the value of their assets are lower than the due amount. The value of assets of a borrower are represented as a variable whose value can change over time. The change in value of the borrower's assets is modeled as a normally distributed random variable. Since the default threshold and the borrower's PD are connected through the normal distribution function, the default threshold is inferred from the PD by applying the inverse normal distribution function to the average PD. As well, the conservative value due to the systematic risk factor can be derived by applying the inverse of the normal distribution function to the high quantile value. A correlation-weighted sum of the default threshold and the conservative value of the systematic factor yields a conditional default threshold.

In a following step, the conditional default threshold is used as an input into the original Merton model and is used to derive a conditional PD via the normal distribution function. It is noteworthy to point out that the ASRF model calculates a requirement from zero to the VaR value of 99.9 % and hence through regular credit provisioning, EL is subtracted from the ASRF so as to not double count the EL value.

The asset correlation values (R) represents the degree of an obligor's exposure to the systematic risk factor. They are asset class dependent and follow supervisory mandated values. A final tack on is a maturity adjustment multiplier that is again calibrated and mandated by supervisors. The reason for its use is to address that capital should reflect the fact that long-term credits are riskier than short-term credits and hence should increase with maturity. The form of the maturity adjustment has been derived using a mark-to-market credit risk model. From a matrix of decreasing PD

grades in rows and increasing time horizons along the columns, a statistical regression model was fit and the output has been given to the industry to be used in the formula.

The takeaway is that setting modeled credit risk capital requirements is different than modeled market risk capital requirements and, as we will see, modeled operational risk capital requirements which is the focus of our work is different than both of them. For modeled operational risk, an actuarial approach has been borrowed called the Loss Distribution Approach (LDA). It is again a VaR method but taken at an even higher quantile. As we will see, there are added layers to the baseline LDA model which results in a nice blend of quantitative and qualitative methodologies and is at the intersection of art and science in terms of modeling. Operational risk research has lagged in importance compared to its counterparts but is now starting to gain momentum. Before we go deeper into operational risk methodologies, we motivate the importance of the subject in the next section.

2.2 Importance of Operational Risk

We argue that both from a bank's perspective as well as from a research perspective, operational risk as a whole has unjustly been placed on the back-burner when it should not have. For the Canadian market, we may motivate the importance of operational risk by looking at the big five Canadian banks' [Bank of Montreal (BMO), Scotiabank (BNS), Canadian Imperial Bank of Commerce (CIBC), Royal Bank of Canada (RBC) and Toronto-Dominion Bank(TD)] supplementary financial statements. Each quarter, banks make public on their websites capital held against market, credit and operational risk as well as disclosing key operating capital ratios. We have briefly covered mathematical approaches that are used to attach a dollar value to the corresponding risk. This value was referred to as a regulatory capital requirement but with a short extension, we may look at the related concept of risk weighted assets (RWA). For market and operational risk, the simple relation is to multiply the capital

requirement by 12.5

$$\text{market or operational RWA} = 12.5 * \text{capital requirement}. \quad (2.2)$$

The 12.5 arises from a legacy value of the reciprocal of the minimum capital ratio of 8% as seen in BCBS (1988) [15]. The capital requirement value could be derived either from a standardized method or a modeled method and in both instances, would be multiplied by 12.5 to form a RWA. RWA represents a bank's assets or off-balance sheet exposures weighted by riskiness. From either standardized or modeled methods, less risky assets would result in smaller capital requirements and hence would equate to smaller RWAs. For modeled credit risk, we saw how to determine the capital requirement risk weight K . As an intermediary step, we must multiply K by an exposure at default (EAD) amount first and then multiply by 12.5. The EAD is the notional amount in dollars that represents exposure for a given obligor. Hence for credit risk RWA, we have

$$\text{credit RWA} = 12.5 * \text{EAD} * \text{risk weight}. \quad (2.3)$$

Another slight technicality is that unlike market or operational RWA, there is a split between the treatment of standardized and modeled credit risk RWA. There is a scalar of 1.06 applied to modeled credit risk

$$\text{credit RWA} = \text{credit RWA}_{\text{Standardized}} + 1.06 * \text{credit RWA}_{\text{Modeled}}. \quad (2.4)$$

In brief, the factor was designed to offset the expected decrease in the capital requirement resulting from the change in the capital formula from our expected loss plus unexpected loss orientation, to an unexpected loss-only orientation. The size of the scaling factor was derived based on the results of the third Quantitative Impact Study conducted by the BCBS (2003) [17].

At this stage it is enough to grasp that capital requirements are directly proportional to RWA. Thus, comparing how RWAs are split among market, credit and operational risks would convey a sense of size/importance. However, rather than purely focus on RWA, the discussion can be framed in terms of regulatory capital ratios for which RWA is a key input. Institutions are expected to meet minimum regulatory risk based capital requirements for exposure to market risk, credit risk and operational risk. The generic formula for a capital ratio is given as

$$\text{Risk Based Capital Ratios} = \frac{\text{Capital}}{\text{Total RWA}} \quad (2.5)$$

where

$$\begin{aligned} \text{Total RWA} = & \text{Credit RWA}_{\text{Standardized}} \\ & + 1.06 * \text{Credit RWA}_{\text{Modeled}} \\ & + 12.5 * \text{Operational Risk Capital} \\ & + 12.5 * \text{Market Risk Capital}. \end{aligned} \quad (2.6)$$

The three most important types of capital take the form of either common equity tier 1 (CET1), total tier 1, or total capital. All three ratios are again publicly disclosed in the supplementary financials. The most scrutinized ratio is the CET1 ratio and is defined as

$$\text{CET1 ratio} = \frac{\text{CET1 capital}}{\text{Total RWA}}. \quad (2.7)$$

The qualifying instruments that make up CET1 capital (subject to regulatory adjustments) are reproduced from the BCBS guidance on capital disclosure (2012) [25] in Table 2.2. Without getting into the accounting of what instruments qualify as viable CET1 instruments, CET1 capital represents the highest quality of assets which may be drawn upon if a bank needs to fulfil its obligations. CET1 capital is a dollar value and total RWA is a simple sum of total RWA for market, credit and operational risk. From a regulatory perspective, the ratio is one of the most important metrics used to monitor a bank's solvency. In Table 2.3, we see that operational risk is the second

Common Equity Tier 1 Capital
• Directly issued qualifying common share (and equivalent for non-joint stock companies) capital plus related stock surplus
• Retained earnings
• Accumulated other comprehensive income (and other reserves)
• Directly issued capital subject to phase out from CET1 (only applicable to non-joint stock companies)
• Common share capital issued by subsidiaries and held by third parties (amount allowed in group CET1)
• Regulatory adjustments applied in the calculation of Common Equity Tier 1

Table 2.2: Common Equity Tier 1 qualifying instruments.

most capital intensive capital requirement and is on average across the big five Canadian banks almost two times greater than market risk RWA.

\$ Million	BMO	BNS	CIBC	RBC	TD
CET1 capital (A)	22,340	28,499	13,347	32,998	27,803
Credit Risk RWA	198,803	253,196	118,548	253,799	263,971
Market Risk RWA	14,494	16,714	4,170	44,055	13,177
Operational Risk RWA	26,779	32,160	*17,787	43,898	35,824
Total RWA (B)	240,076	302,070	140,505	341,752	312,972
CET1 ratio (A/B)	9.3	9.4	9.5	9.7	8.9

Table 2.3: Big five Canadian bank risk weighted assets and common equity tier 1 ratios for Q1 2014. *Reports modeled operational risk.

Now, it is easy to form an argument in support of the importance of operational risk management and research. A lot of attention has been focused on trading activities, sophisticated models used to value financial instruments, and associated models used for risk management. However, resources should also be adequately allocated for operational risk activities. A more risk-sensitive method via AMA allows management to get a handle on operational risk vulnerabilities and potentially implement mitigating strategies to ward off future losses. The AMA model also allows for a more granular drill-down view to track capital allocation commensurate with the losses observed in different areas of the bank.

2.3 Basic Indicator Approach and The Standardized Approach

The BCBS defines two standardized methods to set regulatory capital requirements for operational risk. The two methods are referred to as Basic Indicator Approach (BIA) and The Standardized Approach (TSA). For BIA, the amount of capital to set aside is 15% of the 3-year average of positive gross income reported by a bank. If there are year(s) of negative gross income, then the average is taken only on the remaining positive year(s). There are technicalities that are needed to produce a gross income value, however for our purpose, it is enough to treat this as a starting point. The thinking behind the BCBS for using gross income as a metric to proxy operational risk is that it conveys information on the size and activities a bank may partake in. Hence as gross income increases, operational risk capital should increase proportionately. A by-product of this is that during a boom of an economic cycle, banks would hold increasing amounts of operational risk capital. If a crisis was to occur and gross income for a bank would fall, then a bank would be required to hold a proportionately reduced amount of capital. Hence the bank would be able to draw-down on the capital reserve and partially weather the crisis. A simple example demonstrates the use of BIA.

Example 2.3.1. *A bank has positive gross income of \$1,000,000, \$3,000,000 and \$900,000 over three consecutive years. Under the BIA, the bank would need to hold 15% of $\frac{(\$1,000,000+\$3,000,000+\$900,000)}{3}$ or $0.15 * \$1,633,333$ or \$245,000.*

For TSA, the BCBS specifies to segment a bank into 8 business lines (according to the Basel II suggested classification). Gross income is calculated/allocated for each of these segments (each segment should be assigned some gross income; whereas before we looked at a top-of-the-house number). However, now each gross income increment is weighed by some factor β . This is illustrated in Table 2.4. To outline the application

of TSA, we again proceed via an example.

Business Line	β value
Corporate Finance	18 %
Trading and Sales	18 %
Retail Banking	12 %
Commercial Banking	15 %
Payment and Settlement	18 %
Agency Services	15 %
Asset Management	12 %
Retail Brokerage	12 %

Table 2.4: TSA weights across business lines.

Example 2.3.2. *A bank has gross income per Basel business line as indicated in Table 2.5. (Note, this is an independent example from that in Example 2.3.1 and the result should not indicate that in all instances, a bank would receive a reduction/enlargement of capital requirements from using the TSA approach opposed to the BIA approach). In this example, taking the 3 year total weighted gross income values and averaging would yield $\frac{(\$15000,000+\$200,000+\$100,000)}{3}$ or \$150,000.*

2.4 Advanced Measurement Approach

The modeled method used to calculate operational risk capital is referred to as AMA. While in the early phases most banks initially started reporting operational risk capital using the BIA and TSA method (TSA being preferred in some instances because it allows for more risk differentiation), AMA is seen as the next step in robust risk management. At its core, there are four elements to AMA modeling: Internal Loss Data (ILD), External Loss Data (ELD), Business Environment and Internal Control Factors (BEICF) and scenarios analysis. The AMA model, as we will see, is a blend of both rigorous quantitative/statistical analysis as well as qualitative analysis. Detailed descriptions of AMA modeling are found in Aue and Kalkbrener (2006) [10], Cruz *et*

Year 1	Gross Income	β	Weighted Gross Income
Corporate Finance	\$200,000	18%	\$36,000
Trading and Sales	\$100,000	18%	\$18,000
Retail Banking	\$350,000	12%	\$42,000
⋮	⋮	⋮	⋮
Year 1 Total	\$ 5,000,000		\$150,000
Year 2			
Corporate Finance	\$150,000	18%	\$27,000
Trading and Sales	\$400,000	18%	\$72,000
Retail Banking	\$250,000	12%	\$30,000
⋮	⋮	⋮	⋮
Year 2 Total	\$ 6,000,000		\$200,000
Year 3			
⋮	⋮	⋮	⋮
Year 3 Total	\$ 3,000,000		\$100,000

Table 2.5: Gross income by business line weighted by β .

al (2015) [47], Dutta and Perry (2006) [58] and Panjer (2006) [101]. Actuaries are most familiar with the LDA method to model aggregate claims. Empirically, losses tend to occur at random times (think frequency) and random magnitudes (think severities). By fitting separate frequency distributions and separate severity distributions, both random events may be combined to produce an estimated loss amount. A BEICF assessment is a post-modeling capital adjustment that adjusts a baseline LDA capital requirement. BEICF is an acronym whereby each letter represents an attribute that could be used to develop a view on an operational risk characteristic of a business unit. The attributes make up a scorecard and the weights assigned to each item in the scorecard are completely up to a bank. Finally, and perhaps the most controversial element, scenario analysis requires business line experts to quantify potential events that may occur at a financial institution. There are several difficulties that arise from this element alone. For instance, what questions should be asked of business line experts in order to extract relevant information to quantify a scenario? What necessary and sufficient information is required to form a scenario?

Once formed, how do you use the scenario in the AMA model? More specifically, what are direct or indirect methods that could be used to take the scenario information obtained from workshop and integrate it to a baseline LDA model? It is this topic of scenario analysis that will occupy much of our attention. Whereas it may appear at first to be an area that should not pose much debate, there is much ambiguity and room for rich application of mathematical and statistical tools. As there is little consensus within the finance community on scenario analysis, this is a wide-open area to explore techniques of scenario quantification and integration. However, before we explore methods for scenario analysis, we must fully introduce all elements of the AMA model and ancillary components that effect the AMA model. It is only then will the scenario analysis piece, which will be our focus, be holistically integrated.

2.4.1 Element 1: Internal Loss Data

Internal loss data is the first stepping stone in building the AMA model. (It is helpful to keep reminding ourselves of the applied nature of this problem and the real-world application it serves. Loss data is booked by risk managers into a central database within a bank and the onus is on them to be timely and accurate with recording the losses.) The BCBS provides guidance within Basel II (2006) [19] in Annex 8 how to segment a bank and record operational risk losses by 8 business lines (same as in the TSA method) and by 7 event types which is a level 1 classification. Each segment is referred to as a unit of measure (UoM) or alternatively a cell. The cell is in reference to an entry in the 56 cell matrix. We note that we may further refine the 8 business lines into 19 categories and 7 events types into 20 categories which is then referred to as a level 2 classification. For our analysis, let's consider the 8 by 7 matrix with business lines as rows and event types as columns which is depicted in Table 2.6. A convention that is commonly used in the literature is to refer to a particular business line as BL and then its position in the matrix and similarly with event type as ET. Hence, keeping the same convention we have the mapping in Table 2.7.

Business Line/Event Type	Internal Fraud	External Fraud	Employment Practices and Workplace Safety	Clients, Products & Business Practices	Damage to Physical Assets	Business Disruption and System Failures	Execution, Delivery, Process Mgmt
Corporate Finance							
Trading and Sales							
Retail Banking							
Commercial Banking							
Payment and Settlement							
Agency Services							
Retail Brokerage							
Asset Management							

Table 2.6: 56 cell matrix for recording operational risk losses.

Now over the course of time, it is possible to think of each of cell in Table 2.6 being populated with a loss event. An example of simulated data is located in Table 2.8 and Table 2.9. In practice, an operational risk manager may be tasked to oversee the accurate recording of loss events into a central database. Most banks in Canada use a \$10,000 threshold as a cut off value. For example, the retail banking business could experience an external credit card fraud (BL3ET2) of \$40,000 which would be recorded in the third row and second column in Table 2.6. However, a loss of \$5,000 would not be recorded in a cell because it falls below the threshold value and reasoned not to inform the tail of the distribution too much. From a modeling perspective, it is not always easy to obtain a plentiful data set from internal losses alone. Regulation does allow for the collapsing of cells in order to increase the number of data points across business lines or event types provided that neighbouring cells have similar characteristics. The combined cells would then be treated as one and yield a more coarse structure than the Basel-recommended 56 cell classification. However, to help prevent numerous situations for collapsing, external loss datasets have arose to, among other possible uses, help with data scarcity.

Business Line	
BL1	Corporate Finance
BL2	Trading and Sales
BL3	Retail Banking
BL4	Commercial Banking
BL5	Payment and Settlement
BL6	Agency Services
BL7	Asset Management
BL8	Retail Brokerage
Event Type	
ET1	Internal Fraud
ET2	External Fraud
ET3	Employment Practices and Workplace Safety
ET4	Clients, Products and Business Practices
ET5	Damage to Physical Assets
ET6	Business Disruption and System Failures
ET7	Execution, Delivery and Process Management

Table 2.7: Business Line/Event Type mapping.

Date of Discovery	Business Line	Event Type	Gross Loss (\$)
09-Jan-13	BL3	ET2	84,880
10-Jan-13	BL3	ET2	90,355
14-Jan-13	BL3	ET2	71,620
16-Jan-13	BL3	ET2	85,056
23-Jan-13	BL3	ET2	75,129
25-Jan-13	BL3	ET2	106,295
29-Jan-13	BL3	ET2	91,039
29-Jan-13	BL3	ET2	103,552
29-Jan-13	BL3	ET2	200,431
06-Feb-13	BL3	ET2	64,186
06-Feb-13	BL3	ET2	83,426
13-Feb-13	BL3	ET2	1,417,927
19-Feb-13	BL3	ET2	162,538

Table 2.8: Retail Banking/External Fraud data.

2.4.2 Element 2: External Loss Data

External loss data is another element in the AMA model that addresses the issue of limited internal loss data. In the form of public and/or pooled industry data, ELD

Date of Discovery	Business Line	Event Type	Gross Loss (\$)
17-Jan-13	BL3	ET7	70,000
18-Jan-13	BL3	ET7	42,281
25-Jan-13	BL3	ET7	34,463
29-Jan-13	BL3	ET7	104,903
30-Jan-13	BL3	ET7	49,250
01-Feb-13	BL3	ET7	197,320
05-Feb-13	BL3	ET7	177,623
07-Feb-13	BL3	ET7	50,000
08-Feb-13	BL3	ET7	31,782
14-Feb-13	BL3	ET7	64,000
18-Feb-13	BL3	ET7	39,666
25-Feb-13	BL3	ET7	51,769

Table 2.9: Retail Banking/Execution, Delivery and Process Management data.

allows a bank to incorporate information that could supplement their internal loss experience. Intermixed with ILD, ELD could first be scaled appropriately depending on the scale of business or bank from which the data comes from. It could then be directly used in the capital calculation. Indirectly, ELD may also be used as a guide during brainstorming for scenario analysis which we will see later on.

The two main sources for ELD are vendor databases and consortia databases as stated by the OCC (2012) [95]. Vendor databases such as IBM-Algo FIRST and SAS OpRisk Global are compiled from public data and have a story-line approach to describing losses. The losses are extracted from newspapers, court records, journals, etc. Consortia databases such as ORX (Operational Risk Exchange) and the American Bankers Association (ABA) operate on a give-and-take basis allowing banks to contribute operational loss data in order to receive anonymized loss data from peers groups. Attributes for both ELD types are shown in Table 2.10 and Table 2.11. An ongoing debate among institutions before using ELD is the difficulty in agreeing on a metric to scale the data, especially when data from consortiums are made anonymous. Current proposals are to scale using asset size or revenue of the institution/business line which is an indicator made available.

	SAS OpRisk Global	IBM-Algo FIRST
Total number of observations	22,000	12,000
Industries covered	21	20
Financial services covered	6,361	4,519
Loss threshold	USD \$100,000	USD \$1,000,000
First year of losses	1971	1972

Table 2.10: Vendor databases.

2.4.3 Loss Distribution Approach

Having established the data for which an AMA model is built, it is now appropriate to showcase the LDA model. Using the data, separate probability distributions for loss frequency and severity are fit at the business line/event type level and then combined. That is for each cell ($k = 1, \dots, m$) where m is suggested to be 56, the aggregate loss distribution is given by

$$S_k = \sum_{i=1}^{N_k} X_{k,i} \quad (2.8)$$

where N_k represents realizations from a frequency distribution and $X_{k,1}, \dots, X_{k,N_k}$ represent random draws from the severity distribution X_k which are modeled separately over a one year time horizon. Hence the aggregate loss for a single cell is obtained by the convolution of severity where the weights are the frequency mass

	ORX	ABA
Total number of observations	160,000	27,535
Severity of losses	€55 billion	\$2.7 billion
Member banks	54	10
Location of banks	18 countries	U.S. only
First year of losses	2002	2003
Loss threshold	€20,000	USD \$10,000

Table 2.11: Consortia databases.

probabilities

$$\begin{aligned}
F_S(x) &= Pr(S \leq x) \\
&= \sum_{n=0}^{\infty} p_n Pr(S \leq x | N = n) \\
&= \sum_{n=0}^{\infty} p_n F_X^{*n}(x)
\end{aligned} \tag{2.9}$$

where $p_n = Pr(N = n)$ and F_X^{*n} is the n -fold convolution of the cumulative distribution function (CDF) of X . To define the convolution in greater detail, we adapt the explanation from Klugman *et al* (2012) [81]. We focus on one UoM with the understanding that the same procedure applies to all other cells. The random sum for one cell is given as

$$S = X_1 + X_2 + \dots + X_N \tag{2.10}$$

where index N is represented by a counting distribution. The distribution of S is called a compound distribution or the aggregate distribution for a cell. From (2.9), we define $F_X^{*n}(x)$ iteratively as

$$F_X^{*0}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \tag{2.11}$$

and

$$F_X^{*k}(x) = \int_{-\infty}^{\infty} F_X^{*(k-1)}(x-y) dF_X(y) \quad \text{for } k = 1, 2, \dots \tag{2.12}$$

A simplification which applies to AMA modeling occurs when X is a continuous random variable with probability zero on nonpositive values, then (2.12) reduces to

$$F_X^{*k}(x) = \int_0^x F_X^{*(k-1)}(x-y) f_X(y) dy \quad \text{for } k = 2, 3, \dots \tag{2.13}$$

For $k = 1$, (2.13) reduces to $F_X^{*1}(x) = F_X(x)$. Upon differentiating the probability density function (PDF) is recovered

$$f_X^{*k}(x) = \int_0^x f_X^{*(k-1)}(x-y)f_X(y)dy \quad \text{for } k = 2, 3, \dots \quad (2.14)$$

It is (2.14) that will reappear in a later chapter and serve as a foundational technique used to integrate scenarios into a baseline LDA model. This equation conveys that in order to compute the n-fold convolution, the integral is computed iteratively using previously calculated density functions. Then a VaR measure at a one-tailed 99.9 percentile is taken from the corresponding compound loss distribution for each cell denoted by S_k . We refer the reader to Panjer (2006) [101] or Cruz *et al* (2015) [47] for details on forming the compound loss distributions. Before moving on, since VaR will be our risk measure, we may define this in more detail.

Definition 2.4.1. (*Risk measure*) *A risk measure is a mapping of a random variable representing risk to a real number. Hence, a general risk measure related to the risk X is denoted as $\rho[X]$.*

The choice of risk measures are numerous and in practice, the decision to use one over the other may prove difficult. Before rigorously defining our risk measure, which is seen as one of the simplest and most common risk measures, we turn to a list of properties that a good (coherent) risk measure should satisfy. These properties were introduced in Artzner *et al* (1999) [9] and reproduced here.

Definition 2.4.2. (*Coherent risk measure*) *A coherent risk measure, $\rho[X]$, is defined to have the following properties for any two random variables X and Y*

- *Translation invariance:* for any constant c , $\rho[X + c] = \rho[X] + c$;
- *Monotonicity:* if $X \leq Y$ for all possible outcomes, then $\rho[X] \leq \rho[Y]$;
- *Subadditivity:* $\rho[X + Y] \leq \rho[X] + \rho[Y]$;

- *Positive homogeneity: for positive constant c , $\rho[cX] = c\rho[X]$.*

Since a risk measure outputs a capital requirement, we may interpret these properties in terms of capital. Translation invariance indicates that adding a fixed amount to a collection of risks will cause the capital requirement to be adjusted by the same fixed amount. Here loss is defined as a positive value and the fixed amount does not impart further/new capital requirements. Monotonicity ensures risks that lead to smaller losses in all instances require less capital. Subadditivity is perhaps the most popularized because VaR fails to meet this condition in some instances. The property, however, allows for a diversification benefit. Loosely interpreted, this property indicates that the merger of two collection of risk will not create extra risk. Hence, the sum of the individualized capital requirement serve as the upper bound. Since this concept is of such great importance, risk managers and regulators are constantly concerned with the diversification coefficient.

Definition 2.4.3. (*Diversification coefficient*) For a collection of risks X_1, X_2, \dots, X_n , the diversification coefficient is defined as

$$D = 1 - \frac{\rho[X_1 + X_2 + \dots + X_n]}{\rho[X_1] + \rho[X_2] + \dots + \rho[X_n]}. \quad (2.15)$$

The coefficient is positive if there is a diversification benefit and negative if diversification fails. The negative case then leads to a discussion of superadditivity which will not be covered here. Finally, positive homogeneity indicates that increasing a risk by a factor c should increase capital by the same c . Having described properties of a risk measure, we define the two most popular choices: VaR and expected shortfall (ES).

Definition 2.4.4. (*Value-at-Risk*) The VaR of a random variable $X \sim F(x)$ at the

α^{th} probability level is defined as the α^{th} quantile of the distribution of X

$$\begin{aligned} \text{VaR}_\alpha[X] &= F^{-1}(\alpha) = \inf\{x : \mathbb{P}[X > x] \leq 1 - \alpha\} = \inf\{x : F(x) \geq \alpha\} \\ &= \sup\{x : F(x) < \alpha\}. \end{aligned} \tag{2.16}$$

Hence VaR is the minimum threshold exceeded by X with probability at most $1 - \alpha$.

VaR, as a quantile of a distribution, has become the de facto choice in Basel II and used as the risk measure of choice throughout market and operational risk. It is easy to understand and is capable of being backtested. That is, a VaR model produces an estimate for potential loss at some future time horizon. Moving forward in time, the actuals (profit and loss, realized losses, etc.) can be compared to the estimate and the number of times the quantile is surpassed over a time horizon signifies a backtesting breach. As pointed out before, VaR may fail the subadditivity property and hence is not a coherence risk measure. However, expected shortfall is a coherence risk measure but does come with its own drawbacks.

Definition 2.4.5. (*Expected shortfall*) The expected shortfall, or tail VaR, of a random variable $X \sim F(x)$ at the α^{th} probability level is

$$ES_\alpha[X] = \frac{1}{1 - \alpha} \int_\alpha^1 \text{VaR}_p[X] dp. \tag{2.17}$$

Hence ES is the arithmetic average in the tail having exceeding the value α .

Expected shortfall is becoming increasingly important for market risk quantification as the BCBS seeks to adopt ES as a complementary risk measure. The preliminary guidance is referred to as the fundamental review of the trading book and has not yet been fully adopted. For operational risk, ES is not commonly used. As pointed out by Embrechts *et al* (2014) [62], backtesting VaR is straightforward (hit-and-miss) while for ES one may assume an underlying model such as Extreme Value Theory. Addi-

tional complications of backtesting ES opposed to VaR is discussed therein. Moreover, backtesting an average metric opposed to a single point is seen as a difficult exercise as a series of breaches over a quantile may be warranted provided this behaviour does not repeat indefinitely and allows for other losses below a threshold so that the average is within an acceptable range.

Hence for reasons of simplicity and tradition, VaR will be our risk measure of choice. Hence, applying (2.16) to each UoM loss in (2.8), the final regulatory capital requirement is achieved by aggregating up each VaR result to an enterprise level for the bank. While it is important to have a good understanding of the parts used to form the LDA model, we must revisit each component and provide a bit more detail. It is not within the scope of our work to provide a thorough discussion on the multitude of options of building blocks but rather refer the reader to Cruz *et al* (2015) [47] for a comprehensive review. However, we do focus on the most common and popular modeling choices in the next section.

2.4.4 Calibration and Selection

The most common probability distribution choices for modeling the frequency component of operational losses are the Poisson and Negative Binomial distributions. The convention is to use discrete probability distributions in that losses are not typically seen to occur on a continuous scale. The Poisson distribution has one parameter whereas the Negative Binomial distribution has two parameters which has more flexibility in shape. We define both distributions in [Appendix A: Probability Distributions](#). The Negative Binomial distribution may be viewed as a gamma mixture of a Poisson distribution. That is, a Poisson distribution whose parameter λ is a gamma-distributed random variable. It allows for the case of over-dispersion where the mean is greater than the variance. In Böcker and Klüppelberg (2005) [35], we learn there exists a closed-form approximation for operational risk VaR and in the formula, over-dispersion is of minor importance. Hence, if a modeler would like

to benchmark an AMA model against an approximation method, the choice of frequency distribution should not matter that much. In that sense, we default to using the Poisson throughout our analysis. However, what is more important is the choice of severity distribution.

There are several choices for a severity distribution. Here we allow for continuous probability distributions as they are better suited to modeling the size of losses on a continuous scale. A suggested class of sub-exponential distributions are recommended in the BCBS Supervisory Guidelines (2011) [22] and listed as follows:

- Lognormal,
- Log-Gamma,
- Weibull (shape parameter less than 1),
- Generalized Pareto,
- Burr.

The definitions for each are produced in [Appendix A: Probability Distributions](#). These distributions have the property that the tails decay slower than any exponential tail. The concept of heaviness is also outlined in Böcker and Klüppelberg (2005) [35]. To define the heaviness characteristic, we have

$$\lim_{x \rightarrow \infty} \frac{P(X_1 + \dots + X_n > x)}{P(\max(X_1, \dots, X_n) > x)} = 1 \quad \text{for } n \geq 2. \quad (2.18)$$

The above states that the tail of the sum of n sub-exponential random variables has the same order of magnitude as the tail of the maximum variable among them. As a consequence, one term could explain the reason why the sum of n independent and identically distributed (i.i.d.) severities are large and hence a single large loss would drive operational risk capital. Banks may also choose to model the body and tail of the distribution separately. Hence, there maybe combinations of the above distributions which are spliced together at an optimum splice point.

Common methods to determine the parameters and thus calibrate a severity curve range from using method of moments, percentile matching, or maximum likelihood estimation (MLE). As pointed out in Panjer (2006) [101], the method of moments and percentile matching tend to yield poor performing estimators even though the methodology is easily implemented. Hence for our discussion, we will default to using MLE. Once a candidate list of severity curves are calibrated, goodness of fit tests are used to rank best fits. Common goodness of fits tests used are two-sample Kolmogorov-Smirnov (KS), right-tail Anderson-Darling (AD), and other useful statistics such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). It is common within the banking industry to use MLE for parameter estimation and a host of goodness of fit tests for the selection of champion frequency/severity distributions. For a full discussion on parameter estimation and model selection, we refer the reader to Cruz *et al* (2015) [47], McNeil *et al* (2005) [87] and Panjer (2006) [101]. For context, we reproduce the MLE methodology and a few goodness of fit tests from the references.

Maximum Likelihood Estimation

Suppose a random vector $X = (X_1, \dots, X_n)$ has joint probability density in some parametric family $f_X(x; \theta)$ indexed by a parametric vector $\theta = (\theta_1, \dots, \theta_p)$ in a parameter space Θ . If we assume the independence of the random variables X_1, X_2, \dots, X_n we can interpret the joint probability density function as

$$\begin{aligned}
 & P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\
 &= P(X_1 = x_1)P(X_2 = x_2) \dots P(X_n = x_n) \\
 &= f(x_1; \theta)f(x_2; \theta) \dots f(x_n; \theta).
 \end{aligned} \tag{2.19}$$

We consider our data to be a realization of X for some unknown value θ . The *likelihood function* for the parameter vector θ given the data is $L(X, \theta) = f_X(X; \theta)$ and the maximum likelihood estimator $\hat{\theta}$ is the value of θ maximizing $L(\theta; X)$. For ease

and because it occurs often, we select to maximize the logarithm of the likelihood function $l(\theta; X) = \ln L(\theta; X)$. To be explicit, we write this estimator as $\hat{\theta}_n$ when we want to emphasize its dependence on the sample size n . For large n we expect that the estimate $\hat{\theta}_n$ should be *close* to the true value θ .

For our application to AMA, we consider the classical situation where X is assumed to be a vector of i.i.d. components with univariate density f so that

$$\ln L(\theta; X) = \ln \prod_{i=1}^n f(X_i; \theta) = \sum_{i=1}^n \ln L(\theta; X_i). \quad (2.20)$$

Two-sample Kolmogorov-Smirnov

The two-sample KS test assess the degree of which two probability distributions differ. In this case, the test compares empirical loss data and simulated loss data based on a calibrated fit. Assume two samples X and Y with sample size m and n drawn from continuous distributions. The empirical cumulative distribution functions of both samples may be given as

$$F_X(x) = \frac{\#i : X_i \leq x}{m} \quad \text{and} \quad F_Y(x) = \frac{\#i : Y_i \leq x}{n} \quad (2.21)$$

which counts the number of observed values that are less than or equal to x . The KS statistic is given as

$$D_{m,n} = \sup_x |F_X(x) - F_Y(x)| \quad (2.22)$$

where the smallest distance among a selection of candidate distributions represents the best fit distribution. The null hypothesis is that both sets of data are from the same distribution. The alternative hypothesis is that they are from different distributions. The null hypothesis is rejected at a critical value α where

$$D_{m,n} > c(\alpha) \sqrt{\frac{n+m}{nm}} \quad (2.23)$$

and $c(\alpha)$ maybe looked up from a table.

Anderson-Darling

Another goodness of fit test arises from a set of quadratic empirical distribution function statistics. Let $F_n(x)$ be the empirical CDF based on n observations and $F_0(x)$ be the hypothesized CDF. Then determine the hypothesis for goodness of fit testing where the null claims loss data are from a hypothesized distribution $F_0(x)$

$$H_0 : F(x) = F_0(x), \quad \forall x \quad (2.24)$$

versus the alternative claim that the observed losses are not realizations from $F_0(x)$

$$H_A : F(x) \neq F_0(x), \quad \forall x. \quad (2.25)$$

Then a convenient measure of the discrepancy or distance between two distribution function is detailed in Anderson and Darling (1952) [6] and the test statistic defined as

$$Q_n = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \psi(x) dF(x) \quad (2.26)$$

where $\psi(x)$ is a preassigned weight function. When the weight function is $\psi(x) = 1$, the statistic is called the Cramér-von Mises (CVM) criterion. When there is more emphasis on the tail, the statistic is referred to as Anderson-Darling and given as

$$Q_n = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{[F(x)(1 - F(x))]} dF(x). \quad (2.27)$$

Finally to complete the test, determine the p -value for the test under the null hypothesis given by considering p -value = $P[|Q_n| \geq q_n | H_0]$. For an AMA model, fitting candidate distributions that place emphasis on the tail are important in that operational risk losses that drive capital requirements are usually low frequency/high severity losses that occur far out in the tails of the distribution.

Akaike Information Criterion and Bayesian Information Criterion

The AIC and BIC are measures of relative quality of a statistical model for a given data set. AIC and BIC do not convey any warning if a model fits poorly, however, are useful values to use in conjunction with MLE during the calibration process. AIC is defined as

$$AIC = 2k - 2L \quad (2.28)$$

where L represents the maximized value of the log-likelihood and k represents the number of parameters in the statistical model. BIC is defined as

$$BIC = k \log(n) - 2L \quad (2.29)$$

where n denotes the number of data points. In both instances, goodness of fit is traded off with model complexity. Hence these measures offer a relative estimate of the information lost when a given model is used to represent the process that generates the data. The preferred model is the one that has the minimum AIC and BIC values.

Hence a bank may use a combination of these tests to select a best fitting severity distribution. A customized scorecard may be used to assign weights to each test and a final champion distribution may be selected.

2.4.5 Loss Aggregation Approaches

The computation of the compound distribution (2.8) is not a trivial task. We highlight a few methods that may be used. They are categorically listed as

1. Analytic Approach

- Single Loss Approximation,

2. Numerical Approach

- Panjer Recursion,
- Fast Fourier Transform,

3. Simulation

- Monte Carlo.

For an in-depth review of these methods, we refer the reader to Panjer (2006) [101]. In [Appendix B: Loss Aggregation](#), we provide a short description of each aggregation method. Since our work will utilize the Monte Carlo simulation method, we describe the algorithm here.

Monte Carlo Algorithm

To implement the procedure, we define the following

- N = random variable representing loss frequency with CDF F_N ,
 - X = random variable representing loss severity with CDF F_X ,
 - S = random variable representing annual aggregate loss with CDF F_S ,
 - M = total number of times to run Monte Carlo simulation (e.g. 1 million).
1. For $m = 1, \dots, M$, draw from the frequency distribution a realization N_m with associated CDF F_N .
 2. Draw N_m severity random numbers $(X_1, X_2, \dots, X_{N_m})$ with associated CDF F_X .
 3. Sum all losses represented as $S_m = \sum_{i=1}^{N_m} X_i$.
 4. Repeat up to M times. The loop results in M independent simulated annual aggregate losses S_1, S_2, \dots, S_M with CDF $F_S(s) = \mathbb{P}[S \leq s] = \sum_{m=1}^{\infty} \mathbb{P}[S_m \leq s | N = m] \mathbb{P}[N = m]$ with $S_k = X_1 + X_2 + \dots, X_{N_m}$.

5. Pick up the 99.9 percentile from the simulated annual aggregate loss distribution as the operational risk capital requirement.

Thus far, there has been many assumptions with respect to intra-cell i.i.d. frequency draws, severity draws and aggregate loss draws. Moreover, once each VaR 99.9 estimate has been calculated for each cell, a straight forward addition of each VaR estimate to determine bank-level capital assumes a perfect dependence among cells/UoM. The notion of dependence will be discussed in the next section.

2.4.6 Correlation, Diversification

Thus far we have concerned ourselves with the method to compute a VaR 99.9 capital requirement for a single UoM without any regard for dependencies with other UoMs. The types of dependencies that may exist in LDA modeling are discussed in Aue and Kalkbrener (2006) [10] and Frachot *et al* (2004) [71]. In the LDA framework, dependencies among events and losses may occur in numerous ways. In Aue and Kalkbrener (2006) [10], we learn of intra-(within) and inter-(between) cell correlation that may arise.

intra-cell

- dependence between frequency of losses N_k within a cell,
- dependence between frequency of losses N_k and the severity of losses X_{k,N_k} within a cell,
- dependence between severity of losses $X_{k,1}, X_{k,2}, \dots, X_{k,N_k}$ within a cell,

inter-cell

- dependence between the frequency distributions N_1, \dots, N_m in different cells,
- dependence between the loss distributions S_1, \dots, S_m in different cells.

There are some combinations that are difficult to model due to a lack of data or they violate certain principles. For example, for loss events occurring within a cell that do not occur independently, the frequency distribution can not be Poisson. In addition, from Frachot *et al* [71] we learn that severity correlation is difficult to tackle under the

LDA framework. Severity between cells may be evident (e.g. internal fraud losses are high when external fraud losses are high); however a basic feature of actuarial models assume that losses within a cell are jointly independent. Hence it would be difficult to assume severity independence within each cell and severity correlation between cells. While Aue and Kalkbrener (2006) [10] decided to model inter-cell dependencies between frequencies for Deutsche Bank, most banks choose to model dependencies on the aggregate loss level as stated by the OCC (2012) [95].

Having focused the attention to dependencies between aggregate losses, the most relevant way to incorporate this relationship is through a copula. What is needed first though is to define a marginal distribution. We do this by first defining the joint cumulative probability distribution function of X and Y as

$$F(a, b) = P\{X \leq a, Y \leq b\}, \quad -\infty < a, b < \infty. \quad (2.30)$$

Then the distribution of X can be obtained from the joint distribution of X and Y as

$$\begin{aligned} F_X(a) &= P\{X \leq a\} \\ &= P\{X \leq a, Y < \infty\} \\ &= P\left(\lim_{b \rightarrow \infty} \{X \leq a, Y \leq b\}\right) \\ &= \lim_{b \rightarrow \infty} P\{X \leq a, Y \leq b\} \\ &= \lim_{b \rightarrow \infty} F(a, b) \\ &= F(a, \infty). \end{aligned} \quad (2.31)$$

The CDF obtained is referred to as the marginal distribution which is the CDF without reference to other values of other variables. We have seen this already as the individual aggregate loss distributions characterizing each UoM. This is the required element for the copula. Formally we may then define the copula as taken from Panjer (2006) [101].

Definition 2.4.6. (*Copula*) A d -variate copula C is the joint distribution function of d Uniform $(0,1)$ random variables. If the d random variables are listed as U_1, U_2, \dots, U_d , then the C may be written as

$$C(u_1, \dots, u_d) = P(U_1 \leq u_1, \dots, U_d \leq u_d). \quad (2.32)$$

Considering continuous random variables X_1, X_2, \dots, X_n with distribution functions F_1, F_2, \dots, F_d respectively with joint distribution given by F , then the probability integral transform $F_1(X_1), F_2(X_2), \dots, F_d(X_d)$ are each distributed as Uniform $(0,1)$ random variables. Hence copulas can be seen to be joint distribution functions of Uniform $(0,1)$ random variables. Thus a copula evaluated at $F_1(x_1), F_2(x_2), \dots, F_d(x_d)$ can be written as

$$C(F_1(x_1), \dots, F_d(x_d)) = P(U_1 \leq F_1(x_1), \dots, U_d \leq F_d(x_d)). \quad (2.33)$$

With the inverse (or quantile function) defined as

$$F^{-1}(u) = \inf_x \{F_j(x) \geq u\}, \quad (2.34)$$

the copula evaluated at $F_1(x_1), F_2(x_2), \dots, F_d(x_d)$ can be rewritten as

$$\begin{aligned} C(F_1(x_1), \dots, F_d(x_d)) &= P(F^{-1}(U_1) \leq x_1, \dots, F_d^{-1}(U_d) \leq x_d) \\ &= P(X_1 \leq x_1, \dots, X_d \leq x_d) \\ &= F(x_1, \dots, x_d). \end{aligned} \quad (2.35)$$

Sklar's theorem (1959) states in a formal way that for any joint distribution function F , there is a unique copula C that satisfies (2.35). Conversely, for any copula C and any distribution function $(F_1(x_1), \dots, F_d(x_d))$, the function $C(F_1(x_1), \dots, F_d(x_d))$ is a joint distribution function with marginals $(F_1(x_1), \dots, F_d(x_d))$.

There are two main reason why we choose to integrate copulas into our talk of AMA modeling

- risk can be split into two parts: the individual risk and the dependence structure between them,
- a dependence structure may be defined without reference to the modeling specification of individual risks.

While the discussion of dependence is of great importance, the questions surrounding a best practice is another area of research in itself. For example, choosing the “right” copula to model the dependencies unfortunately has no obvious answer according to Embrechts (2009) [59]. For example, one encounters selection from such choices as

- Gaussian/Normal copula,
- Student t copula (with associated degrees of freedom),
- Gumbel copula,
- Clayton copula,
- Frank copula.

What this equates to is banks who seek to report a diversified operational risk capital requirement may favour such a copula that reduces capital. For instance from Embrechts (2012) [60], we see that joint tail dependence is a copula property, regardless what the marginals are. In addition, under asymptotic independence joint extremes are very rare. This is shown in Figure 2.2. We point out that the top-right quadrant under the Gumbel copula realizes greater joint tail dependence as shown by an increased number of hits. This is contrasted to the Gaussian copula that has less joint realizations in the top-right quadrant.

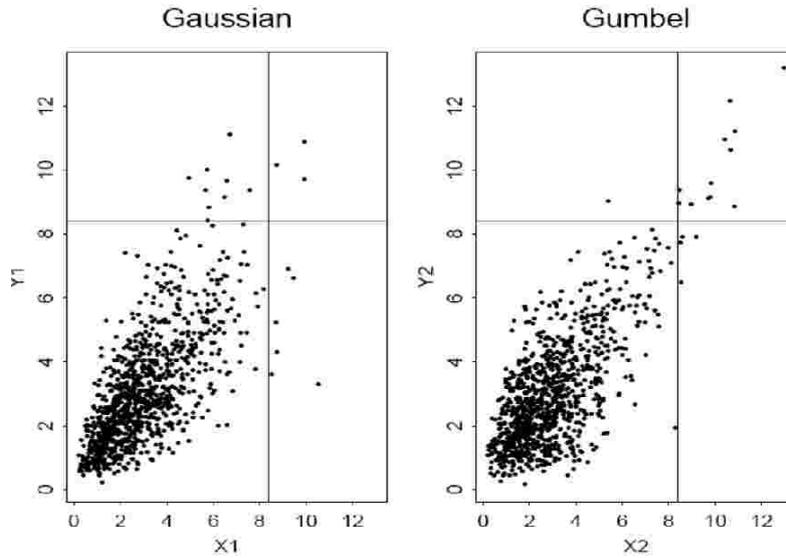


Figure 2.2: 1000 random variates from two distributions with identical Gamma(3,1) marginal distributions and identical correlation $\rho = 0.7$, but different dependence structures.

Hence the dependence structure directly explains the value that diversification granted

$$VaR_{diversified} = (1 - D)VaR, \quad 0 < D < 1, \quad (2.36)$$

where often D is sought to be anywhere from $[0.1, 0.3]$ or translated into a 10-30% reduction in diversified capital as pointed out by Embrechts (2012) [60]. We have seen this before in (2.15) where now we use a copula to perform the aggregation and perhaps receive a diversification benefit.

2.4.7 Element 3: Scenario Analysis

Much of our focus and contribution will be in the area of scenario analysis. In terms of regulatory guidance from the BCBS, there is minimal guidance for financial institutions to follow. Preliminary instructions for banks choosing to develop an AMA

model amounted to a single paragraph in Basel 2 (2006) [19]. Paragraph 675 states that

“A bank must use scenario analysis of expert opinion in conjunction with external data to evaluate its exposure to high-severity events. This approach draws on the knowledge of experienced business managers and risk management experts to derive reasoned assessments of plausible severe losses. For instance, these expert assessments could be expressed as parameters of an assumed statistical loss distribution. In addition, scenario analysis should be used to assess the impact of deviations from the correlation assumptions embedded in the banks operational risk measurement framework, in particular, to evaluate potential losses arising from multiple simultaneous operational risk loss events. Over time, such assessments need to be validated and re-assessed through comparison to actual loss experience to ensure their reasonableness.”

Since then, the BCBS has released added guidance in Principles for the Sound Management of Operational Risk (2011) [23] and Operational Risk Supervisory Guidelines for the Advanced Measurement Approaches (2011) [22], but with little clarification pertaining to methodology implementation. Essentially, the guidance is principle-based at best with the intent that scenario analysis provides a forward-looking view of operational risk exposures. It is only in the past 15 years that operational risk research has gained momentum and helped being shaped by both academia and industry.

A thorough development of the nuances of scenario design and the problems surround quantitative integration will be covered later. Nonetheless, the high-level intent of conducting a scenario analysis program is to elicit expert opinions of plausible futures losses from business line experts within a bank. A consequence of AMA modeling is that data sets are limited in their loss history and the requirement to capitalize for a 1 in 1000 year event is a difficult, if not an impossible task. It is rationalized that running workshops comprised of subject matter experts (SMEs) who have a detailed knowledge of their business are in a good position to theorize potential, future losses.

By showing a workshop of SMEs data from a variety of sources, the workshop participants may be able to formulate stress data points that could be used to modify a baseline LDA model.

Trying to formalize this process mathematically, the industry has converged on collecting estimates of frequency/duration of potential events and severity of events. That is, respondents are being asked to quantify worst-case losses that happen, e.g. one in five, one in ten or even one in forty years. Corresponding to a specific frequency estimate, a dollar impact to a particular business line/event type within a UoM is attached. Following this elicitation process, the method to integrate scenarios into a baseline LDA model is still subject to debate. As we will see, methods such as a change of probability measure, Bayesian techniques, and credibility theory techniques are just some mathematical and statistical tools being applied to the areas of scenario integration. The end-goal is the difficult task to satisfy both senior management within a bank and the independent financial regulator the utility that scenario analysis provides and that indeed the spirit of the guidance is being met.

If not controversial enough, the debate does differ across geography as well. On one end of the spectrum, it was shown by the Federal Reserve Bank of New York (2003) [67] and Colombo and Desando (2008) [44] that purely basing operational risk capital requirements on scenarios is a valid modeling approach. This was conferred upon from a working group composed of representatives from various financial institutions: Banca Intesa, Barclays Bank, Credit Suisse First Boston, Dresdner Bank, Fortis Bank, Halifax Bank of Scotland, Lloyds TSB, The Royal Bank of Scotland Group, UFJ Holdings Inc., and Euroclear. Moreover, an Italian bank has even gone as far as developing a scenario-based AMA model. In Canada, most banks strive to develop a four element AMA model with regulators adopting the view that scenarios are a value-added risk management tool. On the other end of the spectrum was the preliminary view by US regulators through guidance on Interagency Guidance on AMA

(2011) [33] that the subjective nature of scenario analysis may not warrant mixing of synthetic (scenario) data and observational (internal and external) data elements. Hence a second, benchmark model may be developed and a weighting between two models may provide the best answer. However, as research and new ideas started to take root, US regulators provided an updated view via Supervisory Guidance for Data, Modeling, and Model Risk Management for AMA (2014) [34] and stated that scenario analysis may be expanded upon its previous restricted use.

As a final comment on the debate of expert opinion used in quantitative models, there is a sort of awakening or tolerance for trying new things. There is a realization that models are inherently an abstraction of a phenomenon that come with limitations and assumptions. In that sense, a hybrid quantitative/qualitative model may provide a new direction in the future of modeling. This idea was strengthened by a comment made by the Bank of Canada Governor Stephen Poloz in the Globe and Mail in reference to (economic) models (2014) [86]

“The bank is now testing a variety of new models and methodologies to get a better handle on where the economy is headed, and updating its forecasts eight times a year. We are working hard to refine those models, but this experience is also leading us to put increased emphasis on anecdotal evidence - real conversations with real Canadians making economic decisions. That includes increasing its use of surveys, meeting with business associations and regular roundtables with business CEOs - to add colour to our economic analysis.”

Hence with an extension to quantitative modeling where expert judgement may be utilized, further research into incorporating this element may be a direction for new applications.

2.4.8 Element 4: Business Environment and Internal Control Factors

As a final element in AMA modeling, a BEICF assessment is usually treated as a post-modeling adjustment that is applied after all the statistical and mathematical modeling has taken place. The adjustment is usually applied to the allocated diversified capital and applied uniquely at each business line. Much like the guidance surrounding scenario analysis, Basel II (2006) [19] offered a single paragraph clarifying the expectation of what an assessment must entail. Subsequently, little added guidance has been provided in the BCBS' Supervisory Guideline (2011) [22] and BCBS' Sound Management Guideline (2011) [23]. The intent is that a BEICF adjustment is again a forward-looking assessment which focuses on business risk factors as well as a bank's internal control environment. It serves to recognize both improvements and deterioration in the operational risk environment. Hence it provides management the incentive to monitor and improve operational risk management practices. A methodology to transition from an assessment of business processes to reporting on results has been covered in Anders and Sandstedt (2003) [4]. The concept of an operational risk scorecard is defined therein outlining the necessary attributes that a scorecard must contain to be effective. A specialized set of such attributes that could be monitored are referred to as key risk indicators (KRIs) and defined in Davies *et al* (2006) [49]. KRIs are defined to be measurable metrics or indicators that track exposures or losses. KRIs should facilitate decision-making and provide a mechanism for monitoring the assignment of quantitative limits escalation thresholds if risks are viewed as elevated. In Table 2.12 and Table 2.13, a general overview of the BEICF acronym is explained and corresponding factors that could assigned, measured and scored.

Having established a possible number of factors to assess, banks have the ability to

Business Environment	
Factor	Examples
External Environment	Sovereign risk, market volatility, position in credit/business cycle
Internal Environment	Senior management change, lay offs, hiring
Regulatory/Legal	Change in regulation, pending lawsuits
Technology	Bank-wide upgrades, increased cyber attacks, automation

Table 2.12: BE factors.

Internal Control	
Factor	Examples
Audit scores	Audit rating and number of findings
Compliance	Identified conditions and time to resolution
Regulatory	New regulatory initiatives that impact internal control requirements
Technology	Infrastructure upgrades/deficiencies that impact internal controls

Table 2.13: IC factors.

design the weighting scheme as they deem fit. For example, let the realized score for each BE factor be labelled x_i where $i = 1, \dots, 4$. Each x_i may be assigned a grade on a scale from 1 to 10 with 10 indicating high risk. Similarly, each realized score for each IC factor be labelled x_i where $i = 5, \dots, 8$ with the same grading. Next weight each score

$$\sum_{i=1}^8 x_i w_i, \quad (2.37)$$

such that $\sum_{i=1}^8 w_i = 1$. Depending on the final BEICF score, a capital adjustment may be made. Typically in the industry, a symmetric range is used to adjust diversified regulatory capital anywhere from -10% to +10%. While the overall value of a BEICF assessment is high for risk management purposes, the quantification process and application for mathematical modeling is limited. Each bank would design unique scorecards and rely on simple weighting and summations to arrive at a score and an ultimate capital adjustment.

However, what seems to have been overlooked within the industry is the justification to modify post-modeling capital at the 1-year, 99.9 percentile. Within Davies *et al* (2006) [49], the argument is made that it is impossible for management to conceptualize and quantify key exposure indicators that influence extreme measures of risk. A suggested measure is the worst year loss predicted over a 10 year horizon, or unexpected loss (UL_{10}). This is interpreted as a probabilistic risk measure using a confidence interval between 90% and 95%. Hence the KRIs and management information used to manage the risk-and-control environment identify the more routine unexpected risk that occur once every 5-10 years. Adapted from Davies *et al* (2006) [49] is a graph showing the quantiles at which various AMA modeling elements would add value. This is depicted in Figure 2.3. Referring to 1. Risk mitigation programmes, we seldom observe in the industry banks trying to use insurance policies to offset AMA capital requirements. Whereas the regulation allows for up to 20% offset of total operational risk capital charge under AMA, banks do not actively petition to use insurance. Possible reasons for this is the difficulty in trying to justify the mapping of an insurance policy to a particular operational risk event. We actually touch on this in a later chapter and offer a transparent methodology for which insurance could be used. Business continuity planning (BCP) and control strategies technically influence the total loss distribution, however, the authors focus on the mitigation of the “less regular” operational risks.

Under 2. Tools & mechanisms, we see the elements used to form a BEICF assessment: KRI reporting and risk maps/control self-assessments (CSAs). This is to our knowledge the first discussion surrounding the proposal to limit the qualitative factor adjustments to UL_{10} rather than at $UL_{99.9}$. Conceptually, this is an intriguing proposal, however, Davies *et al* (2006) [49] do not provide a methodology to incorporate this information at the specific quantile. This is a potential area that may be further investigated and add value to the industry. We also observe that scenario analysis

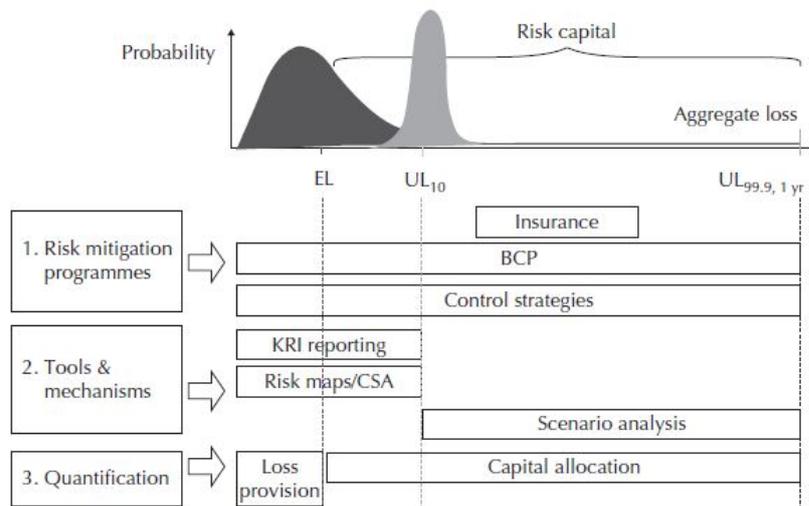


Figure 2.3: Quantiles for which AMA elements impact.

informs the tail of the loss distribution and hence inform capital requirements.

Corresponding to 3. Quantification, we infer that loss data in the form of ILD and ELD used to produce an aggregate loss distribution is used at every quantile. Under loss provisioning, EL occurs at the mean of the distribution for which risk managers and accountants would be highly interested to set aside an allowance to provision for expected loss. UL is of interest for regulatory capital for which the capital requirement is the sum of both EL and UL. This was highlighted in Figure 2.1.

3 Scenario Integration in AMA modeling

3.1 Overview

There is little debate that the starting building block of an AMA model is built using the LDA approach as seen in (2.8) and calibrated using some type of internal or external operational loss data. Having formed a loss distribution, scenario analysis and the methods to include the results of expert opinions obtained from scenario workshops into a data-driven AMA model does lend itself to an area of research and discussion. It is this area of scenario integration that is the focus of this chapter. We propose a method that recycles the concept of convolution and is used as a method of blending of two distributions. The algebraic properties of convolution also make it very flexible and open to many applications.

This technique has been published in The Journal of Operational Risk and authored by Aroda, P., Guergachi, A., and Huang, H. (2015) [8]. The title of the paper is *Application of convolution operator for scenario integration with loss data in operational risk modeling*. Minor formatting has been made in order to integrate the paper into this thesis. Namely, the introduction from the paper has been removed as the necessary material has already been covered in greater detail earlier in this thesis.

The organization of this chapter is as follows. Section 3.2 introduces scenario as-

assessment as a modeling requirement and covers the current direction of scenario quantification. Section 3.3 covers a literature review of scenario integration methods into a baseline LDA model. Section 3.4 revisits the convolution operator and investigates its use as a scenario integration technique. Stylized methods for scenario quantification are covered with the intent to show the mechanics of the method. Using descriptive statistics from a retail bank taken from the operational risk loss literature, the proposed scenario integration method is implemented with hypothetical scenario responses. Section 3.5 concludes with a discussion of the pros and cons of the method.

3.2 Scenario Formulation

We have already stated that the guidance provided to the financial community in terms of the design and use of scenario analysis was limited at best. Reading between the lines, the industry has interpreted the guidance as to assemble a team of business line experts and run workshops to elicit opinions of future, potential losses. By virtue of having seasoned experts and referencing internal loss history in conjunction with external losses faced by other banks, the workshop should be able to quantify exposure of the bank to high-severity events. In trying to distill what form a scenario would take, the banking community has come to some consensus on fundamental elements of a scenario. As pointed out in Ergashev (2012) [64], each scenario needs to be assigned a duration. That is, duration represents the number of years during which a particular scenario happens only once on average. However, when assigning loss severities to scenarios, the range of practice does vary somewhat. Banks may choose to collect scenarios with a range via an upper and lower bound, just a lower bound, or a point estimate. With a point estimate, it is also possible for a modeler to add a confidence interval around the estimate before inclusion. However, a simple way to arrive at some form of consensus around the magnitude is to always select the highest value of the range, interval or estimate. That is to say that with all frequencies being equal, the larger the loss, the larger the impact on capital.

Building on this preliminary concept of a scenario, Chaudhury (2010) [40] highlights many other challenges of scenario analysis and provides perhaps the most comprehensive review of the key issues in operational risk modeling as a whole. Among the many areas that pose a challenge for modeling, scenario/workshop data is highlighted as one such area. Chaudhury (2010) [40] points to issues that arise from the very method of designing a workshop to asking the necessary questions in order to extract sufficient information to be passed onto a modeling team to integrate the scenarios. Issues may also arise when dealing with potential cognitive biases or difficulty in mapping loss events to different probability values as pointed out by Barberis (2013) [14] and Kahneman and Tversky (1979) [79]. To mitigate problems with trying to quantify extreme probabilities, we limit ourselves to scenario responses anywhere from 1 in 2 to 1 in 40 year events. We believe it is easier for workshop participants to quantify these sorts of estimates as these losses may happen in the foreseeable future.

Chaudhury (2010) [40] also points out that it would be quite taxing on experts to give opinions directly on loss distributions since the loss distribution is seldom achieved by closed-form distributions. Hence, seeking opinions on frequencies and severities is most logical. Even with this clarified goal there exists some challenges. To calibrate severity distributions, one may ask participants to provide their opinion about the expected loss given that the loss exceeds various severities. This task of collecting key statistics would help calibrate the conditional probability of severity which would help calibrate the right tail of a severity distribution. As well, quantile information may be sought by fixing a severity and asking how often (once in a t -year event) a loss occurred. Finally, the author also talks about the method of dividing the severity domain into several buckets and asking participants the percentage of times they expect the severity of a loss to fall into different buckets.

We provide more detail into one such scenario elicitation method. The description

of the 1-in- t years method is very clearly explained in De Jongh *et al* (2015) [50]. A probabilistic explanation presented slightly differently regarding the same concept has been covered in the change of measure method in Dutta and Babbel (2010) [57]. The method starts off by asking scenario workshop participants what loss level L is expected to be exceeded once every t years. We envision that the true underlying annual loss frequency is Poisson distributed with parameter λ and the severity distribution is distributed according to continuous probability distribution $G(x, \theta)$. If the workshop has the ability to see into the future and predict the intended loss profile that has not yet materialized from observed internal loss data alone and therefore λ and G are known, then the assessments they provide should line up exactly with

$$L = G^{-1}\left(1 - \frac{1}{\lambda t}\right), \quad (3.1)$$

where G^{-1} denotes the inverse CDF of G . From this interpretation, we see that with probability $1 - \frac{1}{\lambda t}$, the level of L will not be superceeded. Stated differently, L is the $1 - \frac{1}{\lambda t}$ upper quantile of G .

It is noteworthy to point out that what we have labelled as true λ and G are not to be confused with the parameters calibrated on realized historical loss data. We denote these estimates by $\hat{\lambda}$ and a possibly different severity distribution $F(x, \hat{\theta})$. Hence if scenario respondents were to provide 1 in t year estimates that perfectly align with the realized historical loss experience, then their loss prediction \hat{L} would align up perfectly with

$$\hat{L} = F^{-1}\left(1 - \frac{1}{\hat{\lambda} t}\right). \quad (3.2)$$

We will exploit this difference later on when formulating a probability distribution used to quantify the scenario loss distribution. Another observation to be mindful of when collecting scenario estimates is to note that, for example, a 1 in 35 year event corresponding to a unit of measure that has $\lambda = 10$, one may look at the $1 - \frac{1}{350} = 0.997$ or 99.7th percentile of a CDF to find the value that will not be superceeded. So a

natural question that would arise is whether assessments go up to $t = 50$, $t = 100$ or beyond. Operational risk capital is mainly driven by infrequent large losses but human judgement may become unstable when deciding on the exactness of infrequent events.

3.3 Literature Review

Having shown just some of the challenges in arriving at a set frequency and severity, there are still a host of options to integrate scenario analysis with loss data. A thorough summary of the methods are detailed in Cope (2012) [45]. A popular method for scenario integration is the change of measure approach proposed by Dutta and Babbel (2010) [57]. The method starts from a baseline assuming that a modeler has fitted appropriate frequency and severity distributions in order to convolve to obtain a loss distribution. However before performing the n -fold convolution, the density function $f(x)$ of the severity distribution is adjusted to reflect the impact of scenarios. Scenarios are collected with frequency estimates taking the form $m \div t$ where m is the number of times the event is expected to occur in t years. Severity estimates are collected with a loss range $[a, b]$. The methodology aims to answer the question, “*Given that the scenarios are tail events, how much does $f(x)$ need to be adjusted so that its probability for those events ‘match’ the probabilities implied by the scenario*”. The implied severity distribution incorporating the scenarios will have losses from the range $[a, b]$ occurring m times in t years. It is noteworthy to point out that further clarification was given to the financial community in Babbel (2010) [12] that defended the change of measure approach as a reweighting of the probability measure. That is, the numerous ways to select a reweighting is accompanied by an objective – the authors choosing to reweight historical probabilities of tail events to match those given by scenarios. Cope (2012) [45] pointedly described the unique method to reweight as not a method to integrate of scenarios per-se, but rather to override and replace historical estimates of frequency with scenario estimates of frequency and an ensuing

update of the severity distribution.

Ergashev (2012) [64] introduced a method of scenario integration that filters scenarios whereby only worst-case scenarios influence capital estimates. Scenarios are collected with a duration which represents the number of years during which a particular scenario happens only once on average. A scenario is defined as a “once-in-a- M -year” scenario if it has a duration of M years. The convention is that among two worst-in-a-certain-year events, the one with a larger duration must be more severe than the other. For a severity estimate, a scenario expert need only assign a lower bound of an unknown loss amount. The scenario is envisioned to be a random realization with unknown loss amount and duration where a scenario probability distribution unifies all scenarios to have a common unobserved continuous distribution. Hence Ergashev (2012) [64] differentiates between two probability distributions: one from the base model which typically follows from the loss distribution approach and the scenario probability distribution. The filtering method determines which scenarios are “concordant” or “discordant” with the base model’s severity distribution. The rule is essentially an inequality that is to be checked to see if at a particular quantile, any adjustment is necessary. Scenarios that are concordant with the base model are deemed uninformative in that no adjustments would be necessary to the base model’s severity distribution at the corresponding quantile. Only scenarios that are discordant with the base model signal a change in the institution’s risk profile and the underlying loss distribution. The paper concludes by presenting five alternative approaches to scenario integration which all rely on the scenario filtering rule: three direct methods via a scenario-adjusted CDF and two direct methods via a closest-curve type minimization technique.

Changing direction, there are an increasingly number of proposals to the industry at using Bayesian inference methods for scenario integration. One such method has been detailed in Shevchenko and Wüthrich (2006) [106]. The method proceeds by

allowing expert opinions to be incorporated into the analysis by specifying prior distributions for model parameters. The prior is to be estimated via scenario analysis using expert opinions and external data to guide the quantification. Bayes' theorem is used to weight the prior distribution accordingly in order to achieve a posterior distribution. Finally the predictive distribution, both frequency and severity, of future observations conditional on all present information is derived. The benefit of the method is that there is a dynamic ability for experts to reassess prior distributions as new information is brought online.

A Bayesian approach combined with a least square approximation has led to “the greatest accuracy credibility theory”. A nice treatment of the topic is covered in Klugman *et al* (2012) [81] arising from an insurance context. Therein, full and partial credibility approaches are described. In brief, when faced with choices between two estimators (e.g. expected value of a risk with sample mean \bar{X} and class mean μ), it is possible to assign full weight to one element or a split between the two via a credibility factor. A paper by Bühlmann *et al* (2007) [37] shows an example of a “toy” model where a credibility approach is adopted to estimate frequency and severity distributions for low frequency/high severity cells by using ILD/ELD data and expert opinion. The tiered approach starts with MLE to calibrate parameters based on data in a risk cell, followed by an improved credibility estimator using bank-wide data and finally an improved credibility estimator using industry data. The expert opinions are interpreted as scalars to parameter estimates which augment the risk profile of the cells.

Yet another use of advanced credibility theory was shown in a paper by Agostini *et al* (2010) [2]. This confirms that options are still being presented to the operational risk community in order for practitioners to chose among viable options. The authours form VaR estimates based on historical loss data and another subjective VaR using expert opinion collected during scenario workshops. The integration oc-

curs at the frequency and severity parameter level whereby partial credibility is used and the credibility weights are given by the Bühlmann-Straub model. The parameter estimates are driven by historical losses and corrected using information coming from expert opinions. The frequency parameter, coming from a Poisson distribution, is easily weighted using advanced credibility theory. The tail of the severity distribution is modeled by a Generalized Pareto Distribution which follows from Extreme Value Theory over a stated threshold. The three parameter distribution is simplified by setting both scale and threshold parameters to coincide with historical values and the free shape parameter is thus back-out of a simple quantile relation. An obvious short-coming of the method is that worst-case scenarios are defined to coincide to the 99.9th percentile for which scenario experts would have a hard time comprehending. The authors do provide a case study based on an Italian banking group using scaled loss data over five years. When focusing on a particular event type where 9 scenarios were available, only 7 provided sufficient information to calibrate the free parameter for the severity distribution because expert opinions fell below the threshold and hence fell in the body. Hence, calibrating the distribution requiring such a high quantile proved difficult. Nonetheless, the final integrated VaR did provide a trade-off between both historical VaR and subjective VaR by using the credibility theory to weight each parameter and provide a single loss approximation to produce an integrated VaR.

As another application using a Bayesian framework, Cope (2012) [45] suggests a method of integrating data and scenarios at the loss-generating mechanism (LGM) level by using Bayesian techniques and the Dirichlet process as a prior distribution. In that sense, there is a two-fold introduction of “new concepts”: the LGM and the Dirichlet prior. Cope (2012) [45] asserts that within the unit of measure, there are heterogenous processes that are evident and are usually overlooked. Hence, there are distinct underlying processes that, as a collective superposition, define the unit of measure. This translates to taking the assumption that the severity of losses in

(2.8), which are assumed to be independent and identically distributed according to a common distribution F_i , are further represented as a mixture of severities from each underlying LGM: $S_i = \sum_{j \in \text{LGM}_i} \sum_{k=1}^{N_{i,j}} X_{ijk}$. In that sense, a scenario is defined on a more granular scale where scenarios are linked to a given LGM. Next, priors are formed which captures the uncertainty around a single parameter. For the loss frequencies, a Poisson model is assumed with the associated conjugate prior being the gamma distribution over the Poisson parameter λ . In terms of a nonparametric form for the severity, the Dirichlet process is represented in terms of a base distribution estimate and a scalar parameter. That is a scenario-based estimate of the severity distribution, denoted $\hat{F}(x)$, is made together with a certainty parameter (measure) c . The resulting posterior distribution is a weighted average of the prior estimate $\hat{F}(x)$ with weight $\frac{c}{(c+n)}$ and empirical CDF of the data with weight $\frac{n}{(c+n)}$. Nonetheless, an implicit assumption when working with an LGM is that there exists an underlying process governing a loss profile that does not vanish over time. Hence, no break would be given to an organization implementing a control to mitigate a string of observed operational losses.

3.4 Convolution to Integrate Scenarios

The previously discussed methods have varying degrees of difficulty in the sense that a necessary goal of a modeler is to be able to get buy-in from non-technical experts in addition to providing a robust method of scenario integration. At the core of our proposed method is a simple take on scenario integration. Our method essentially is a convolution between a loss density derived from frequency and severity estimates from scenario opinions and the loss density obtained from the convolution of frequency and severity distributions coming directly from operational loss data. That is, the process of adding two continuous random variables is akin to the convolution of their probability densities. Hence, to quantify the effects of ILD/ELD and scenarios, summation is the natural operation to obtain a model that incorporates all elements.

We contrast our method with other techniques for scenario integration.

1. On the one hand, one may seek to refine or modify the parameters of the underlying distributions used to build the loss distribution. For example, based on the information gained using scenarios, frequency and severity parameters are adjusted using Bayesian inference methods. Alternatively, the change of measure method could be used to reweight the probability measure for the severity distribution.
2. On the other hand, we propose that scenarios represent the missing piece of information that must be added. For example when working with the loss distribution, we claim that the final loss distribution to be used for regulatory capital = baseline LDA + “missing piece”. This may also work on the parameter level for which the convolution between two distributions may occur. For example, final Poisson parameter = baseline λ + “missing piece”.

At a high level, the method satisfies a number of useful characteristics for scenario integration:

- The convolution was already used in determining the loss function for a unit of measure. That is, both frequency and severity distributions were combined through a n -convolution of the severity distribution with itself, where N is a random variable that follows a frequency distribution. Hence, the convolution technique may be recycled again for scenario integration.
- Drawing on the analogy when working with deterministic L^1 functions on \mathbb{R}^n , there exists the notion of “flip and drag” whereby one function is used to smooth another via the convolution. When extending this concept to probability distributions, the intuition is still a mixing effect that has the visual effect of smoothing as well. We will define this later.

- Convolution is commutative. Hence there is no set ordering necessary to first determine loss distribution based on loss data and then layer on scenarios. The end result is a symmetric application of the convolution operator.
- There are many possibilities for using the convolution operator. For example, it is possible to convolve three separate curves: (i) internal loss data, (ii) external loss data, (iii) scenarios. In addition, it is possible to blend-in other types of external loss data sets or other scenario libraries.
- The regulatory guidance surrounding scenarios calls for the need to avoid anchoring of responses, bias or uncertainty through estimates. The convolution resolves this problem in that we are not constrained to the baseline LDA parametric assumption.

3.4.1 Convolution Operation

Within the fields of signal processing and medical imaging, convolution is widely used as a method of smoothing noisy signals. The concept of a filter is defined as any operation that maps inputs to outputs, for which the convolution may also be classified as a filter. A thorough treatment of signal processing, transforms and filtering are described in Epstein (2008) [63]. An intuitive primer on convolution is detailed by Osgood (2007) [99]. To motivate the application to scenario integration, we recall the definition from a deterministic case.

Definition 3.4.1. (*Convolution - deterministic*) *If f is an L^1 -function defined on \mathbb{R}^n and g is bounded, locally integrable function then the convolution product of f and g is the function on \mathbb{R}^n defined by the integral*

$$(g * f)(t) = \int_{\mathbb{R}^n} g(t - x)f(x)dx. \quad (3.3)$$

In the deterministic function sense, Osgood (2007) [99] points to the notion of “flip

and drag”. That is for a fixed value t , the graph $g(x - t)$ has the same shape of $g(x)$ but shifted to the right by t . Hence $g(t - x)$ flips the graph (left-right) about the line $x = t$. Then the two functions $g(t - x)$ and $f(x)$ are multiplied and integrated with respect to x . The result can be thought of a smoothing or averaging technique. That is, g is used to smooth f in $g * f$.

Definition 3.4.2. (*Convolution - random*) When X and Y are independent random variables with probability densities f_X and f_Y respectively and $Z = X + Y$, then the density of f_Z is given by the convolution

$$f_z(Z) = f_Y * f_X = \int_{-\infty}^{\infty} f_Y(z - x)f_X(x)dx. \quad (3.4)$$

Since the resulting convolved density will yield a corresponding CDF that stochastically dominates the baseline LDA model as seen in Levy (2006) [83] (adding a positive random variable with non-degenerate distribution would have $F_Z > F_X$), there is always the ability to scale the LDA or scenario model to temper/moderate the effect. This introduces an additional problem to calibrate the weighting coefficients. That is, we can take a linear combination or scale any individual random variable. This detail is required in order to perform the convolution and scale the associated PDF/CDF. For this, we will also require the definition for scaling a random variable.

Definition 3.4.3. Let X be a continuous random variable with PDF $f_X(x)$ and CDF $F_X(x)$. Let $Y = \theta X$ with $\theta > 0$. Then $F_Y(y) = F_X(\frac{y}{\theta})$, $f_Y(y) = \frac{1}{\theta}f_X(\frac{y}{\theta})$.

Before moving on and having now formally defined the convolution operation for continuous random variables, we further justify this method. One take on interpreting the problem occurs when Z represents the aggregate loss distribution incorporating data X and scenarios Y , where X is known and Y is missing. One can view scenario analysis as the “missing data” of the same distribution represented by the base case or originated from a different sources. In the former, if both scenario and base case

are from the same distribution, the question becomes what is the distribution? If some scenarios have already occurred, we have already fit Z through the observed data. In that case, $X = Z - Y$ which gives scenario-free data.

However, we have approached this problem from the latter stance. The case $Z = X + Y$ where X and Y are originated from different sources is another interpretation and resonates with the intent of producing scenarios that have not occurred. Currently banks have calibrated X , if based purely on internal loss data (ILD), on 10-15 years of data. The intent of Y is to help steer the risk quantification to align with those 1 in a 1,000 year events. One may assume that both X and Y have some dependence but based on the gap that is trying to be bridged and the uncertainty involved in human judgement, independence (and hence summation of random variables) may have merit.

Embrechts and Hofert (2011) [61] state that scenarios may provide a forward-looking method for capturing tail events that may not have occurred in the bank's loss history. By extension, rather than supplementing the same distribution to model the tail, viewing our technique through another lens may have ILD modeling quantify a backward-looking view and scenarios a forward-looking view. Cited within the same context in Embrechts and Hofert (2011) [61] is the BCBS principles for sound stress testing practices and supervision guidance (2009) [20] which is relevant for motivating this topic. Reproducing some of the guidance, we learn that stress testing should provide a complementary and *independent risk perspective* to other risk management tools such as value-at-risk and economic capital. In addition, as stress testing allows for the simulation of shocks which have not previously occurred, it should be used to assess the robustness of models to possible changes in the economic and financial environment. As well, banks should also simulate stress scenarios in which the model-embedded statistical relationships break down. Since we interpret scenarios to be the stress testing element within the AMA model, by assuming independent shocks our

scenario integration technique could be well served.

3.4.2 Scenario Distribution Formulation

The basis of using this method relies on the fact that a probability distribution can be formulated based on the responses from a scenario workshop. That is the necessary information to build a probability distribution would be a series of estimates of frequency and severity. In its most general form, both parameters could be collected with a range

$$\left(\left[\frac{1}{t}, \frac{1}{u} \right], [a, b] \right) \quad (3.5)$$

with an event occurring with a 1 in t or 1 in u year duration with magnitude $[a, b]$. From a modeling perspective, selecting the minimum frequency with the largest end severity loss could add a layer of conservatism for which the ordered pair would be

$$\left(\frac{1}{t}, b \right). \quad (3.6)$$

As seen from (3.2), a 1 in a t year event may be mapped to the quantile of an severity distribution. As from Ergashev *et al* (2013) [65], elicitation to calibrate a symmetric probability distribution may take the form of responses about the mean and standard deviation. It was pointed out that elicitation to arrive at a skewed distribution is easier if expert opinion is sought on quantiles instead; for example, first quantile, median, third quartile. In addition, for practical reasons, frequency estimates limited to 1 in 30 to 1 in 40 year events are realistic and coincide with an employee's career within an institution. For human contributions to be of value, there is a need to keep things simple. For reasons of stability, Dutta and Babbel (2010) [57] notice that scenarios with short frequencies (between 20 and 25 years) cause for little concern for the change of measure approach. It is when scenarios with long term frequency estimates are used (e.g. 100 year), that complications arise in terms of an adequate sample size needed to implement the change of measure approach.

In terms of the number of scenarios required to calibrate a frequency distribution, a Poisson distribution would require only an estimate of λ and the minimum number of scenarios to calibrate a two-parameter severity distribution would require two loss estimates to calibrate (e.g. μ, σ). While this does seem overly simplistic, we do see in Frachot *et al* (2003) [70] a similar requirement in a worked-example. Therein, three different scenarios were required to calibrate the three parameters λ, μ, σ . The calibration proceeded via a quadratic criterion to be minimized which could also be used when a system is over-determined.

To offer practitioners tasked to conduct workshops and collect suitable information to calibrate scenario probability distributions, we offer additional references. Shevchenko and Wüthrich (2006) [106] also needed the ability to estimate structural parameters subjectively to determine the prior distributions for their Bayesian inference method when combining loss data with expert opinion. To overcome this shortcoming, the reference pointed out was that of Berger (1985) [31]. Applicable to our example, the method of matching a given functional form was adopted. That is, finding scenario parameters assuming some functional form to match subjective beliefs. Related to this also would be the CDF determination technique that produces directly a smooth curve. As well in Cruz *et al* (2015) [47], we see examples of how expert judgement could be used to calibrate probability distributions. We see examples of fixing an annual frequency and asking for likelihoods of losses exceeding certain thresholds to calibrate a severity distribution. Conversely, opinions of loss frequencies in pre-defined severity buckets allow for estimates of annual frequencies. Both methods could be subjected to goodness of fit tests depending on the number of data points collected. This would avoid the distributional assumption we have made from the onset. Nonetheless, we made this simple assumption just to show the mechanics of our proposed method.

Comparing the frequency probability distributions between the distribution derived from ILD/ELD and scenarios, more often than not the Poisson distribution may be selected for simplicity and consistency. Hence when comparing severity distributions, it is possible to also match the choice of severity distribution derived from ILD/ELD (chosen from several goodness of fit statistical tests) and inferring the same, albeit different parameters, severity distribution for the tail/scenario region. However, some banks do make an ex-ante assumption that the scenarios follow a fat-tailed (eg, Log-normal, Weibull,...) distribution in order to calibrate a severity distribution.

In the following example, we can see by using the quantile function of a corresponding probability distribution that we may recover a calibrated distribution to the severities of scenarios. We must stress that this is not a comprehensive method to calibrate a probability distribution. We simply introduce such a method as a conversational starting point to novice practitioners who would like to build-in sophistication once the mechanics are understood. We acknowledge that additional methods exist that yield smoother probability distributions. Timely to this discussion, in De Jongh *et al* 2015 [50] we see an investigation into two cases: the Generalized Pareto Distribution (GPD) motivated from Extreme Value Theory (EVT) (a three parameter distribution where one parameter serves as the EVT threshold) and the three parameter variant of the Burr type XII distribution. Whereas the former case was motivated by modeling the tail, the latter case was motivated by selecting a best fitting probability distribution based on data from the SAS OpRisk Global database. Once the foundation was laid, we then see the discussion of using the minimum number of points for calibration: namely 1 in 7, 1 in 20 and 1 in 100 year loss estimates. Subsequently, we see alternative methods of scenario integration in De Jongh *et al* 2015 [50]. We reiterate that for our discussion, we required a continuous probability loss distribution and hence needed some building block to begin with. Without the luxury of dealing with a single financial institution with a particular risk profile, loss experience, scale and geographical footprint, any example must be purely hypothetical and in that sense,

we resort to the most simplest of starting points.

Example 3.4.1. *Consider the case where a view is taken that the scenarios follow a Weibull distribution. To make this more tractable, consider the case where the shape parameter (k) is less than 1 to ensure that the distribution is representative of a common heavy-tailed distribution. Recall that the PDF of a 2-parameter Weibull distribution is given by*

$$f(t) = \frac{k}{b^k} t^{k-1} \exp\left(-\left(\frac{t}{b}\right)^k\right) \quad t > 0,$$

with shape parameter k and scale parameter b . The corresponding quantile function is given by

$$F^{-1}(p) = b(-\ln(1-p))^{\frac{1}{k}} \quad 0 < p < 1.$$

Hence a scenario impact at a particular quantile is given by $F^{-1}(p)$. Then what is left to solve is a system of two equations with two unknowns where p_1 and p_2 are percentiles corresponding to severities v_1 and v_2 respectively

$$v_1 = F^{-1}(p_1) = b(-\ln(1-p_1))^{\frac{1}{k}}, \quad (3.7)$$

$$v_2 = F^{-1}(p_2) = b(-\ln(1-p_2))^{\frac{1}{k}}. \quad (3.8)$$

From (3.7) we have

$$b = \frac{v_1}{(-\ln(1-p_1))^{\frac{1}{k}}}. \quad (3.9)$$

Then substituting (3.9) into (3.8) we have

$$\begin{aligned}
v_2 &= \frac{v_1}{(-\ln(1-p_1))^{\frac{1}{k}}} \cdot (-\ln(1-p_2))^{\frac{1}{k}} \\
\frac{v_2}{v_1} &= \left(\frac{-\ln(1-p_2)}{-\ln(1-p_1)} \right)^{\frac{1}{k}} \\
\left(\frac{v_2}{v_1} \right)^k &= \frac{\ln(1-p_2)}{\ln(1-p_1)} \\
k \ln\left(\frac{v_2}{v_1} \right) &= \ln\left(\frac{\ln(1-p_2)}{\ln(1-p_1)} \right) \\
k &= \frac{\ln\left(\frac{\ln(1-p_2)}{\ln(1-p_1)} \right)}{\ln\left(\frac{v_2}{v_1} \right)}. \tag{3.10}
\end{aligned}$$

So from (3.9) and (3.10) we have recovered the parameters for our density/cumulative distribution function.

This example has shown a simple, idealized method to calibrate a probability density function. This combined with the Poisson process allows for the characterization of the loss distribution that may be used in the convolution method to integrate scenarios. For three or four parameter probability distributions, a similar method could be used but more points along the curve would need to be extracted from scenario respondents.

To see an alternate method using specific quantiles, we may again use a Monte Carlo sub routine to produce a distribution. For example, for a given frequency, two opinions would need to be collected regarding the quantiles of the severity distribution for an assumed a-priori probability distribution fit. Assuming this time the Lognormal distribution is used, then the median of the Lognormal distribution is given as e^μ where μ is associated with the mean/median/mode of the normal distribution. From (3.6) we have

$$\mu = \ln(b_{0.5}) \tag{3.11}$$

where $b_{0.5}$ is an estimate of a median loss. To find σ , we use the definition that the Lognormal quantile function is given by

$$F^{-1}(p) = e^{(\mu + \sigma\Phi^{-1}(p))} \quad 0 < p < 1. \quad (3.12)$$

Hence the loss at the third quartile implies $b_{0.75} = F^{-1}(0.75)$. Solving for σ we have

$$\begin{aligned} b_{0.75} &= e^{(\mu + \sigma\Phi^{-1}(0.75))} \\ \ln(b_{0.75}) &= \mu + \sigma\Phi^{-1}(0.75) \\ \ln(b_{0.75}) - \ln(b_{0.5}) &= \sigma\Phi^{-1}(0.75) \\ \frac{\ln(b_{0.75}) - \ln(b_{0.5})}{\Phi^{-1}(0.75)} &= \sigma. \end{aligned} \quad (3.13)$$

Having calibrated a frequency distribution and a severity distribution, the Monte Carlo sub routine may be performed whereby for a given realization of the Poisson process, a corresponding draw of that number could be taken from the severity distribution. The sum of realizations for a given draw would represent an annual loss for a given year. Iterating for many times and plotting a histogram would show the probability distribution of the scenario loss distribution.

3.4.3 Application of Convolution Method

We are now in a position to apply the convolution operator using our building blocks. To recap, we have at our disposal Table 3.14 and outline the approach to integrate scenarios as follows:

Convolution approach to integrate scenarios

1. From an internal/external data set, calibrate frequency probability mass function (PMF) and severity PDF $p(x)$ and $f(k)$ respectively.
2. Obtain aggregate loss $G(x)$ from the n -fold convolution of severity and frequency distributions (associated PDF is denoted as $g(x)$).

Building block	Definition
$p(k)$	Fitted frequency PMF for a unit of measure.
$F(x)$	Fitted severity CDF for a unit of measure, with corresponding PDF $f(x)$.
$G(x)$	Convolution of severity and frequency to obtain loss distribution, with corresponding PDF $g(x)$. The associated random variable for this loss will be labelled X .
$H(y)$	Fitted loss distribution to scenario data, with corresponding PDF $h(y)$. The associated random variable for this loss will be labelled Y .

Table 3.14: Elements needed to perform baseline LDA and scenario convolution.

3. Calibrate scenario PDF $h(y)$ based on responses from a scenario workshop.
4. Convolve $g(x)$ and $h(y)$ to obtain $c(z)$ for which $C(z)$ represents the aggregate loss distribution including scenarios.

In our case we used a simple form of obtaining a fitted distribution to scenario data by working with minimum number of points. However, the method is flexible enough to work with any level of robustness in calculating the distribution. What we propose is that the convolution of $h(y)$ and $g(x)$ given as

$$c(z) = (h * g)(z) = \int_{-\infty}^{\infty} h_Y(z - x)g_X(x)dx \quad (3.14)$$

allows for the quantification of the CDF $C(z)$, which taken at the 99.9 percentile is the $\text{VaR}_{99.9}$.

While (3.14) conveys the method and the concept of independence between the two loss distributions, we make the adjustment of taking the mean $(X + Y)/2$ while maintaining independence.

$$c(z) = \frac{1}{2}(h * g)(z) \quad (3.15)$$

We observed the effect of scaling PDFs in Definition 3.4.3. We rationalize this by noticing if $Y \equiv X$, we would want the loss quantified by X and not $2X$. Again, the

choice of the scalar is subject to debate and calibration. By taking the mean, we obviously address the case where capital requirements would more than double if no scalar were in place.

To provide the mechanics of this method, we provide a numerical illustration with parameters from Jiménez-Rodríguez *et al* (2011) [78] and apply the technique above. We reiterate by reminding readers that operational risk loss data is difficult to obtain. Data consortiums such as ORX operate on a give-and-take basis whereby international banks submit their loss data to ORX where it is anonymized and pooled to be made available only to contributing members. IBM Algo FIRST contains a case-study approach to operational risk loss events and provides a qualitative description of publicly disclosed losses. Again, the data set is available at a cost and limited in its distribution for research.

Jiménez-Rodríguez *et al* (2011) [78] provide descriptive statistics of loss data from a medium-sized Spanish bank that operates within the retail banking sector. The sample is comprised of seven years of historical operational risk losses. The paper provides a sensitivity analysis on varying levels of data collection thresholds of 25, 50, 100, 1000, 5000, and 10,000 Euros. The Poisson distribution is selected for frequency calibration and candidate severity distributions are selected from Lognormal, Weibull, and Pareto and calibrated using the MLE technique. The Kolmogorov-Smirnov (KS) test is used as a goodness-of-fit test to choose severity distributions. The parametric best fits are shown in the paper together with the parameter estimates.

For our purpose, we may collapse all loss data across event types and assign the data to the retail banking business line as was done in the paper. The champion fit from the paper is provided in Table 3.15. It is possible to recover the loss distribution solely based on this data. Monte Carlo methods are used to construct the compound distribution of using frequency and severity densities.

Threshold	Frequency	Severity	LDA VaR 99.9%
€10,000	PO ($\hat{\lambda} = 16.73$)	LN ($\hat{\mu} = 10.129, \hat{\sigma} = 0.862$)	€1,542,567

Table 3.15: Champion fit with PO = Poisson and LN = Lognormal.

Next, hypothetical scenarios were formulated to be combined with the LDA model. The way we form these estimates is we perturb the quantiles that are implied from the underlying data. This idea was motivated from the paper by De Jongh *et al* (2015) [50]. For instance, we use $\hat{\lambda}$ and \hat{F} from Table 3.15. Using (3.2), we ask ourselves for 1 in t year events where $t = 5$ and $t = 35$, what are the implied loss amounts? We select these durations as they are within the range of what banks typically use. We call these \hat{L}_5 and \hat{L}_{35} which when backed out are €175,589 and €312,580 respectively. We then perturb these loss amounts up and down and allow ourselves to fit a new \hat{F} . Note, since we are not limited to any a-priori family of distributions, we may resort to the Weibull distribution. This could be based on a variety of reasons: high severity losses as collected from ELD may be better suited with an alternative distribution or a more robust scenario elicitation method could be used for which the Weibull is a better fit. Nonetheless, we fix $\hat{\lambda}$ and the 1 in t year events, although we could change these as well. We carry out the n-fold convolution of frequency and severity to arrive at a loss distribution for scenarios. We use (3.15) to produce a convolved capital requirement. These results are summarized in Table 3.16.

Comparing the VaR estimates in Table 3.15 and Table 3.16, we get a sense of the contribution of each possible scenario. When shocking the first loss estimate, we see the effect of reducing the loss amount by 30%, fitting a Weibull distribution and using that as the severity distribution to determine capital requirements based on the scenario estimates alone. However, when combined with the LDA model using the independence and averaging assumption, we naturally see a bias downward due to the contribution of the scenario. We then notice a general pattern of increasing

1 in 5 year	1 in 35 year	Scenario VaR(99.9%)	Convolution VaR(99.9%)
$0.7\hat{L}_5$	\hat{L}_{35}	1,138,193	1,281,665
$0.8\hat{L}_5$	\hat{L}_{35}	1,081,799	1,284,248
$0.9\hat{L}_5$	\hat{L}_{35}	1,113,394	1,320,921
\hat{L}_5	\hat{L}_{35}	1,237,189	1,387,259
$1.1\hat{L}_5$	\hat{L}_{35}	1,438,433	1,484,045
$1.2\hat{L}_5$	\hat{L}_{35}	1,755,268	1,630,826
$1.3\hat{L}_5$	\hat{L}_{35}	2,214,494	1,848,069
\hat{L}_5	$0.7\hat{L}_{35}$	2,172,263	1,806,347
\hat{L}_5	$0.8\hat{L}_{35}$	1,575,216	1,531,662
\hat{L}_5	$0.9\hat{L}_{35}$	1,318,888	1,425,858
\hat{L}_5	\hat{L}_{35}	1,237,189	1,387,259
\hat{L}_5	$1.1\hat{L}_{35}$	1,233,446	1,381,631
\hat{L}_5	$1.2\hat{L}_{35}$	1,285,395	1,398,278
\hat{L}_5	$1.3\hat{L}_{35}$	1,422,416	1,436,329

Table 3.16: VaR due to scenarios alone and VaR from convolved baseline LDA and scenario model.

VaR as we increase the 1 in 5 year loss estimate keeping everything else fixed. It is noteworthy to point out that by using (3.1), a 1 in 5 year event and 1 in 35 year event are scaled by the Poisson parameter ($\hat{\lambda} = 16.73$). Thus, it is natural to see how improvements could be made to improve the calibration of the scenario loss distribution by collecting losses at various quantiles. Finally, for the 1 in 35 year loss, we see a peak capital requirement at the lower shock level due to this lower loss occurring at the same percentile.

3.5 Summary

When using the AMA methodology to set regulatory capital requirements for operational risk, scenario analysis is a modeling element that is used to layer-in expert judgement for exposure to high-severity events. We have proposed the use of convolution which has wide applications in the fields of signal processing, electrical en-

gineering, mathematics and statistics. As a mathematical operation, one function is used to smooth and average the other as it is translated through the domain.

There are many advantages to using the convolution to integrate scenarios. The averaging effect is an easy method to convey the concept to non-experts. Moreover, the mathematical operations are straightforward - a model that integrates scenarios is simply a sum of two random events to produce one combined result with some scaling to be determined. In addition, since we interpret the regulatory guidance around scenario design with the goal to avoid anchoring and strive for independent assessments, using independent probability distributions and not necessary the same family of distributions and incorporating them into a baseline LDA model meets this spirit. This may also be useful in stress testing exercises. Ever since the 2007/2008 financial crisis, there has been increased focus on stress testing. One such requirement is for banks to conduct a macroeconomic stress test either to meet a regulatory requirement or an internal mandate of enterprise-wide stress tests. One such proposal to evolve operational risk capital is to produce changes to operational risk capital over a projected future time horizon. The inclusion of subsequent scenario curves may be one such area that could be investigated and using unforeseen, regulatory prescribed exogenous shocks would resonate well the independence assumption.

Nonetheless, this technique is not without its faults. The very core of the assumption is also a contestable subject: the assumption of independence between scenario and data-derived distributions. This assumption may be accepted if scenarios are viewed as a stress testing element that are implemented for the very reason to augment underlying assumptions based on limited data. Also, the application of the methodology fundamentally relies on the ability to calibrate a probability distribution to scenario data. While the frequency calibration has always been somewhat simplistic, the calibration of severity and aggregate loss distributions could be subject to much improvement. The applications of optimization algorithms or statistical

sampling to produce improved distributions around scenarios is an area for advanced research. Then, any refinements may be easily incorporated as plug-and-play into this framework.

4 Enterprise-Wide Scenario Analysis

4.1 Overview

Up until now, we have seen scenarios been narrowly defined at the UoM level of granularity. That is to say, scenarios are asked to be developed for a specific business line and then depending on the nature of the scenario, slotted into a particular event type. Hence, the scenario becomes associated with a particular cell in the 56 cell matrix, as see in Table 2.6. Procedurally, this process has naturally evolved from the functional requirement given to the operational risk management team within a bank. The typical steps in a scenario analysis program have been outlined in Haackert and Regling (2012) [73]. Tasked to hold workshops, assemble relevant internal loss data, relevant external loss data, and key information on business strategy, political environment and the economy as a whole, the scenario design team would systematically proceed business line by business line. The siloed approach to scenario design does not encourage enough collaboration among different experts from different business lines to think of aggregate potential loss to a bank. This is partly to be blamed with a lack of research in the area and lack of guidance. We have already seen various proposed methodologies for scenario integration that provided the banking industry with options to think of scenarios on the UoM level. Hence, this could have partly steered the industry in a direction of cell by cell scenario design. However, in a few instances, we do see some banks starting to develop enterprise-wide scenarios.

When banks do formulate enterprise-wide scenarios, we have not seen any structured

process that details roles and responsibilities. Moreover, we see that enterprise-wide scenarios are seen as something of an after thought. Preliminarily, we see that the only use of an enterprise-wide scenario is to verify if the severity impact of the scenario falls below the calculated operational risk capital requirement as stated by the OCC (2012) [95]. As a hypothetical example, if an enterprise-wide scenario is thought to effect the entire bank and has an impact of \$800 million but the LDA model calculates a capital requirement of \$3 billion, then in this particular case the bank would not have to alter the LDA model. The scenario usage could be seen as a validating tool that has validated the capital requirement as being more than adequate of absorbing the catastrophic loss. Although this seems somewhat elementary, these are the types of simplistic interpretations banks are making.

While there is value in thinking of scenarios on the UoM level, by the very nature of operational risk capital being highly sensitive to idiosyncratic, low frequency/high severity events, devastating catastrophic scenarios spanning multiple cells is something to be taken into consideration. For example, we have already seen some low frequency/high severity events in the past that have been mapped to single UoMs. Examples are detailed in Table 4.17. Although referring to these types of large losses and trying to formulate a potential impact to one's host institution is part of scenario analysis, there is a missing line of thinking.

If anything, catastrophic losses that systemically effect multiple areas of a bank are possible and need to be explored. In the damage to physical assets event type, examples are listed in Basel II (2006) [19] such as human losses from external sources (terrorism and vandalism) and natural disasters. The losses caused by the 9/11 terrorist attack in 2001, hurricane Katrina in 2005, and the earthquake in Japan in 2011 are just some examples pointed out by Horkenko *et al* (2011) [75] that are classified as operational risks. If focused directly on a financial headquarter, these events could severely damage or wipe out multiple business lines. In addition, in the business

December 2008	
Loss Amount:	\$7.5 billion
Firm:	Airfield Greenwich Group
BIS event type:	Theft and fraud (ET2)
BIS business line:	Asset management (BL7)
Airfield Greenwich Group lost as much as \$7.5 billion due to Bernard Madoff's \$50 billion Ponzi scheme. The US investment fund had the largest exposure of the more than 50 banks and hedge funds affected by Madoff's fraud. Other financial institutions with exposures in excess of \$1 billion included Banco Santander, Mass Mutual Financial Group, Access International Advisers, HSBC Bank, Union Bancaire Privée and UniCredit.	

June 2011	
Loss Amount:	\$85 million
Firm:	Wells Fargo
BIS event type:	Internal fraud (ET1)
BIS business line:	Retail banking (BL3)
The US Federal Reserve Board fined Wells Fargo \$85 million for mortgage fraud. Wells Fargo Financial, the firm's former subsidiary, rewarded employees who processed the most mortgages. Employees falsified mortgage documents to help unqualified applicants receive loans and convinced borrowers to take subprime loans even if they qualified for better rates. Wells Fargo closed the subsidiary and agreed to compensate affected borrowers.	

February 2013	
Loss Amount:	\$325 million
Firm:	Royal Bank of Scotland Group
BIS event type:	Clients, Products and Business Practices (ET4)
BIS business line:	Trading and Sales (BL2)
The CFTC fined RBS \$325 million for manipulating London Interbank Offered Rate (LIBOR). Between 2006 and 2010, RBS traders shared confidential information interdepartmentally and influenced interest rates both internally and at other banks. When RBS employees discovered a pending internal investigation, they concealed their activities by making oral requests instead of written ones. At least 21 RBS employees were involved in the scandal.	

Table 4.17: Select large operational risk losses as reproduced from Operational Risk and Regulation (2009, 2011, 2013) [96–98] respectively.



Figure 4.4: Single vs. multiple UoM scenario.

disruption and system failures event type, we see examples of hardware, software, telecommunications and utility outages. Such real life events were highlighted in McConnell and Blacker (2013) [84]. For instance, there were extended system failures encountered by Royal Bank of Scotland in 2012 in which a software update turned out to be corrupted and resulted in disrupted payments and other transactions. In 2010, the Development Bank of Singapore also experienced a massive IT failure which impacted customers and business. Faced with other cyber risk or systemic IT failures, multiple business lines could also be vulnerable to such risks and hence would impact multiple UoMs.

This chapter aims to migrate from a single UoM scenario to a scenario that spans multiple UoMs. This is graphically displayed in Figure 4.4. We develop the end-to-end process that enables the development of enterprise-wide scenarios. This idea of approaching scenario analysis on an aggregate level was formulated in the 2014 Fields-Mprime Industrial Problem Solving Workshop held in Toronto, ON. There, open-ended questions were asked of participants such as:

- How best should scenario workshops be conducted in order to elicit necessary and sufficient information to formulate scenarios for modeling?
- How should questions be framed?
- What are realistic questions to ask on frequency and severity?
- What sort of data should be referenced in order to guide scenario respondents?

- How can scenarios be updated/refreshed over time to ensure their reasonableness?
- Are scenarios a valuable element for AMA modeling?

The direction the group took the problem was that of scenarios spanning multiple UoMs. The method detaches the requirement for bankers to quantify all aspects of a scenario. Disaster planning experts are seen as experts in quantifying frequency estimates of large-scale, low frequency events. Corresponding to each event would be different intensity levels such as low, medium, high. Next, depending on the level of intensity, severity distributions would need to be calibrated to quantify the loss characteristics. This is where the banking experts could add value. Bankers would be called upon to quantify financial impacts to their institution depending on the severity of the event. That is, different severity levels of loss conditioned on a particular catastrophic event would impact a bank differently. In order to account for different types of enterprise-wide scenarios materializing, a Monte Carlo routine was used to cycle through different scenarios where samples were pulled from calibrated distributions and aggregate capital could be produced. Aggregate capital was calculated by compounding frequency and severity to determine a capital requirement derived from the various scenarios.

We build upon this framework and generalize the process. We contrast the workshop results and our contribution in Table 4.18. We build in areas for sophistication where previous assumptions were overly simplistic. We generalize to allow for different scenarios to have unique intensity levels with unique probability of occurrences. Moreover, different scenarios would naturally effect different areas of the bank and should not be limited to the same effected cluster of UoMs. Hence, increased granularity would be required. Whereas in the workshop multiple scenarios were allowed to happen simultaneously (even with small probability), we take a different approach so that different scenarios are treated independently. This changes the conditional

Attribute	Workshop	Contribution
Frequency	<ul style="list-style-type: none"> • Constant intensity • Allow for two scenarios 	<ul style="list-style-type: none"> • Constant intensity • Discuss real disaster study to extract intensity • Develop theory for n type of scenarios • Reinforce with example using two scenarios
Probability of Occurrence	<ul style="list-style-type: none"> • Low, Medium, High intensity levels • Constrained to use same three levels for different scenarios • Claim $P(L > x)$ conditioned on two scenarios plus intersection of both events 	<ul style="list-style-type: none"> • Generalized for m intensity levels • Allow flexibility for different levels depending on scenarios • Reinforce with example using (L,M,H) and (L,H) levels • Altered $P(L > x)$ to ensure only one scenario can happen at a time
Unit of Measure	<ul style="list-style-type: none"> • Same UoMs effected regardless of scenarios • Constant loss distribution across all UoMs 	<ul style="list-style-type: none"> • Unique number of UoMs effected per scenario • Generalize to allow unique loss distribution per UoM
Scenario Integration Application	None	Reapply to convolution method
Capital	<ul style="list-style-type: none"> • Calculate aggregate loss across all UoMs then take VaR • No discussion of capital allocation 	<ul style="list-style-type: none"> • Calculate aggregate loss across all UoMs then take VaR • Introduce and apply capital allocation methodology

Table 4.18: Contrast between workshop and new contribution.

probability and results in a different framework. We also complete the problem by proposing a partition methodology that disseminates capital to individual business lines. To our knowledge, the methodology proposed to quantify enterprise-wide scenarios has not been covered in the literature. In addition, even though the capital allocation discussion is not new, its application in an operational risk context resulting from catastrophic scenarios to our knowledge has not received attention. This collective framework is something of value that could be offered to the financial community in present day to complement operational risk modeling. We also note that once formulated, the scenario loss distribution may be applied in the context of the integration method of the previous chapter. In that sense, this is yet another contribution.

The organization of this chapter is as follows. Section 4.2 covers a literature review of the developed area of scenario quantification on a single cell level. In addition, the underlying motivation for the need of such enterprise-wide scenarios is shown through a discussion of various references. Section 4.3 shows how disaster planning experts may provide input into the frequency quantification of a scenario. A California earthquake study is investigated to show how to extract the relevant parameter estimates. Section 4.4 discusses how to quantify the severities of the loss distribution. Section 4.5 generalizes the simulation algorithm using a Monte Carlo framework. Section 4.6 works through hypothetical catastrophic scenarios using all required parameters for simulation. Section 4.7 links the results to an application of scenario integration using convolution. Section 4.8 shows a different treatment by computing aggregate loss and looks at capital allocation to contributing UoMs. Section 4.9 summarizes the chapter.

4.2 Literature Review

When it comes to modeling and incorporating scenarios into an AMA model, it is pointless if there is no proposed method to translate the story or details of a scenario

into the quantitative model. However, a less structured approach could be taken for what-if scenarios that add value in another way. Chapelle (2014) [39] differentiates between two different types of scenarios: (i) capital scenarios and (ii) management scenarios. Capital scenarios are used as explicit inputs into the calculation of regulatory capital and require precise quantification. These are relevant to our work. However, management scenarios are not integrated into an AMA model. They require management response and close monitoring without necessarily needing additional funding. They typically relate to the absence of expected benefits or loss of upside potential. Examples of management scenarios would be failure of a project or reputational damage due to poor communication.

Since we are concerned with the modeling and methodology of capital scenarios, we will simply refer to capital scenarios as just scenarios and exclude any further discussion of management scenarios. As we have already noticed, the literature surrounding scenario quantification is usually at the UoM level. Dutta and Babbel (2010) [57] define a scenario with frequency in the form of $\frac{m}{t}$ which means that m events are likely to occur in a period of t years with corresponding severity range $[a, b]$. The change of measure scenario integration methodology revises the historical probability distribution so that the number of occurrences of the event in a sample equivalent to t years of losses is equal to m . This method is applicable for a single scenario or a set of scenarios to form an implied severity distribution. The methodology is reinforced via a real-world example in Dutta and Babbel (2010) [57] where for a single UoM, a scaled and anonymized set of 16 scenarios are used to revise a historical probability distribution corresponding to five candidate severity distributions.

Ergashev (2012) [64] defines a scenario to have both a duration and a lower bound. For a particular UoM, a set of scenarios is defined as $S = (S_1, \dots, S_k)$ where $S_i = (M_i, L_i)$, M_i is a natural number indicating that scenario S_i is a once-in-a- M_i -year event, $L_i > 0$ is the lower bound of the loss associated with the scenario and k is the total number

of scenarios for a UoM. Hence the scenario integration methodology relies on filtering the scenario set to arrive at the select few that shift the risk profile of the base model's severity distribution for a particular UoM.

Also, the use of Bayesian methods as detailed in Shevchenko and Wüthrich (2006) [106], Lambrigger *et al* (2007) [82], and Cope (2010) [45] operate on the condition that the parameters of the prior distribution can be estimated subjectively using expert opinion and used to refine/update the posterior distribution. Since parameters of probability distributions are unique for each UoM, the scenario integration methodology is again on a very granular level.

In the context of AMA modeling for regulatory capital quantification, there is only high-level discussions of catastrophic scenarios with no formulaic approach to AMA modeling. As discussed in McDermott *et al* (2012) [85], when an interview panel was asked if the financial community would, “see any kind of convergence either in the number and type of scenarios used or in the way they are being used, as well as the goals that institutions are trying to achieve by using them”, McDermott states that because of the diversity of natural catastrophes in the US, there does not appear to be any convergence in the near future. Also, “it is still a fairly young discipline”. The panel as a whole agreed that it might be natural to think of convergence at some point in the future, but practitioners are just in the experimenting phase. We see in Tozer-Pennington (2011) [109] the terminology of a multivariate scenario being introduced. It was reasoned that assessing simple single variable scenarios by themselves does not give insight to what may unfold in a multifaceted loss event. It was pointed out that the 2011 devastating earthquake in Japan, ensuing tsunami and potential nuclear crisis had been termed a one-in-a-thousand year event. The crisis started off with a 9.0 earthquake off the Pacific coast of Tōhoku, which was followed by a tsunami off the northeast coast of Japan, and then resulted in the Fukushima Daiichi nuclear disaster resulting in a nuclear meltdown of three of six nuclear reactors. Operational

risk managers are now becoming increasingly aware that such devastations are becoming more common. It was also suggested in Tozer-Pennington (2011) [109] that firms need to be better prepared to quantify impacts of extreme scenarios and perhaps the risk management community as a whole could collaborate in its development.

Branching out to see if the concept of multivariate scenarios are being used in the insurance field, we see only limited discussion of using scenarios to aid in operational risk capital quantification. The Solvency II Directive (2009) [66], referred to as “Basel” for insurers, is somewhat similar to banking regulations of Basel II (2006) [19]. Like Basel regulations, Solvency II has undergone revisions to take into account current developments in insurance, risk management, finance techniques, international financial reporting and prudential standards. The revision has come into effect January 2016. That being said, we see that within the insurance field there is talk about advancing operational risk capital quantification using scenarios. Although the framework is not the same as the AMA modeling, scenario analysis does provide the same forward-looking risk assessment relevant for the insurance industry. In Meek (2012) [88], we learn that at the Swiss general insurance group Zurich, the insurance industry is also faced with the scarcity of loss data which implies that a curve-fitting methodology used to quantify other risk types is not feasible for operational risk. Although the capital rules may be different, banks/insurance companies usually run internal models for risk management reasons to inform internal views of capital adequacy. These may be similar or different to regulatory capital models. At Zurich, the scenario analysis method runs through a set of 50 scenarios. The top-down scenario analysis program runs the scenarios across the company and initiates discussions with a business head. The brainstorming sessions, for instance, may look at specific fraud events or fines, which yield a few data points. The data points are then passed onto the analytic group which is tasked to convert data points via an undisclosed methodology into parameters for quantification. The end goal is that the “mathematical curve” that is used to quantify the risk may not align with the risk profile of the scenarios. Hence

Zurich may then adopt any other type of loss curve that better fits the risk profile.

The reason this example was shown was two fold: 1) to show that the contribution to the development of a enterprise-wide scenario framework and methodology will be new and 2) banking and insurance are concerned with the same problem for operational risk capital quantification. Moreover, as our method calls for disaster planning experts to input their view into scenario design and these experts are naturally associated with the insurance and actuarial fields, the cross-over is natural.

4.3 Frequency Quantification

We concern ourselves with catastrophic scenarios that seldom occur, however when they do, the severity level is high. Natural Resources Canada monitors natural hazards in Canada and provides information about hazardous events, as well as provide information to help Canadians understand, prepare and reduce losses due to hazardous events. Examples of hazards that are reported include earthquakes, space weather, marine geological hazards, floods and landslides. In the United States, the United States Geological Survey (USGS) is a scientific agency of the US government that also monitors natural hazards. The six science programs covered by USGS are coastal and marine geology, earthquake hazards, geomagnetism, global seismographic network, landslide hazards, and volcano hazards. Using the US as an example, one such notable publication was a study done by Field *et al* (2009) [68] investigating earthquake forecasts for the state of California called the Uniform California Earthquake Rupture Forecast (UCERF). A collection of scientists and engineers built a series of four models, one of which was a probability model, that gave a probability of occurrence for each earthquake during a specified (future) time interval. The authors assume time-independence through a Poisson model in the sense that the probability of each earthquake rupture was completely independent of the timing of all others. This was in contrast to other potential modeling choices which assume

time-dependence. Under a different framework, the time-dependent earthquake rupture forecasts condition event probabilities on the date of the last rupture. Such models were motivated by elastic rebound theory of the earthquake cycle and were based on stress-renewal models in which probabilities drop immediately after a large earthquake. The release of tectonic stress reduces the probability, whereas the probability of an earthquake rises as stress re-accumulates due to constant tectonic loading of the fault line. However, luckily for us, the time-independence assumption aligns well with our AMA modeling framework and will serve us in motivating our application.

Results of probability calculations are taken from Field *et al* (2009) [68] and summarized in Table 4.19. Interpreting the first entry, we see that an earthquake with

Magnitude (M)	Frequency	Probability of occurrence
$M \geq 6.7$	1 in 30 year	99.7%
$M \geq 7.0$	1 in 30 year	94.0%
$M \geq 7.5$	1 in 30 year	46.0%
$M \geq 8.0$	1 in 30 year	4.5%

Table 4.19: Summary frequencies and probabilities of earthquake predictions.

a magnitude greater than or equal to 6.7 on the Richter scale is virtually assured in California during the next 30 years (with a 99.7% probability of occurrence). In order to extract the relevant information needed to run an LDA model, we recall the basic properties of the Poisson probability distribution. We encountered the Poisson distribution in Section 2.4.4 used for frequency modeling. We turn to the homogeneous Poisson process which is characterized by an intensity parameter λ such that the number of events in time interval $(t, t + \tau]$ follows a Poisson distribution with corresponding parameter λt . The expression is given by

$$P[N(t + \tau) - N(t) = k] = \frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!} \quad k = 0, 1, \dots \quad (4.1)$$

Magnitude (M)	Poisson parameter λ
$M \geq 6.7$	0.12
$M \geq 7.0$	0.094
$M \geq 7.5$	0.021
$M \geq 8.0$	0.0015

Table 4.20: Summary Poisson parameters for corresponding earthquake magnitudes.

where $N(t + \tau) - N(t) = k$ represents the number of events in the time interval $(t, t + \tau]$. Hence in our case, λ is the long-term rate of the earthquake rupture, t can be set to zero, and $\tau = 30$ represents the forecast duration. In the following example, we show how to recover the necessary intensity value.

Example 4.3.1. *Using (4.1), we find the associated λ from the first row in Table 4.19. Since the entries provide us with information of an earthquake in the next 30 years with a probability of occurrence, we can solve the complement*

$$\begin{aligned}
P[N(t + 30) - N(t) \geq 1] &= P[N(t + 30) - N(t) = 1] + P[N(t + 30) - N(t) = 2] + \dots \\
&= 1 - P[N(t + 30) - N(t) < 1] \\
&= 1 - P[N(t + 30) - N(t) = 0] \\
&= 1 - \frac{e^{-30\lambda}(30\lambda)^0}{0!} \\
&= 1 - e^{-30\lambda}
\end{aligned}$$

$\implies e^{-30\lambda} = 0.03$ or $\lambda = 0.12$. Hence we can say that we expect 0.12 earthquakes of $M \geq 6.7$ in a given year.

We solve the same expression and summarize the corresponding intensities parameters in Table 4.20. To summarize, this frequency estimate represents the first necessary piece of information required for the scenario analysis program. We have tasked the disaster planning experts to use their expertise to provide best estimates to pass along to the banking professionals to quantify a financial impact due to the catastrophic

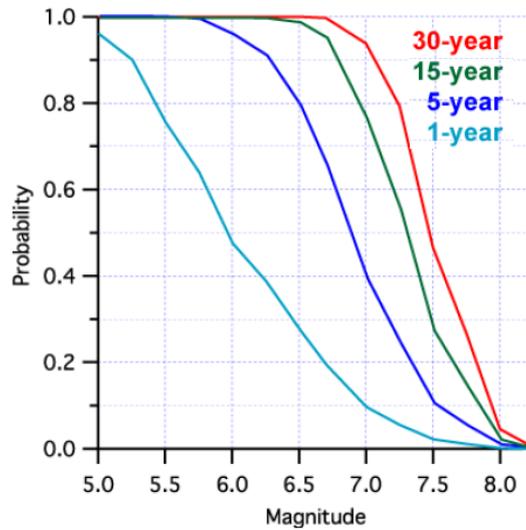


Figure 4.5: Cumulative probabilities for occurrence for one earthquake ranging over various magnitudes.

scenario.

While this is one method to collect the frequency estimate, our methodology is a slight modification of the above method. In Table 4.20 we had a different λ for different magnitude levels. What was reported were cumulative probabilities. That is, the arrival of a single earthquake during a 30 year period while cycling through various magnitudes. Figure 4.5 takes a ‘traditional’ CDF plot and has increasing probabilities from right to left where lower magnitude earthquakes are more likely. Moreover, there is a hidden third dimension which is ‘typical’ of a CDF plot where for a Poisson distribution, the possibility of k events happening are taken into account. However, we propose to fix a λ and cycle through different severity options. That is, there are distinct type of events that may occur that are mutually exclusive of one another.

We can achieve distinct type of events by simple subtraction. While $M \geq 8.0$ with probability 4.5 % is one type of earthquake, we may classify a different type of earthquake with magnitude $7.5 \leq M \leq 8.0$ with probability $46\% - 4.5\% = 41.5\%$. Hence

Magnitude (M)	Frequency	Probability of occurrence
$M \leq 6.7$	1 in 30 year	0.3%
$6.7 \leq M \leq 7.0$	1 in 30 year	5.7%
$7.0 \leq M \leq 7.5$	1 in 30 year	48.0%
$7.5 \leq M \leq 8.0$	1 in 30 year	41.5%
$M \geq 8.0$	1 in 30 year	4.5%

Table 4.21: Summary by magnitude ranges.

we modify Table 4.19 and produce distinct earthquake events in Table 4.21. We interpret earthquakes with magnitude less than 6.7 to be a catch-all segment.

We formalize our mathematical algorithm in three boxes (Frequency, Severity and Monte Carlo framework) to follow. The generalization of this algorithm calls for different disaster planning experts to quantify different scenarios (e.g. earthquake, flood, hurricane) and hence allow for n possible scenario types.

Frequency framework: Let E_i represent a catastrophic scenario event where there are $i = 1, \dots, n$ different types of events. Let the corresponding frequency estimate be represented by λ_i . For E_i , let there be j corresponding severity levels where $j = 1, \dots, m$. We append the index on E_i and write $E_{i,j}$ to represent the i^{th} event with severity level j . For instance if $j = 2$, we would interpret a catastrophic event occurring with either a low or high severity impact. We then require given that event E_i has occurred, the conditional probability of a specific severity level is given as $\mathbb{P}(E_{i,j}|E_i) = p_{i,j}$ where $\sum_{j=1}^m p_{i,j} = 1$.

The flexibility of this framework will be made clear later by virtue of an example.

4.4 Severity Quantification

Having a frequency estimate in-hand for a particular scenario event with a specific severity level, a banking professional would then be better suited to determine a corresponding financial impact. It would still be up to the scenario design team to research and translate the type of scenario and severity into a storyline. Using the earthquake as an example with two severity levels (low and high), a remote low magnitude earthquake may lead to a mild financial set-back effecting bank branches with downed power lines and thus cause business disruption. However, a centralized high magnitude earthquake focused in a financial headquarters may have multiple impacts: (i) damage to physical assets for the office building and workstations, (ii) business disruption for the capital markets, corporate and retail business, (iii) potential external fraud in the form of opportunistic theft, (iv) workplace safety losses in the form of general liability (slip and fall accidents).

Much like in equations (3.11) and (3.13), we can proceed by asking participants two opinions regarding the median loss (50%) and third quartile (75%) of a candidate severity distribution. In that example we chose, a-priori, a Lognormal distribution. Although we need not necessarily limit ourselves to that distribution, we may select from a host of two, three, or four parameter distributions. We rationalize that we may limit our investigation to a select number of two-parameter distribution choices. As an aside, it was stated in a publication from the BCBS, Operational risk - Revisions to the simpler approaches (2014) [28], that the Basel working group has formulated an Operational Risk Capital-at-Risk (OpCaR) calculator. The calculator is an in-house model aimed to benchmark operational risk proxy indicators (like gross income) and to calibrate coefficients for standardized approaches. The basis of the model is an AMA model using the LDA approach using four candidate distributions used for severity fitting. Two parameter Pareto, Lognormal, Loglogistic and Log-Gamma are used. Hence we believe that since the BCBS is using two parameter probability dis-

tributions for calibration and benchmarking, then by extension we may constrain ourselves in the same manner and motivate our technique accordingly. Moreover, we elected to use a Lognormal distribution based on the range of practice observed at scenario-based AMA banks as detailed in the publication from the BCBS, Operational Risk - Supervisory Guidelines for the Advanced Measurement Approaches (2014) [22]. There it was stated that banks used only a Lognormal curve to fit scenario data regardless of the business, size and complexity of an UoM. Hence the use of a single curve across all UoMs indicated that the only driver of variation in the operational risk exposure lied in the scenario-induced parameter estimates corresponding to different Lognormal distributions.

Our goal here is to calibrate a severity distribution based on scenario responses stemming from a single catastrophic event. The severity loss (or simply loss) is unique for a single UoM that is part of the bigger picture. While we still are concerned with a single scenario impacting multiple business lines and multiple event types, our quantification starts at the individual cell level. Alternatively, it is also possible to think in terms of one single aggregate loss and disseminate capital ex-post, however, this simplification can be seen as a corollary to the more granular method.

Severity framework: Let $L_{i,j,k}$ represent the loss corresponding to the i^{th} scenario event with severity level j coming from UoM k . Recall we allow for $i = 1, \dots, n$ different types of scenarios, $j = 1, \dots, m$ different severity levels and $k = 1, \dots, 56$ possible UoMs. If $k = 56$, this means that the scenario effects every UoM in the bank. Usually we have $k < 56$ indicating that the effect is localized to a portion of the bank. Each loss is then distributed according to some probability distribution with parameters $\tilde{\theta} = (\theta_1, \theta_2, \dots, \theta_h)$ where h represents the number of parameters needed to be calibrated. The associated quantile function for percents \tilde{p} is solved where

$$Q_{\tilde{\theta}}(\tilde{p}) = \inf\{\tilde{x} \in \mathbb{R} : \tilde{p} \leq F(\tilde{x})\}. \quad (4.2)$$

Since the dimension $\dim \tilde{p} = \dim \tilde{\theta} = h$, we have a system of h equations with h unknowns to solve.

Thus once $\tilde{\theta}$ has been obtained, we are able to apply our Monte Carlo simulation to find our aggregate loss impact.

4.5 Monte Carlo Simulation

Recall that the end goal is to compute the operational risk capital requirement which is a VaR measure at the 99.9 percentile. Mathematically, we need to find the specific quantile of the annual loss distribution. That is, our goal is to find the value x such that

$$\mathbb{P}(L > x) = 1 - 0.999. \quad (4.3)$$

Since this a VaR measure purely attributed to the use of scenarios, an approximation can be made when computing the VaR attributed to the occurrence of each catastrophic event. Mathematically, conditioned on the possibility of multiple scenarios

occurring we have

$$\begin{aligned}
\mathbb{P}(L > x) &= \mathbb{P}(L > x | E_1 E_2 \dots E_n) \mathbb{P}(E_1 E_2 \dots E_n) + \\
&\mathbb{P}(L > x | E_1^c E_2 \dots E_n) \mathbb{P}(E_1^c E_2 \dots E_n) + \\
&\mathbb{P}(L > x | E_1 E_2^c \dots E_n) \mathbb{P}(E_1 E_2^c \dots E_n) + \dots + \\
&\mathbb{P}(L > x | E_1 E_2 \dots E_n^c) \mathbb{P}(E_1 E_2 \dots E_n^c) + \dots + \\
&\mathbb{P}(L > x | E_1^c E_2^c \dots E_n) \mathbb{P}(E_1^c E_2^c \dots E_n) + \dots + \\
&\mathbb{P}(L > x | E_1^c E_2^c \dots E_n^c) \mathbb{P}(E_1^c E_2^c \dots E_n^c).
\end{aligned} \tag{4.4}$$

From the law of total probability, we have the following

$$\begin{aligned}
\mathbb{P}(L > x) &\approx \mathbb{P}(L > x | E_1) \mathbb{P}(E_1) + \\
&\mathbb{P}(L > x | E_2) \mathbb{P}(E_2) + \dots + \\
&\mathbb{P}(L > x | E_n) \mathbb{P}(E_n).
\end{aligned} \tag{4.5}$$

Hence (4.5) states that $\mathbb{P}(L > x)$ is equal to a weighted average of $\mathbb{P}(L > x | E_i)$, each term being weighted by the probability of the event on which it is conditioned (see Prop. 3.1, Ross (2002) [103]).

Having the problem phrased this way is conducive to solving using Monte Carlo methods. This can be seen easily in the case of a single event ($E_i \equiv E$) where we are left to solve

$$\mathbb{P}(L > x) = \mathbb{P}(L > x | E) \mathbb{P}(E). \tag{4.6}$$

For example, in the case where $\mathbb{P}(E) = 0.01$, we know that we must have the VaR condition of $\mathbb{P}(L > x) = 0.001$. Hence we simply need to find the value x that makes $\mathbb{P}(L > x | E) = 0.1$

$$\underbrace{\mathbb{P}(L > x)}_{0.001} = \underbrace{\mathbb{P}(L > x | E)}_{\text{find } x \text{ that makes this} = 0.1} \underbrace{\mathbb{P}(E)}_{0.01}. \tag{4.7}$$

To make this concrete, we summarize the Monte Carlo framework that generalizes the routine for any number of scenarios. We then provide a two scenario example that shows how the method may be used in practice.

Monte Carlo framework: Collect frequency estimates λ_i corresponding to catastrophic scenario events E_i , $i = 1, \dots, n$. Corresponding to each unique scenario are j severity levels where $j = 1, \dots, m$. Then mapped to each scenario is a calibrated loss $L_{i,j,k}$ which is representative of a loss that would be realized for UoM $k = 1, \dots, 56$. To perform the Monte Carlo loop, initialize a high number of simulations. Determine which scenario, if any, are realized. Do this by generating uniformly distributed random numbers from the interval $[0,1]$. Check, element by element, if each realization is less than λ_i . Note that each λ_i discretizes the interval $[0,1]$ into segments such that depending on the realization from the random draw, only one scenario may occur. Going down the decision tree, next determine which severity intensity level occurs. Each severity intensity has an associated probability of occurrence such that all probabilities sum to one. To determine which intensity occurs, a similar random draw scheme used previously to determine the type of scenario may be applied again. Finally, produce a realization from the unique loss distribution across all effected cells. Sum across all cells and store. After running all simulations, take VaR at the 99.9 percentile (or in this case corresponding to the conditional quantile on non-zero entries that is equivalent to the 99.9 percentile of the full loss distribution). This VaR estimate represents the aggregate loss corresponding to the possibility of n number of scenarios possibly occurring.

4.6 Example: Two Catastrophic Scenarios

To show the mechanics of the method, we provide a hypothetical example and run a full simulation. We provide all required parameters to run the algorithm. The param-

eters used are fictitious because depending on the footprint of a bank geographically and the type of losses that may occur which are only known to its business line experts, the loss experience is very unique.

From disaster planning experts

We start with a hypothetical bank that has offices in both New York and California. We propose two catastrophic scenarios events: a flood affecting New York and an earthquake affecting California. The first choice of scenario was motivated from the 2012 Atlantic hurricane season that resulted in hurricane Sandy and caused flooding of streets and subway lines and cut power around the city. The second scenario was motivated by the previous UCERF study for the state of California and the recurrent number of earthquakes characteristic of the state.

We define the two catastrophic events

- Event 1: New York flood, represented as E_1 ,
- Event 2: California earthquake, represented as E_2 .

Corresponding to each event are different severity levels. For New York, we have

- Level 1: Low - represented as $E_{1,1}$,
- Level 2: Medium - represented as $E_{1,2}$,
- Level 3: High - represented as $E_{1,3}$.

For California, we have

- Level 1: Low - represented as $E_{2,1}$,
- Level 2: High - represented as $E_{2,2}$.

Hence we are left with the mutually exclusive set of scenarios $\{E_{1,1}, E_{1,2}, E_{1,3}, E_{2,1}, E_{2,2}\}$ as defined by the disaster planning experts. We note that we are not limited to having the same severities levels across different scenarios. We envision more choices for severity levels in New York as given by the experts in that a greater density of finan-

cial offices are located there and hence could attract different types of losses. Next, we require frequency estimates from the experts. To keep inline with our reasoning for career-spanning estimates for frequency, we assign the New York flood to be a 1 in 20 year event and the California earthquake to be a 1 in 30 year event. For each catastrophic event, each severity level occurs with different probabilities. Hence we have

$$\mathbb{P}(E_{1,j}|E_1) = p_{1,j} \quad j = 1, 2, 3 \quad (4.8)$$

$$\mathbb{P}(E_{2,j}|E_2) = p_{2,j} \quad j = 1, 2. \quad (4.9)$$

Again, the probability of each severity level occurring must sum to 1. The information gathered from the experts is summarized in Table 4.22 and Table 4.23.

Frequency	$\lambda = 0.05$		
Severity Level	Low	Medium	High
Probability ($p_{1,j}$)	0.5	0.4	0.1

Table 4.22: New York flood parameters.

Frequency	$\lambda = 0.0333$	
Severity Level	Low	High
Probability ($p_{2,j}$)	0.6	0.4

Table 4.23: California earthquake parameters.

From banking experts

Passing along this information to the banking experts, it is reasoned that different events may occur with different probabilities of severity levels and experts within the institution are in the best position to assign a financial impact to the bank in their level of expertise and then aggregated to the bank as a whole. For example, if an earthquake hit California, that may result in a shut down to a particular part of a bank. Depending on the damage to physical assets or business disruptions causing lost revenues, a banker may be able to quantify a reasonable estimate of financial loss

due to the operational risk event. Hence, in our example, we have different UoMs effected in each scenario

$$UoM_{E_{1,j,k}} \quad j = 1, 2, 3, \quad k = 1, 2, \dots, 10 \quad (4.10)$$

$$UoM_{E_{2,j,k}} \quad j = 1, 2 \quad k = 1, 2 \dots, 5. \quad (4.11)$$

Not only do we allow for 10 UoMs in the first scenario and 5 UoMs in the second scenario, we also allow for unique loss distributions to be calibrated for each UoM under each scenario type and severity level. Using the same line of reasoning from the previous chapter, we may take a view that scenarios follow a Lognormal distribution and the minimum number of points need to calibrate such as curve are two points. Moreover, rather than elicit answers regarding the moments, we seek responses in the form of quantiles: namely median (50%) loss and third quartile (75%) loss. We provide hypothetical values for each case in order to show the mechanics of the method. The combined information received from disaster planning experts and banking experts for a single UoM are given Table 4.24 and Table 4.25.

Frequency	$\lambda = 0.05$		
Severity Level	Low	Medium	High
Probability ($p_{1,j}$)	0.5	0.4	0.1
Median	\$100,000	\$500,000	\$1,000,000
75 th percentile	\$300,000	\$1,000,000	\$2,000,000

Table 4.24: New York flood full information for single UoM.

Frequency	$\lambda = 0.0333$	
Severity Level	Low	High
Probability ($p_{2,j}$)	0.6	0.4
Median	\$50,000	\$300,000
75 th percentile	\$150,000	\$700,000

Table 4.25: California earthquake full information for single UoM.

We reiterate that we generalize this process to allow different UoMs being impacted and thus having individual loss profiles. For example, if a flood was to occur, we

could reasonably expect that two event types would occur: ET5 (damage to physical assets) and ET7 (execution, delivery and process management). That is, losses arising from the flood directly causing damage to infrastructure and then losses from failed transactions processing respectively. For arguments sake, we can then say that these ETs occur in the first five business lines: BL1 (corporate finance), BL2 (trading and sales), BL3 (retail banking), BL4 (commercial banking), BL5 (payment and settlement). Hence, this forms the 10 UoM structure (2ET*5BL). Then, to make this example more realistic, we require an individualized loss assessment in each UoM from a business line expert. To this end, we take each 50th and 75th percentile loss and add white Gaussian noise with mean zero and variance 10%. Once the noise is added, we round each value to the ten thousandth as we would expect business line experts not to provide estimates to the dollar level. This information is summarized in Table 4.26 and Table 4.27. This mimics the effect of an individualized loss assessment. Finally, once we have loss amounts for each UoM, we fit individualized Lognormal distributions using (3.11) and (3.13). To be explicit, from Table 4.26, the first entry median loss of \$100,000 together with 75th percentile estimate of \$300,000 produces the Lognormal distribution with parameters $\mu = 11.51$ and $\sigma = 1.63$ to coincide with the first entry in Table 4.28. The calibrated probability distributions corresponding to the loss estimates are summarized in Table 4.28 and Table 4.29.

Frequency	$\lambda = 0.05$					
Severity Level	Low		Medium		High	
Probability ($p_{1,j}$)	0.5		0.4		0.1	
Quantile	Median	75 th percentile	Median	75 th percentile	Median	75 th percentile
UoM1	\$100,000	\$300,000	\$500,000	\$1,000,000	\$1,000,000	\$2,000,000
UoM2	\$120,000	\$170,000	\$590,000	\$1,230,000	\$1,170,000	\$2,450,000
UoM3	\$160,000	\$590,000	\$790,000	\$940,000	\$1,580,000	\$1,870,000
UoM4	\$130,000	\$290,000	\$140,000	\$960,000	\$290,000	\$1,920,000
UoM5	\$110,000	\$370,000	\$640,000	\$1,470,000	\$1,270,000	\$2,940,000
UoM6	\$60,000	\$280,000	\$550,000	\$1,450,000	\$1,100,000	\$2,890,000
UoM7	\$90,000	\$290,000	\$290,000	\$1,450,000	\$590,000	\$2,900,000
UoM8	\$110,000	\$440,000	\$430,000	\$1,210,000	\$860,000	\$2,420,000
UoM9	\$210,000	\$430,000	\$550,000	\$620,000	\$1,110,000	\$1,240,000
UoM10	\$190,000	\$430,000	\$1,070,000	\$1,230,000	\$2,130,000	\$2,450,000

Table 4.26: New York flood UoM loss estimates.

Frequency	$\lambda = 0.0333$			
Severity Level	Low		High	
Probability ($p_{2,j}$)	0.6		0.4	
Quantile	Median	75 th percentile	Median	75 th percentile
UoM1	\$50,000	\$150,000	\$300,000	\$700,000
UoM2	\$60,000	\$170,000	\$350,000	\$780,000
UoM3	\$80,000	\$320,000	\$470,000	\$1,490,000
UoM4	\$60,000	\$90,000	\$380,000	\$400,000
UoM5	\$60,000	\$290,000	\$330,000	\$1,370,000

Table 4.27: California earthquake UoM loss estimates.

Frequency	$\lambda = 0.05$					
Severity Level	Low		Medium		High	
Probability ($p_{1,j}$)	0.5		0.4		0.1	
Parameter	μ	σ	μ	σ	μ	σ
UoM1	11.51	1.63	13.12	1.03	13.82	1.03
UoM2	11.70	0.52	13.29	1.09	13.97	1.10
UoM3	11.98	1.93	13.58	0.26	14.27	0.25
UoM4	11.78	1.19	11.85	2.85	12.58	2.80
UoM5	11.61	1.80	13.37	1.23	14.05	1.24
UoM6	11.00	2.28	13.22	1.44	13.91	1.43
UoM7	11.41	1.73	12.58	2.39	13.29	2.36
UoM8	11.61	2.06	12.97	1.53	13.66	1.53
UoM9	12.25	1.06	13.22	0.18	13.92	0.16
UoM10	12.15	1.21	13.88	0.21	14.57	0.21

Table 4.28: New York flood UoM parameter estimates.

At this point, we have all the information we need to run the Monte Carlo simulation. The method by which we achieve this listed below.

1. Each scenario (flood or earthquake) once it occurs, will impart a unique loss in each effected UoM.
2. The Monte Carlo loop is performed multiple times creating a unique distribution of losses for each UoM.
3. A VaR at the 99.9 percentile is taken from the corresponding loss distribution for each individual UoM. Again, the 99.9 quantile needs to be adjusted to be taken over only those realizations that were non-zero. The VaR representing a high quantile aggregate loss due to the scenario is obtained from the aggregate

Frequency	$\lambda = 0.0333$			
Severity Level	Low		High	
Probability ($p_{2,j}$)	0.6		0.4	
Parameter	μ	σ	μ	σ
UoM1	10.82	1.63	12.61	1.26
UoM2	11.00	1.54	12.77	1.19
UoM3	11.29	2.06	13.06	1.71
UoM4	11.00	0.60	12.85	0.080
UoM5	11.00	2.34	12.71	2.11

Table 4.29: California earthquake UoM parameter estimates.

loss distribution corresponding to the multivariate probability density (obtained from convolution in the independent case).

n =100,000	
Scenario 1: New York flood	VaR 99.9
UoM1	\$5,829,676
UoM2	\$7,551,260
UoM3	\$2,522,637
UoM4	\$51,138,068
UoM5	\$10,862,853
UoM6	\$14,184,212
UoM7	\$50,407,975
UoM8	\$13,094,804
UoM9	\$1,484,313
UoM10	\$3,063,496

(a) New York UoM VaR

n =100,000	
Scenario 2: California earthquake	VaR 99.9
UoM1	\$2,181,451
UoM2	\$2,203,969
UoM3	\$7,357,474
UoM4	\$429,267
UoM5	\$8,775,085

(b) California UoM VaR

Table 4.30: Individual UoM VaR.

Having run our Monte Carlo simulation, we produce the results in Table 4.30. During our simulation, we had independent random draws from each loss distribution across each affected UoM. In order to compute the aggregate VaR, we took the VaR of the corresponding aggregate loss distribution. That is, after simulation we obtained a unique loss distribution for each UoM that may be associated with a PDF $f_i(x)$. Since the UoMs were assumed to be independent, the PDF of the aggregate loss was given by the convolution of the impacted PDFs and stated in Osgood (2007) [99] as

$$f_{agg}(x) = f_1 * f_2 * \dots * f_k. \quad (4.12)$$

The VaR was then found by inverting the CDF at the associated quantile q and given as $F_{agg}^{-1}(q)$. In that sense, if a flood was to occur in New York, the aggregate loss to the effected parts of the bank would be \$121,489,665. If an earthquake was to occur in California, the aggregate loss to the effected parts of the bank would be \$17,198,999. We pause to highlight an important point. It is widely beneficial of having an estimated numeric value due to the possibility of catastrophic scenarios. We explore the concept of risk transfer and for the potential of offsetting risk using insurance. Pointing to Basel II (2006) [19], we refer to paragraph 677 under the section of risk mitigation for operational risk which says that a bank will be allowed to recognize the risk mitigating impact of insurance in the measurement of operational risk used for regulatory minimum capital requirements. The recognition of which would be limited to 20% of the total operational risk capital charge under AMA.

By narrowing down the potential list of disasters and also by focusing on the quantified impact, an appropriate insurance policy could be purchased to realize a capital savings. Importantly, there is a transparency in the modeling that allows for senior management and independent regulators the ability to judge adequacy and appropriateness of insurance coverage. From Ames *et al* (2015) [3] we learn that no US bank has successfully claimed any credit for regulatory capital relief due to insurance. However by contrast, several European banks have successfully achieved capital relief from insurance, but not the full 20%. Thus providing a mechanism as we have done could change this behaviour.

4.7 Scenario Integration Application

Before we continue onto the discussion of deciding what to do with the aggregate capital value, we point that this catastrophic scenario could be integrated into the AMA model by one of the many scenario integration methods as a point estimate or probability distribution. That is, for a particular UoM, the catastrophic scenario

could be just one of the many scenarios developed for the cell. The caveat would be that this scenario is not stand-alone in that other UoMs would be effected simultaneously resulting from the same catastrophic loss.

For example, consider a two UoM banking universe where for UoM_1 , we have five scenarios developed specifically for the UoM and a sixth scenario which was a catastrophic scenario that was developed in conjunction with UoM_2 . For UoM_2 , we consider two stand-alone scenarios and a third scenario that was a catastrophic scenario developed in reference to UoM_1 . While we could take the sixth scenario in the first case, combine it with the third scenario in the second case and look at the aggregate risk, we could simply leave them subject to scenario integration techniques on a cell-by-cell basis. We refer the reader to Dutta and Babbel (2013) [57] where an example is shown in the change of measure context where for a particular UoM, 16 scenarios were developed. The change of measure approach is also amendable to incorporate multiple scenarios for a single UoM.

Referring back to our convolution technique from the previous chapter, we operated under the context of having a unique loss distribution characterizing each scenario. Hence, looking at this enterprise-scenario framework, we may include a catastrophic scenario on a cell level by again performing the convolution. Thus we believe the unique application of the convolution scenario integration technique in the enterprise-wide scenario framework is an added extension that has not been covered before.

4.8 Capital Allocation

This last section views the catastrophic scenario as a stand-alone element subject to insurance offset or other risk mitigating strategies. Then it would be of interest to see the effect on the enterprise and also each contributing UoM. Hence as a final step to the formalization, quantification and solution to this problem, we investigate

the effects of capital allocation. Capital allocation could be used as a mechanism to incentivize better risk management or if not possible, at least highlight vulnerabilities/weaknesses in an organization.

In this section, we adopt the convention that we have a collection of risks $X = X_1, X_2, \dots, X_n$ where each X_i is an independent random variable that is distributed according to a continuous probability distribution and hence uniquely defines the loss profile for each i^{th} UoM. Whereas before in (2.10) X_i was an individual severity, for ease when talking about capital allocation and associated risk measures, simply referring to aggregate X instead of S is more common. Now having adopted a risk measure $\rho[\cdot]$ and having this collection of risks $X = X_1, X_2, \dots, X_n$, we may determine risk capital for each UoM as $\rho_i = \rho[X_i]$. Since we are concerned with aggregate risk for the enterprise, we may combine them to determine total capital as $\rho[X_1 + X_2 + \dots + X_n] \leq \rho[X_1] + \rho[X_2] + \dots + \rho[X_n]$, then we have subadditivity. When aggregate capital is calculated using $\rho[\cdot]$, the natural question arises as to how much risk does the i^{th} UoM contributes to the total capital and hence how much should be allocated to it?

In general, allocation of capital into risk cells is done in such a way that

$$\rho[X] = \sum_{i=1}^n \Pi_i, \quad (4.13)$$

where Π_i denotes the capital allocated to the i^{th} UoM. To formalize this, we define the principle below and introduce a popular method as listed in Cruz *et al* (2015) [47].

Definition 4.8.1. (*Allocation principle*) *An allocation principle is a mapping of a collection of risks X_i , $i = 1, 2, \dots, n$ into unique allocations $\Pi_i = \Pi_i[X_1, X_2, \dots, X_n]$, $i = 1, 2, \dots, n$ such that $\rho[X] = \sum_{i=1}^n \Pi_i$.*

A popular way to allocate capital is based on marginal risk contribution and given as

$$\rho_i^{marg} = \rho[X] - \rho[X - X_i], \quad (4.14)$$

which is the aggregate risk minus the aggregate risk without cell i . According to Tasche (2008) [108]

$$\sum_{i=1}^n \rho_i^{marg} \leq \rho[X], \quad (4.15)$$

and hence to ensure that allocated capital sums to $\rho[X]$, we may have

$$\Pi_i^{marg} = \frac{\rho_i^{marg}}{\sum_{j=1}^J \rho_j^{marg}} \rho[X]. \quad (4.16)$$

Thus the allocation principle will be satisfied as $\rho[X] = \Pi_1^{marg} + \Pi_2^{marg} + \dots + \Pi_n^{marg}$. We will explore various capital allocation principles in greater detail in the next chapter. Hence for now, applying the marginal risk contribution method to our example we obtain the following allocation in Table 4.31.

n =100,000	
Scenario 1: New York flood	Allocated capital
UoM1	\$2,401,784
UoM2	\$2,009,652
UoM3	\$1,667,319
UoM4	\$5,2287,804
UoM5	\$1,478,391
UoM6	\$2,891,061
UoM7	\$5,2141,388
UoM8	\$4,078,614
UoM9	\$963,932
UoM10	\$1,569,720
Total	\$121,489,665

(a) New York allocation

n =100,000	
Scenario 2: California earthquake	Allocated capital
UoM1	\$889,057
UoM2	\$1,006,604
UoM3	\$7,028,155
UoM4	\$274,154
UoM5	\$8,001,029
Total	\$17,198,999

(b) California allocation

Table 4.31: Capital allocated per UoM.

4.9 Summary

In this chapter we explored the idea of catastrophic scenarios effecting the multiple business lines simultaneously. The idea of an enterprise-wide scenario was a departure from the typical scenario formulation that has traditionally been focused on a single cell level. In order to work with such scenarios, we segregated the frequency estimation of the event to disaster planning experts that normally would be external to the bank. We explored a specialized case of an empirical California earthquake study that fit into the AMA framework and leveraged the work of scientists to better predict the occurrence of an earthquake. Next, we relied on banking experts to quantify the loss conditional on the intensity of scenario and the specific business line for which they work in. We generalized the framework and worked through a hypothetical example. A connection was made with the previous chapter whereby the scenario could also be subject to convolution integration on a granular level. We concluded by looking at the aggregate risk and determined one method to allocate capital to the effected UoMs.

5 Optimal Capital Allocation to Sub-portfolios with Dependency

5.1 Overview

As we have seen thus far the complex set of operations to precisely quantify risk commensurate with the activities a bank undertakes is a daunting task in of itself. While we have focused on operational risk and the concept of UoMs, we may broaden our scope into market risk activities where individual trading desks require precise risk quantification that are diversified and rolled-up to an aggregate market risk. Alternatively from a credit risk perspective, individual loans/mortgages or even pools of exposures are aggregated and diversified to form an aggregated credit risk. Hence to be explicit, we draw the similarities in that varied risky activities take place on a segmented level across all market, credit and operational risk types and then are naturally thought-of in an aggregate basis. While much attention of advanced quantitative modeling techniques have been discussed and used to reach the top of the mountain (so to speak), the problem of coming back down and reallocating the risk to the contributing sub-portfolios is an equally important discussion that needs to be had. As articulated by Denault (2001) [53], the partitioning problem arises when there is an offsetting in the form of a diversification effect observed in the measurement of financial portfolios. The allocation problem is thus to apportion the diversification advantage in such a way that each portfolio is treated in a fair manner. The benefit of conducting such an exercise is that it allows for comparisons to take place. That is,

knowing the profit/loss generated and the risk taken by the sub-portfolios of a bank allows for a much more relevant comparison than knowing just the profit/loss. This concept is conveyed by the term risk-adjusted return on capital (RAROC).

In James (1996) [77] we learn of the RAROC system developed by Bank of America along with the economic rationale for allocating capital in a diversified organization. The view was adopted that the capital budgeting process was a useful exercise because it suggested that a businesses' contribution to the overall variability of the cash flows of the bank was an important factor in evaluating the risk of (and therefore the capital allocated to) a specific business unit. A useful tool of RAROC is that the platform may be used for both capital budgeting and management compensation. Said concisely, James (1996) [77] stipulates that RAROC systems allocate capital for two basic reasons: (i) risk management and (ii) performance evaluation. Thus, we see the direct application in an operational risk management setting where the overriding goal of conducting the exercise to allocate capital would be to determine the optimal capital structure for each business unit commensurate with the risk. Loosely, we proceed to develop the case that the process involves estimating how much volatility each business unit contributes to the total risk of the bank and thus the aggregate capital requirement.

Another point may be made here with respect to the difference between regulatory capital and economic capital. On the one hand, regulatory capital represents a mandatory capital requirement imposed by regulators for which top-of-the-house adequacy is of greater importance. The sufficiency of capital is thus important under the Pillar I and Pillar III construct as seen in BCBS (2006) [19] and BCBS (2011) [24]. The three Pillar framework was introduced in Section 2.1. For economic capital which is calculated with the intent to assess capital reasonability of a worst-case scenario, the output is used to directly allocate the cost of maintaining a desired capital structure. Thus this ties into a Pillar II requirement and from a bank's perspective, will always be an

important aspect of internal risk management. Thus added research is required in the allocation of operational risk capital and this is chapter explores possible applications.

While we have focused on the VaR risk measure and the regulatory-prescribed sum or both expected and unexpected losses that is required to be capitalized, we may approach the subject from a generic risk measure perspective. The literature tends to follow the same path. Later on however, we will concern ourselves with a special case when we are just concerned with the expected loss and the capital required to offset those losses that are to be expected. The active, day-to-day oversight into capital used to offset expected loss is immediately of great importance for regulation of internal risk taking activities.

The organization of this chapter is as follows. Section 5.2 covers a literature review of methodologies used to partition aggregate capital to sub-portfolios. Section 5.3 introduces a risk measure defined in terms of shortfall risk and shows the convex optimization problem that seeks to minimize capital allocated to sub-portfolios. The formulation and subsequent analysis is heavily motivated by Armenti *et al* (2015) [7]. Section 5.4 replicates results from Armenti *et al* (2015) [7] and establishes a path to follow for an extension. Section 5.5 provides an application of the already developed theory to an empirical data set but seen in a new context. Section 5.6 takes the work by Armenti *et al* (2015) [7], which includes inter-dependent (systemic) risk between sub-portfolios through a specialized loss function, and extends the problem by allowing for a diverse set of dependencies through the use of copulas. The extension allows for highly customizable stress testing capabilities. The output on optimal sub-portfolio capital requirements is analyzed and discussed. Section 5.7 summarizes the work.

5.2 Literature Review

A thorough treatment of a wide variety of capital allocation techniques is covered in Urban *et al* (2004) [110]. The context is that from the insurance industry where risk capital is held to assure policyholders that claims can be paid in the event that claims are larger than expected. A useful takeaway from the paper is that insolvency concerns a company as a whole and while bankruptcy is not defined on sub-portfolios of the insurer, it is useful to think of risk capital as being allocated to different business units or sub-portfolios. Urban *et al* (2004) [110] is a good starting point in that a linkage is drawn between the choice of risk measure and the allocation methodology. Within the paper five risk measures were defined and five allocation methods introduced leading to 25 allocation principles. We define the nuances that must be introduced in an insurance context but soon realize that the problem, once formulated, is applicable in different settings.

The claims for the next period (usually one year) of the different positions of an insurance portfolio were represented by random variables S_1, S_2, \dots, S_n . Under the assumption that cost, commission and interest of risk capital were paid in advance, the premiums were understood to be equal to $E[S_1], E[S_2], \dots, E[S_n]$. Future income for each claim was represented by $X_i = E[S_i] - S_i, i = 1, 2, \dots, n$ where the problem was to take a quantified risk capital and allocate it in a fair way to the different positions of the portfolio. Again, a risk measure ρ was defined as a mapping

$$\rho : L \rightarrow \mathbb{R}, \tag{5.1}$$

where L was a set of real-valued random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The same discussion of coherence of a risk measure still applied as seen previously in Definition 2.4.2. The paper was concerned with centered random variables X_i , i.e. the restriction of the space L to the space L^0 , which consists of all

real-valued centered random variables. With the added clarification that for a random variable X and $\alpha \in (0, 1)$, the upper α -quantile was denoted by $Q^\alpha(X) = \sup\{x \in \mathbb{R} \mid P[X \leq x] \leq \alpha\}$ and for $x \in \mathbb{R}$, $x_+ = \max(x, 0)$ and $x_- = (-x)_+$, the following five risk measures were defined:

1. **Variance:** $\rho_{var}(X) = var[X] = E[X^2] - (E[X])^2$,

2. **Standard deviation:** $\rho_{sd}(X) = \sigma[X] = \sqrt{var[X]}$,

3. **Semi-variance:** $\rho_{svar} = E[((X - E[X])_-)^2]$,

4. **Value-at-Risk(VaR):** $\rho_{VaR(\alpha)}(X) = -Q^\alpha(X)$,

5. **Expected shortfall(ES):**

$$\rho_{ES(\alpha)}(X) = -\frac{1}{\alpha} \left(E[X \mathbf{1}_{\{X \leq Q^\alpha(X)\}}] + Q^\alpha(X)(\alpha - P[X \leq Q^\alpha(X)]) \right).$$

To provide some context, ρ_{var} and ρ_{sd} are common risk measures used in statistics and insurance pricing and portfolio optimization. Semi-variance (ρ_{svar}) is a downward risk measure in that only events below a certain target value are considered as risky. The consideration of the expectation as the target value for ρ_{svar} will suit our analysis later on in that we will be concerned with the expected value. In this paper, the expectation results in $\rho_{svar} = E(X_-)^2$ since we are dealing with centered random variables. Finally, we have already seen $\rho_{VaR(\alpha)}$ and $\rho_{ES(\alpha)}$ in (2.16) and (2.17) respectively where the later is still the average in the tail of the distribution. For a random variable with continuous distribution function, the second term is zero.

To provide a wider array of allocation methodologies, Urban *et al* (2004) [110] accounts for the size of a portfolio position. This arises from the necessity that small risk capital may be considered to represent a less risky position compared to a position that has assigned high risk capital. Hence the idea is to scale each portfolio position by its size and thus allocate risk capital to each portfolio on a per unit basis. The random vector $X = (X_1, X_2, \dots, X_n)$ of incomes is scaled by size

$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ to yield $X = (X_1(\lambda_1), X_2(\lambda_2), \dots, X_n(\lambda_n))$ where the aggregate loss is given by $Z = X_1(\lambda_1) + X_2(\lambda_2) + \dots + X_n(\lambda_n)$. For example, the size may be determined by the premium volumes which could be seen as an exogenous parameter. Hence it is then possible to scale each portfolio position by a size metric which allows for the allocation of capital to each portfolio position per unit premium.

Using this foundation, we may expand upon Definition 4.8.1 and focus the discussion on the set R of risk measures for centered random variables. Then we may define a capital allocation methodology Φ as a mapping

$$\Phi : R \times (L^0)^n \rightarrow \mathbb{R}^n, \quad (\rho, X_1, X_2, \dots, X_n) \rightarrow \begin{pmatrix} \Phi_1(\rho, X_1, X_2, \dots, X_n) \\ \vdots \\ \Phi_n(\rho, X_1, X_2, \dots, X_n) \end{pmatrix}, \quad (5.2)$$

where $\sum_{i=1}^n \Phi_i = 1$ and $\Phi_i \geq 0$ for $i = 1, 2, \dots, n$. Thus Φ_i are viewed as allocation coefficients that partitions risk capital K . Under this expanded scope of portfolios scaled with size, Urban *et al* (2004) [110] introduces five allocation methods that lead to their respective coefficients.

Definition 5.2.1. (*Allocation coefficients*) Define $Z : \mathbb{R}^n \rightarrow L^0$ with $Z(\lambda) = \sum_{i=1}^n X_i(\lambda_i)$ and ρ a risk measure whose partial derivative with respect to λ_i exist for all $i \in N$. We define various allocation coefficients for $i \in N$:

1. **Proportional:** $\Phi_i^{p,\rho} = \frac{\rho(X_i)}{\sum_{j \in N} \rho(X_j)}$,
2. **Merton and Perold:** $\Phi_i^{MP,\rho} = \frac{\rho(Z) - \rho(Z - X_i)}{\sum_{j \in N} [\rho(Z) - \rho(Z - X_j)]}$,
3. **Myers and Read:** $\Phi_i^{MR,\rho} = \frac{\partial \rho(Z(\lambda))}{\partial \lambda_i} / \sum_{j=1}^n \frac{\partial \rho(Z(\lambda))}{\partial \lambda_j}$,
4. **Shapely:** $\Phi_i^{S,\rho} = \frac{1}{\rho(Z)} \sum_{S \subseteq N} \frac{(|S|-1)!(n-|S|)!}{n!} \left(\rho(\sum_{j \in S} X_j) - \rho(\sum_{j \in S \setminus \{i\}} X_j) \right)$,
5. **Aumann and Shapely:** $\Phi_i^{AS,\rho} = \frac{\lambda_i}{\rho(Z(\lambda))} \int_0^1 \frac{\partial \rho(Z(t\lambda))}{\partial \lambda_i} dt$.

The proportional method is by far the most straightforward which is just the individual contribution compared to the total. The Merton and Perold (1993) [90] and Myers and Read (2001) [93] allocation methods are based on the option pricing model of a firm where risk is measured by the price of an insolvency put option. The put option is the loss to policyholders given that the insurer defaults. The Merton and Perold method is simply the marginal risk capital whereby the risk capital required for the firm without a business is subtracted from the risk capital required for the full portfolio and placed over the total marginal impacts. Within Myers and Read (2001) [93], the context is taken from insurance companies and the paper shows how option pricing methods can be used to allocate required capital across lines of insurance. Within the insurance business, capital is referred to as surplus. The allocation method depends on the marginal contribution of each insurance line to default value. That is, the present value of the company's option to default. Also, the marginal default values add up to the total default value hence the capital allocation are unique and not arbitrary. The benefit of the uniqueness of the solution is that if insurance companies have surplus allocations that are wrong, then in a competitive setting, allocation errors may lead firms to write unprofitable businesses and lose profitable businesses to the market. The derivation of the allocation method proceeds by a general expression for marginal default values and obtained by taking derivatives with respect to line-by-line liabilities. The paper shows a general formalization where there is no assumption about the joint probability distribution of line-by-line losses and return of the company's portfolio of assets. The second formulation walks through an example in which the probability distribution of losses and asset values are joint Log-normal and again derives the formula for marginal default values by line of business.

Shapley (1953) [105] and the Aumann and Shapley (1974) [11] method arise from co-operative games in a game theoretic context, the latter of which will be explored a bit more later as it is a commonly used method. The first of these methods has the risk measure ρ denoting the worth of the coalition or rather the total expected sum

of payoffs the members (denoted by S) can obtain by cooperation. Thus for brevity, according to the Shapley value, the amount that player i receives given a coalitional game with value ρ and set N with n players (ρ, N) is given by $\Phi_i^{S,\rho}$. The sum extends over all subsets S of N not containing player i . The formula can be interpreted as imagining the coalition being formed by one player at a time, with each player demanding their contribution $\left(\rho(\sum_{j \in S} X_j) - \rho(\sum_{j \in S \setminus \{i\}} X_j)\right)$ as a fair compensation, and then for each player to take the average of this contribution over the possible different permutations in which the coalition can be formed. The Aumann and Shapley (1974) [11] method is the most commonly used method and is often referred to as the Euler method. The allocation can be summarized as the average of the marginal cost of the i^{th} -portfolio as the level of activity increases uniformly for all portfolios.

Returning back to the discussion in Urban *et al* (2004) [110], the paper is again of the utmost importance because it takes the various risk measures and risk allocations and proceeds via an example. Seven different claim distributions represented by random variables $S_i, i = 1, \dots, 7$ were specified. Again, this structure can be seen as a seven sub-portfolio structure with each sub-portfolio having a unique loss profile. The first four claims were characterized by a compound Poisson model with Pareto claim sizes (similar to the LDA approach). Hence, the parameters produced were based on realistic calibrations to business data: (i) storm, (ii) earthquake, (iii) engineering (major loss) and (iv) fire (major loss) events. The last three claims were modeled via a Lognormal distribution (representing the loss distribution with no preliminary specified compound process) and characterized as (v) general liability (basic losses), (vi) engineering (basic losses) and (vii) fire (basic losses). While the losses were assumed to be independent, the basic losses (last three claims) were assumed to have a dependency. The dependency was a Gaussian copula with pairwise rank correlation 0.14. Hence having specified loss distributions, tables of allocation coefficients for the proportional, Merton and Perold, and Shapley method could be computed for each of the five risk measures. The Myers and Read and Aumann and Shapley methods were

not computed because of the instability of the derivative approximations for small Δ . As well, in the final section of the paper when interpreting results, it was found that under certain combinations for certain claims, expecting that pairwise positively correlated processes should always yield higher allocation coefficients did not always hold. As we will see later on, this type of analysis with dependency and resulting capital allocation will form the extension from Armenti *et al* (2015) [7].

As promised, the most popular capital allocation method of Aumann and Shapley will be discussed further. Much like the work by Artzner *et al* (1999) [9] which took an axiomatic approach to defining what properties a “good” risk measure should have, Denault (2001) [53] follows the same approach to define what “good” properties an allocation principle should have. Denault provides an axiomatic description of what an allocation principle should follow, labels this as a coherent allocation and then shows that the Aumann and Shapley value is both coherent and a practical allocation method. As an aside, Kalkbrener (2005) [80] also takes an axiomatic approach to capital allocation and builds upon the theory to include discussions on completeness, existence and allocation formulae for particular classes of risk measures. Before reproducing the properties for a coherent allocation from Denault (2001) [53], for completeness we introduce some terminology: N represents the set of all portfolios of the firm, A is the set of risk capital allocation problems where pairs (N, ρ) are composed of a set of n portfolios and a coherent risk measure ρ . The mapping $\Pi : A \rightarrow \mathbb{R}^n$ is the allocation principle that maps each allocation problem (N, ρ) into a unique allocation. Finally, $K = \rho(x)$ represents the risk capital of the firm. Then the following definitions follow.

Definition 5.2.2. (*Coherent Allocation Principle*) *An allocation principle Π is coherent if for every allocation problem (N, ρ) with N , the allocation $\Pi(N, \rho)$ satisfies the three properties:*

- *No undercut:* $\forall M \subseteq N, \sum_{i \in M} K_i \leq \rho\left(\sum_{i \in M} X_i\right)$,

- *Symmetry: If by joining any subset $M \subseteq N \setminus \{i, j\}$, portfolios i and j both make the same contribution to the risk capital, then $K_i = K_j$,*
- *Riskless allocation: $K_n = \rho(\alpha r_f) = -\alpha$ where the n^{th} portfolio is a riskless instrument.*

Formally this does not rule out negative allocations where if there is a requirement for $K_i \geq 0, \forall i \in N$, then this is referred to as a non-negative coherent allocation. The interpretation of the above is as follows. The no undercut properties rules out the case when an undercut occurs which results when a portfolio's allocation is higher than the amount of risk capital it would face as a separate entity separate from the bank. So if a portfolio joins the bank, the total risk capital increases by no more than the portfolio's own risk capital. Hence the portfolio can not justifiably be allocated more risk capital than it could possibly have brought to the bank. The symmetry property ensures that a portfolio's allocation depends only on its contribution to risk within the bank. The riskless property indicates that a portfolio should be allocated exactly its risk measure. Thus a portfolio that increases its cash offset should see its allocated capital decrease by the same amount. This can be seen from the translation invariance property where for all $\alpha \in \mathbb{R}$ and random variable X , $\rho(X + \alpha r_f) = \rho(X) - \alpha$ where r_f is the price, at some point in the future, of a reference, riskless investment whose price is 1 today.

The remaining paper continues on casting the problem in a game theoretic framework where the capital allocation problem is modeled as a game between portfolios. The notion is extended to fuzzy games (here fuzzy implies divisibility of players and hence the allowance for fractional players) and thus the optimal allocation principle is the Aumann-Shapley value (Euler allocation) which results in a coherent allocation methodology. In a game setting framework, portfolios are akin to players and a risk measure is akin to a cost function.

Definition 5.2.3. (*Coalitional game*) A coalitional game with fractional players (N, Λ, r) consists of

- a finite set N of players with $|N| = n$,
- a positive vector $\Lambda \in \mathbb{R}_+^n$, each component representing for one of the n players involvement,
- a real-valued cost function $r : \mathbb{R}^n \rightarrow \mathbb{R}, r : \lambda \rightarrow r(\lambda)$ such that $r(0) = 0$.

Hence the ratio $\frac{\lambda_i}{\Lambda_i}$ denotes an activity level for player i . The random variable X_i represents the net worth of the portfolio i at a future time T . Then the cost function r is identified with the risk measure ρ through

$$r(\lambda) = \rho\left(\sum_{i \in N} \frac{\lambda_i}{\Lambda_i} X_i\right), \quad (5.3)$$

so that $r(\Lambda) = \rho(N)$. Thus $\frac{X_i}{\Lambda_i}$ represents the per-unit future net worth of portfolio i . Introducing a vector $k \in \mathbb{R}^n$ to represent the per unit allocation of risk capital to each portfolio, the total capital allocated to each portfolio is obtained by $\Lambda k = K$ componentwise. As one final terminology introduction, we call f a k -homogeneous function, i.e. $f(\gamma x) = \gamma^k f(x)$, then $\frac{\partial f(x)}{\partial x_i}$ is $(k - 1)$ -homogeneous. Thus the stated optimal Aumann-Shapley allocation result is given by

$$k_i^{AS} = \int_0^1 \frac{\partial r}{\partial \lambda_i}(\gamma \Lambda) d\gamma, \quad (5.4)$$

for player i of N . The per-unit cost k_i^{AS} is thus the average of the marginal costs of the i^{th} -portfolio, as the level of activity increases uniformly for all portfolios. As our cost function is r 1-homogeneous ($r(\gamma \lambda) = \gamma r(\lambda)$), we simply write $\frac{\partial r}{\partial \lambda_i}(\Lambda)$. Thus Aumann-Shapley prices are traditionally referred to as

$$k^{KS} = \nabla r(\Lambda). \quad (5.5)$$

Our intent is not to go through the full derivation. However, one added clarification is that this method is also commonly referred to as Euler's allocation because of Euler's theorem which states that if F is a real, n -valued homogeneous function of degree k , then $x_1 \frac{\partial F(x)}{\partial x_1} + x_2 \frac{\partial F(x)}{\partial x_2} + \dots + x_n \frac{\partial F(x)}{\partial x_n} = kF(x)$.

While aspects of the Euler method has been discussed separately in many different papers, a comprehensive overview of the method has been covered in Tasche (2008) [108]. The presentation of the method reinforces that the technique may be applied to any risk measure that is homogenous of degree one and differentiable in an appropriate sense. Moreover, it is stated the Euler's method is well-suited as a tool for the detection of risk concentrations and the treatment goes as far as defining a marginal diversification index. In addition, Tasche (2008) [108] summarizes six ways at which different authors discuss and critique the Euler method. In terms of application, formulae that are needed to calculate Euler contributions were also introduced for standard deviation, VaR, and ES risk measures. In a specialized example, a new approach was also introduced that identified the contributions of underlying names to expected losses of collateralized debt obligation (CDO) tranches using the Euler method. In a final application of the Euler method, non-linear risk impacts on portfolio-wide risks were further analyzed by virtue of a toy portfolio using a set of factors.

As we have seen, the Euler method is a versatile tool in the discussion of capital allocation. If we allow to further expand our universe to include a new notion of a convex measure of risk, this opens up the discussion to a new problem of solving for a concept called the bounded shortfall risk which is in essence an optimization problem. This idea was developed in depth in the seminal paper by Föllmer and Schied (2002) [69]. Actually, Dhaene *et al* (2003) [55] also discuss risk measures based on convex functions and formulate the similar optimization to be solved but do so in a less rigorous manner. As we will need to fully develop the notation and theory in

the next section, we briefly describe the necessary ingredients and their contribution to the optimization problem. We start by assuming we have a financial position X which represents the discounted net worth at the end of a given period. We then apply a convex loss function (to be defined later which overweighs losses) of X . This ensures we have a certain form which allows for the determination of the optimal amount of capital. This is done in such a way that the expected loss is bounded. The convexity allows us to use the theory of convex analysis in that local minimums are actually a global minimums. The shortfall risk is defined as the expectation risk measure applied to the convex loss function on the value X . The bounded condition is an imposed condition to ensure that the risk is not exceeded by some fixed value. This problem has been extended by Armenti *et al* (2015) [7] to solve the multivariate shortfall risk and hence can be viewed as a multivariate extension of Föllmer and Schied (2002) [69]. However, the extension goes further to include systemic risk among different portfolios and continues to solve a modified optimization problem under different loss functions. Our work will further build on the systemic risk component and hence extend the concept further.

As a final note on the coverage of capital allocation, we stay away from the discussion on the dynamic and multi-period problems of capital allocation. That is, Cheridito *et al* (2006) [41] study dynamic monetary risk measures that depend on bounded discrete-time processes of financial values. The paper shows in a very technical manner how time-consistent dynamic risk measures can be constructed by pasting together one-period risk measures. It points out the obvious limitation of Föllmer and Schied (2002) [69] where the setting is static in the sense that the risk of financial values are only measured at the beginning of the time-period and an allocation is made. In the multi-period setting, cash-flow streams or financial processes model the evolution of financial values and risk measurements can be updated as new information becomes available over time. An extension is further made by Cherny (2009) [43] in that the discrete-time coherent risk measure is extended to included unbounded

processes.

5.3 Shortfall Risk Allocation

In this section, we introduce the necessary terminology and background needed to solve our impending capital optimization problem. The necessary material will be reproduced from Föllmer and Schied (2002) [69] and where the multivariate extension is necessary, the material will be reproduced from Armenti *et al* (2015) [7]. We do not claim any originality when reproducing the necessary background. However, we will annotate and summarize concepts in order to facilitate a simpler interpretation where needed.

We start with the inherent interest in a quantitative assessment of the risk attributed to a financial position. In a univariate setting, we may describe the position as the discounted net worth at the end of a give period which is represented by a real-valued function X on some set Ω of possible scenarios. That is, $X : \Omega \rightarrow \mathbb{R}$ where $X(\omega)$ is the net worth if scenario $\omega \in \Omega$ is realized. Then the measure of risk is given by the mapping ρ from a certain space \mathcal{X} of functions on Ω to \mathbb{R} . Actually, the risk $\rho(X)$ of the financial position X can be thought of as the minimal amount of capital that should be added to the position in order to make that position acceptable. We have seen in Artzner *et al* (1999) [9] the properties for a coherent measure of risk. This was introduced previously in Definition 2.4.2 where we had translation invariance, monotonicity, subadditivity and positive homogeneity. A coherent measure of risk ρ arises from a family \mathcal{Q} of probability measures on Ω by computing the expected loss under $Q \in \mathcal{Q}$ and then taking the worst results as Q varies over \mathcal{Q}

$$\rho(X) = \sup_{Q \in \mathcal{Q}} E_Q[-X]. \quad (5.6)$$

Having established the risk measure, we account for the case where the risk of a position might increase in a nonlinear way with the size of a position. The example provided is a liquidity shock that may arise if a position is multiplied by a large factor. Thus, the coherent risk measure properties of subadditivity and positive homogeneity may be relaxed and instead get substituted by the weaker property of convexity

$$\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y) \quad \lambda \in [0, 1]. \quad (5.7)$$

The convexity property implies that diversification does not increase risk and is less than or equal to the linear weighted average of the individual risks.

Extending the concept of a convex risk measure, a representation theorem can be made. That is, sets of acceptable positions can be shown to be isomorphic to a convex risk measure. This notion has been fully developed in Föllmer and Schied (2002) [69] in section 2 of the paper. To start things off, we may define such sets.

Definition 5.3.1. (*Acceptance set*) Let \mathcal{X} be a linear space of functions on a given set Ω of possible scenarios. Assume \mathcal{X} contains all constant functions. Any risk measure $\rho : \mathcal{X} \rightarrow \mathbb{R}$ induces an acceptance set \mathcal{A}_ρ defined as

$$\mathcal{A}_\rho := \{X \in \mathcal{X} \mid \rho(X) \leq 0\}. \quad (5.8)$$

Conversely the way to go back is that for a given class \mathcal{A} of acceptable positions, the associated risk measure $\rho_{\mathcal{A}}$ is defined as

$$\rho_{\mathcal{A}}(X) := \inf\{m \in \mathbb{R} \mid m + X \in \mathcal{A}\}. \quad (5.9)$$

With this relation in hand, we can now state the representation theorem for convex measures of risk. We first consider the special case in which \mathcal{X} is the space of all real-valued functions on some finite set Ω .

Theorem 5.3.1. *Suppose \mathcal{X} is the space of all real-valued functions on a finite set Ω . Then $\rho : \mathcal{X} \rightarrow \mathbb{R}$ is a convex measure of risk if and only if there exists a penalty function $\alpha : \mathcal{P} \rightarrow (-\infty, \infty]$ such that*

$$\rho(X) = \sup_{Q \in \mathcal{P}} (E_Q[-X] - \alpha(Q)). \quad (5.10)$$

The function α satisfies $\alpha(Q) \geq -\rho(0)$ for any $Q \in \mathcal{P}$, and it can be taken to be convex and lower semicontinuous on \mathcal{P} .

For the proof, see Föllmer and Schied (2002) [69]. As we will see, this problem admits a dual representation as is possible in convex optimization problems. We will state this in a theorem later on.

5.3.1 Risk Measures Represented as Shortfall Risk

In order to facilitate a solution to (5.6), we resort to not just loss of X but convex representations of X . Suppose that $\ell : \mathbb{R} \rightarrow \mathbb{R}$ is an increasing convex loss function which is not identically constant. For a position $X \in L^\infty(\Omega, \mathcal{F}, \mathcal{P})$ we introduce the expected loss

$$E_P[\ell(-X)], \quad (5.11)$$

and refer to this as shortfall risk. Note that we constrain ourselves to the practical application where \mathcal{X} is given by the space $L^\infty(\Omega, \mathcal{F}, \mathcal{P})$ of bounded functions on a general probability space. At this stage, we may fix a point c in the range of ℓ (or rather $c > \inf \ell$) and define the acceptance set. That is, a position $X \in L^\infty(\Omega, \mathcal{F}, \mathcal{P})$ is acceptable if the expected loss is bounded by the threshold c and denoted by

$$\mathcal{A} := \{X \in L^\infty(\Omega, \mathcal{F}, \mathcal{P}) \mid E_P[\ell(-X)] \leq c\}. \quad (5.12)$$

The set \mathcal{A} is thus convex and the convex risk measure corresponding to \mathcal{A} is $\rho := \rho_{\mathcal{A}}$. The corresponding minimal penalty function $\alpha_0(\cdot)$ can be expressed in terms of the Fenchel-Legendre transform of ℓ given by

$$\ell^*(z) := \sup_{x \in \mathbb{R}} (zx - \ell(x)). \quad (5.13)$$

This is sometimes referred to as the convex conjugate. Geometrically, this means we seek a point x on the function $\ell(x)$ such that the slope of the line z passing through $(x, \ell(x))$ has a maximum intercept on the y axis. Thus, this is the point on the curve that has a slope z which is tangent at the point $z = \ell'(x)$. Again the transformation maps the $(x, \ell(x))$ space to the space of slope and conjugate $(z, \ell^*(z))$.

Finally, the main result of Föllmer and Schied (2002) [69] is given by the following theorem, the proof of which is very lengthy and we choose to omit. It showcases the dual problem and the optimization problem.

Theorem 5.3.2. *Suppose that \mathcal{A} is the acceptance set given by (5.12). Then, for $Q \ll P$, the minimal penalty function of $\rho = \rho_{\mathcal{A}}$ is given by*

$$\alpha_0(Q) = \sup_{X \in \mathcal{A}} E_Q[-X] = \inf_{\lambda > 0} \frac{1}{\lambda} \left(c + E_P \left[\ell^* \left(\lambda \frac{dQ}{dP} \right) \right] \right). \quad (5.14)$$

This theorem thus states that the infimum over all allocation used for defining $\rho(X)$ is real valued and has the desired properties of a risk measure. The questions of existence and uniqueness are covered in the multivariate setting.

5.3.2 Multivariate Shortfall Risk Allocation with Systemic Risk

While the seminal paper by Föllmer and Schied (2002) [69] laid the foundational work for the convex measure of risk and the notion of shortfall risk, Armenti *et al*

(2015) [7] extends and presents the concepts while focusing on two main issues relevant for systemic risk: (i) the computation of an overall reverse level for a financial network to overcome unexpected stress or default and (ii) the reserve allocation to different sub-portfolios according to their systemic involvement. The structure of the presentation had four main goals: (i) develop a class of systemic risk measures, (ii) assess the impact of the dependence structure of the system on risk allocation, (iii) analyze the sensitivity of the allocation with respect to exogenous shocks and (iv) test efficient numerical schemes. The applied nature of the problem and the treatment of the topic gave an intuitive sense of the wide applicability for further investigation.

We may envision a bank, for instance, for real time monitoring purposes wanting to channel to each trading desk a cost reflecting its responsibility in the overall capital requirement. A central clearing counterparty (CCP), also known as a clearing house, is interested in quantifying the size of the so-called default fund and allocating it in a meaningful way among the different clearing members. The importance of the CCP has drawn a lot of attention since the 2007/2008 global financial crisis. The CCP interposes itself between counterparties to contracts traded in one or more financial markets, becoming the buyer to every seller and the seller to every buyer and thereby ensuring the future performance of open contracts. A CCP becomes counterparty to trades with market participants through novation, an open offer system, or another legally binding arrangement. The shift to using CCPs has the intended impact to clear a lot of over the counter (OTC) derivative trading through CCPs for the purpose of regulatory capital relief. However, non-centrally cleared contracts are still permitted but subject to higher capital requirements. For the purposes of the capital framework, a CCP is a financial institution.

In order to set the groundwork for a multivariate extension, we define some basic notation. Let x_k denote the generic coordinate of a vector $x \in \mathbb{R}^d$. By \geq we denote the lattice order on \mathbb{R}^d , that is, $x \geq y$ if and only if $x_k \geq y_k$ for every

$1 \leq k \leq d$. We denote by $\|\cdot\|$ the Euclidean norm and by $\pm, \wedge, \vee, |\cdot|$ the lattice operations on \mathbb{R}^d . For $x, y \in \mathbb{R}^d$, we write $x > y$ for $x_k > y_k$ componentwise, $x \cdot y = \sum x_k y_k$, $xy = (x_1 y_1, \dots, x_d y_d)$ and $x/y = (x_1/y_1, \dots, x_d/y_d)$.

Let (Ω, \mathcal{F}, P) denote a probability space, where P represents the objective probability measure, with related expectation denoted by E . We denote by $L^0(\mathbb{R}^d)$ the space of \mathcal{F} -measurable d -variate random variables on this space. The space $L^0(\mathbb{R}^d)$ inherits the lattice structure on \mathbb{R}^d , hence we can use the above notation in a P -almost sure sense. For instance, for $X, Y \in L^0(\mathbb{R}^d)$, we say that $X \geq Y$ or $X > Y$ if $P[X > Y] = 1$, respectively. Since we mainly deal with multivariate functions or random variables, to simplify notation we drop the references to \mathbb{R}^d in $L^0(\mathbb{R}^d)$, writing simply L^0 unless a particular dimension is meant, mainly for $L^0(\mathbb{R})$ in the case of univariate random variables.

Let $X = (X_1, \dots, X_n) \in L^0$ represent a random vector of financial losses, that is, negative values of X_k represent actual profits. This is different than in (5.6) where previously we imposed a minus sign on the net positive net worth. It is just a matter of convention that will make simulation more intuitive in that we will need to think in terms of losses. We want to determine an overall monetary measure $\rho(X)$ of the risk of X as well as a sound risk allocation $\Pi_k(X), k = 1, \dots, d$ of $\rho(X)$ among the d risk factors.

As we will see, by virtue of considering a flexible class of risk measures defined by means of loss functions and sets of acceptable monetary allocations, there are resulting properties that allow for the discussion on risk allocation. Inspired by the shortfall risk measure introduced by Föllmer and Schied (2002) [69] in the univariate case, we will require the use of the loss function ℓ defined on \mathbb{R}^d , used to measure the expected loss $E[\ell(X)]$ of the financial loss vector X . This was introduced in (5.11) however now we define some additional properties.

Definition 5.3.2. (*loss function*) A function $\ell : \mathbb{R}^d \rightarrow (-\infty, \infty]$ is called a loss function if

1. ℓ is increasing, that is, $\ell(x) \geq \ell(y)$ if $x \geq y$,
2. ℓ is convex, lower semi-continuous and finite on some neighbourhood of 0,
3. $\ell(0) = 0$ and $\ell(x) \geq \sum x_k$ on \mathbb{R}^d .

A risk neutral assessment of the losses corresponds to $E[\sum X_k] = \sum E[X_k]$. This implies that the third property expresses a form of risk aversion, whereby higher loss are worse and weighted more than compared to a risk neutral evaluation. The first two properties express the concept that for more losses the risk is increased and that diversification should not increase risk. In the next example, we will see examples of loss functions used on \mathbb{R} and then extend them to higher dimensions. To our knowledge aside from this reproduction, little emphasis has been placed on various forms of loss functions. By introducing some interesting examples, it opens up a discussion for numerical analysis and also highlights the potential for systemic risk derived from the form of the loss function.

Example 5.3.1. Let $h : \mathbb{R} \rightarrow (-\infty, \infty]$ be a one-dimensional loss function which is convex, increasing, lower semi-continuous such that $h(0) = 0$ and $h(x) \geq x$ for every $x \in \mathbb{R}$ and $h(x) \geq (x)$ for every $x \in \mathbb{R}$. Classical examples of loss functions are

$$h(x) = \beta x^+, \quad \beta > 1, \quad (5.15)$$

$$h(x) = x + \frac{(x^+)^2}{2}, \quad (5.16)$$

$$h(x) = e^x - 1. \quad (5.17)$$

Referring to (5.15), recall $x^+ = \max(x, 0)$ and we orient our convention such that positive values of x are losses and hence negative values of x are profits. It is easy to

see this is convex, increasing and even continuous. The scaling by β penalizes losses but not as much as (5.17) by comparison. We will eventually use a variant of (5.16) as it is related to a mean-variance penalization of the losses, is smoother than (5.15) while being less abrupt than (5.17), hence yielding a good trade-off for optimization routines.

Using $h(x)$ as building blocks, we obtain the following classes of multivariate loss functions.

$$\ell(x) = h\left(\sum x_k\right), \quad (5.18)$$

$$\ell(x) = \sum h(x_k), \quad (5.19)$$

$$\ell(x) = \alpha h\left(\sum x_k\right) + (1 - \alpha) \sum h(x_k), \text{ where } 0 \leq \alpha \leq 1, \quad (5.20)$$

$$\ell(x) = \sum x_k + \frac{1}{2} \sum (x_k^+)^2 + \alpha \sum_{j < k} x_j^+ x_k^+, \text{ where } 0 \leq \alpha \leq 1, \quad (5.21)$$

$$= \sum x_k + \frac{1 - \alpha}{2} \sum (x_k^+)^2 + \frac{\alpha}{2} \sum (x_k^+)^2. \quad (5.22)$$

Armenti *et al* (2015) [7] uses (5.21) for investigating numerical simulation in two ways. As we will see, if we set $\alpha = 0$, we zero out the co-mingled term and just focus on independent sub-portfolios and solve for the optimal capital allocation problem. If we let α vary, then in essence this dictates a degree of inter-dependence. We note that in the (5.21), the term $\sum_{j < k} x_j^+ x_k^+$ implies that loss functions are permutation invariant. Hence, this addresses the requirement that loss functions should not discriminate some factors against others. Thus the loss function should be invariant under permutation of its variables. We will show the equivalence of (5.21) and (5.22) later in a three-portfolio case that will make working with either representation equivalent. One other clarification that has been pointed out is that for integrability and topological reasons, loss vectors are restricted to the following multivariate Orlicz heart

$$M^\theta = \{X \in L^0 \mid E[\theta(\lambda X)] < \infty \forall \lambda \in \mathbb{R}^+\}, \quad (5.23)$$

where $\theta(x) = \ell(|x|), x \in \mathbb{R}^d$. We will not focus on this requirement, but provide some motivating arguments. We learn in Cheridito and Li (2009) [42] that the study of such $(-\infty, \infty]$ -valued coherent, convex and monetary risk measures on maximal subspaces of Orlicz classes are referred to Orlicz hearts. Loosely, if a Banach space L^p is a vector space of all measurable functions f for which $\|f\|_p$ is finite, then if $F : [0, \infty) \rightarrow [0, \infty)$ is non-decreasing and convex with $F(0) = 0$, then an Orlicz space L_F contain measurable functions f for which $\|f\|_F$ is finite. The benefit of restricting the analysis within the Orlicz heart is that it includes all L^p spaces for $1 \leq p < \infty$ which allow for the duality theory without additional continuity assumptions. Focusing on the notion of acceptable sets of the dual problem, we recast (5.12) and the notion of acceptability in \mathbb{R}^d .

Definition 5.3.3. (*Acceptable monetary allocation*) A monetary allocation $m \in \mathbb{R}^d$ is acceptable for X at the loss level $c > 0$ if

$$E[\ell(X - m)] \leq c, \tag{5.24}$$

and

$$\mathcal{A}(X) := \{m \in \mathbb{R}^d : E[\ell(X - m)] \leq c\}, \tag{5.25}$$

is the set of acceptable monetary allocations.

The link in the multivariate setting is much more intuitive. That is, given an acceptable monetary allocation $m \in \mathcal{A}(X)$, the aggregate capital requirement (or liquidity cost depending on the context) is $\sum m_k$. Hence, the smaller the requirement, the better it is from a capital savings perspective. This then leads to the very problem we will concern ourselves with throughout the chapter which is the multivariate shortfall risk of $X \in M^\theta$

$$\rho(X) = \inf \left\{ \sum m_k : E[\ell(X - m)] \leq c \right\}. \tag{5.26}$$

In the most simplest of terms, we aim to minimize all the m_k such that the aggregate

money m offsets the losses X in $E[\ell(X - m)]$ under the penalty function ℓ . We do this so that we are under c dollars as required by some risk officer or regulator. For some more context, in a centrally-cleared trading setup, each clearing member k is required to post a default fund contribution m_k in order to make the risk of the clearing house acceptable with respect to a risk measure. The default fund contribution can be used in the case of liquidation of any member. For the determination of the default fund contributions, the methodology can be applied to the vector X defined as the vector of stressed profit-and-losses of the clearing members. Hence this can also be used as a stress testing risk management tool.

We restate the duality theorem but in the multivariate setting which aids in the next theorem that guarantees optimal portions of m_k . The theorem will be akin to taking derivatives and solving under certain conditions.

Theorem 5.3.3. *The function*

$$\rho(X) = \inf \left\{ \sum m_k : m \in \mathcal{A}(X) \right\}, \quad X \in M^\theta, \quad (5.27)$$

is real valued, convex, monotone and translation invariant. In particular, it is continuous and sub-differentiable. It admits the dual representation

$$\rho(X) = \sup_{Q \in \mathcal{Q}^{\theta^*}} \{E_Q[X] - \alpha(Q)\}, \quad X \in M^\theta, \quad (5.28)$$

where \mathcal{Q}^{θ^} is the set of measures Q on the product space $\Omega \times \{1, \dots, d\}$ with density Y in L^{θ^*} normalized to d in the sense $E[1 \cdot \frac{dQ}{dP}] = d$, and where the penalty function is given by*

$$\alpha(Q) = \inf_{\lambda > 0} \frac{1}{\lambda} \left(c + E \left[\ell^* \left(\lambda \frac{dQ}{dP} \right) \right] \right), \quad Q \in \mathcal{Q}^{\theta^*}. \quad (5.29)$$

A final restriction on the loss function helps with the existence of a risk allocation.

Definition 5.3.4. (*zero-sum allocation*) A zero-sum allocation is a monetary allocation $u \in \mathbb{R}^d$ such that $\sum u_k = 0$. Also for any zero-sum allocation u

$$\ell(0) = 0 = \sum u_k \leq \ell(u). \quad (5.30)$$

Definition 5.3.5. (*unbiased*) A loss function ℓ is unbiased if for every zero-sum allocation u , $\ell(ru) = 0$ for any $r > 0$ implies that $\ell(-ru) = 0$ for any $r > 0$.

Theorem 5.3.4. If ℓ is an unbiased loss function, then for every $X \in M^\theta$ risk allocations m^* exist. They are characterized by the first order conditions

$$1 \in \lambda^* E[\nabla \ell(X - m^*)] \quad \text{and} \quad E[\ell(X - m^*)] = c, \quad (5.31)$$

where λ^* is a Lagrange multiplier. In particular, when ℓ has no zero-sum direction of recession except 0, the set of solutions (m^*, λ^*) to the first order condition (5.31) is bounded. If ℓ is strictly convex outside \mathbb{R}_-^d along zero-sum allocations, then the risk allocation is unique.

We refer the reader to Armenti *et al* (2015) [7] for the proofs of the last two theorems. However, since this theorem allows us to find optimal allocations, we provide an example in a bivariate case. The paper jumps to cover a two portfolio case but lacked in its description of the steps to follow. We add that in multivariable Calculus terms, the method of Lagrange multipliers indicates that in order to find max/min values of $f(x, y)$ subject to $g(x, y) = c$, we find values of x, y, λ such that

$$\nabla f(x, y) = \lambda \nabla g(x, y) \quad (5.32)$$

$$g(x, y) = c. \quad (5.33)$$

Adapting this to our notation, $f(m) = m_1 + m_2$, $g(m) = E[\ell(X - m)]$ where

$X = (X_1, X_2)$ are vector profit-losses distributed according to some probability distribution and $m = (m_1, m_2)$. Differentiation denoted by ∇ is with respect to m_1 and m_2 so we are solving $f_{m_1} = \lambda g_{m_1}$ and $f_{m_2} = \lambda g_{m_2}$. The constraint imposed is at a level c .

We also make a remark in that the positivity of the risk allocation is not required. That is, we may have instances were in the optimization routine, we achieve negative values of m_k . Stated differently, if the losses observed in the sub-portfolios are not too high or the threshold c set by the overseeing body is generous and high, we may yield a capital surplus in the form of m_k that could be allocated to another sub-portfolio or carried over to the next period. However, if positivity or any other convex constraint is imposed, it can easily be embedded in the set up. In case of positivity, this would modify the definition of $\rho(X)$ into

$$\rho(X) = \inf\left\{\sum m_k : E[\ell(X - m)] \leq c \quad \text{and} \quad m_k \geq 0 \quad \text{for every } k\right\}. \quad (5.34)$$

Before we move onto our unique application and set of numerical simulation in the next section, we break down an example from Armenti *et al* (2015) [7] in order to gain comfort with applying the first order conditions while applying additional commentary and clarity. Also, we get a sense at how the systemic risk component is embedded in the loss function.

Example 5.3.2. *Consider a bivariate framework with the following loss function*

$$\ell(x_1, x_2) = \frac{(x_1^+)^2 + (x_2^+)^2}{2} + \alpha x_1^+ x_2^+ + x_1 + x_2, \quad 0 \leq \alpha \leq 1. \quad (5.35)$$

To solve the problem of determining optimal capital allocations under this loss penalty function, we must use (5.31). We have the liquidity constraint $f(m) = m_1 + m_2$ which is on the left hand side. Thus using the method of Lagrange multipliers in (5.32), we must compute $\nabla f(m)$. The right hand side contains $E[\ell(X - m)]$ for which the

derivative passes through the expectation and we must compute $\lambda E[\nabla \ell(X - m)]$. The constraint term (5.33) is simply $E[\ell(X - m)] = c$. Thus we will determine the allocation given by $(m_1, m_2) = (\Pi_1(X), \Pi_2(X))$. Thus the first equation is given by

$$\begin{aligned} \frac{\partial}{\partial m_1}(m_1 + m_2) &= \lambda \frac{\partial}{\partial m_1} E \left[\frac{((X_1 - m_1)^+)^2 + ((X_2 - m_2)^+)^2}{2} + \right. \\ &\quad \left. \alpha(X_1 - m_1)^+(X_2 - m_2)^+ + X_1 - m_1 + X_2 - m_2 \right] \\ 1 &= \lambda \left[-E[(X_1 - m_1)^+] - \alpha E[(X_2 - m_2)^+ \mathbf{1}_{X_1 \geq m_1}] - 1 \right] \\ \frac{1}{\lambda} - 1 &= E[(X_1 - m_1)^+] + \alpha E[(X_2 - m_2)^+ \mathbf{1}_{X_1 \geq m_1}]. \end{aligned} \quad (5.36)$$

By symmetry we have the other first order condition and the constraint equation which is set equal to c

$$\frac{1}{\lambda} - 1 = E[(X_2 - m_2)^+] + \alpha E[(X_1 - m_1)^+ \mathbf{1}_{X_2 \geq m_2}] \quad (5.37)$$

$$\begin{aligned} c &= \frac{1}{2} [((X_1 - m_1)^+)^2 + ((X_2 - m_2)^+)^2] + \alpha E[(X_1 - m_1)^+(X_2 - m_2)^+] + \\ &\quad E[X_1] + E[X_2] - m_1 - m_2. \end{aligned} \quad (5.38)$$

Thus our system is composed of (5.36)-(5.38). While we can not attempt solve this analytically until some assumptions are made on the distributions of X , we may dissect the first component (5.36). We see that the term $E[(X_1 - m_1)^+]$ involves the loss X_1 exceeding its liquidity requirement m_1 and the second term weighted by α on the loss X_2 in excess over the liquidity requirement m_2 conditioned on the case that the loss X_1 also exceeds its own liquidity requirement m_1 as indicated by the term $\alpha E[(X_1 - m_1)^+ \mathbf{1}_{X_2 \geq m_2}]$. The reason this was highlighted was that the second term clearly shows the need to account for the dependency structure between X_1 and X_2 which was induced from the loss function.

5.4 Numerical Replication

The in-depth numerical analysis covered in Armenti *et al* (2015) [7] focused on the loss function (5.21). To make the discussion focused around d -variate sub-portfolios, we make this explicit

$$\ell(x) = \sum_{k=1}^d x_k + \frac{1}{2} \sum_{k=1}^d (x_k^+)^2 + \alpha \sum_{1 \leq j < k \leq d} x_j^+ x_k^+, \text{ where } 0 \leq \alpha \leq 1. \quad (5.39)$$

As an inroads in the paper, a bivariate case was first investigated. Gaussian distributions with $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$ were first assumed with the interdependency term zeroed out by setting $\alpha = 0$. Analytic representations could be achieved and optimal capital levels calculated. Monte Carlo methods also confirmed the same result as the analytic method. Adding complexity, the systemic factor was set to $\alpha = 1$ and then correlation sensitivities analyzed for the cases $\rho = \{-0.9, -0.5, -0.2, 0, 0, 2, 0.5, 0.9\}$. From there, a trivariate case was investigated under $\alpha = 0$ and $\alpha = 1$ but using a modified covariance matrix. Our walk through will follow a similar path.

Our extension will use the same loss function as in (5.39) where we will build in systemic dependencies between the historical returns/losses of x_k . This may either naturally arise due to an observed effect arising from the natural linkage between portfolios or enforced as a stress testing tool to perform what-if analysis. In order to apply Theorem 5.3.4, we define the following functions

$$\begin{aligned} e_k(m) &= E[(X_k - m)], \\ f_k(m) &= E[(X_k - m)^+], \\ g_k(m) &= E[((X_k - m)^+)^2], \\ h_{j,k}(m, n) &= E[(X_j - m)^+(X_k - n)^+] \quad (\text{needed when } \alpha \neq 0), \\ l_{j,k}(m, n) &= E[(X_j - m)^+ \mathbf{1}_{\{X_k \geq n\}}] \quad (\text{needed when } \alpha \neq 0). \end{aligned} \quad (5.40)$$

Then our first order conditions yield the following system to be solved

$$\begin{cases} \lambda f_k(m_k) + \alpha \lambda \sum_{j=1, j \neq k}^d l_{j,k}(m, n) = 1 - \lambda & \text{for } k = 1, \dots, d \\ \sum_{k=1}^d \left\{ e_k(m_k) + \frac{1}{2} g_k(m_k) \right\} + \alpha \sum_{1 \leq j < k \leq d} h_{j,k}(m, n) = c. \end{cases} \quad (5.41)$$

In order to solve this system, a numerical method is needed to solve the expectations together with a root finding procedure. We select to use Monte Carlo methods to compute the functions in (5.40). The main advantage is a variety of models may be considered such as models with copulas. The Monte Carlo method proceeds by generating all M realizations of the random vector $X = (X_1, X_2, \dots, X_d)$ used in the estimation of (5.40) for arbitrary m, n during the root finding procedure. The estimators provided are typical of Monte Carlo methods by taking the arithmetic mean over M iterations

$$\begin{aligned} I_M^{MC}[e_k(m)] &= \frac{1}{M} \sum_{i=1}^M (X_k^i - m), \\ I_M^{MC}[f_k(m)] &= \frac{1}{M} \sum_{i=1}^M (X_k^i - m)^+, \\ I_M^{MC}[g_k(m)] &= \frac{1}{M} \sum_{i=1}^M ((X_k^i - m)^+)^2, \\ I_M^{MC}[h_{j,k}(m, n)] &= \frac{1}{M} \sum_{i=1}^M (X_j^i - m)^+ (X_k^i - n)^+, \\ I_M^{MC}[l_{j,k}(m, n)] &= \frac{1}{M} \sum_{i=1}^M (X_j^i - m)^+ \mathbf{1}_{\{X_k^i \geq n\}}. \end{aligned} \quad (5.42)$$

Operationally, this is done in MATLAB using anonymous functions and the function `fsolve` to solve the roots of the system.

Method	m_1^*
Analytic	-0.1731
Monte Carlo 2 million	-0.1724
Monte Carlo 10 million	-0.1730

Table 5.32: Bivariate case with systemic risk $\alpha = 0$.

Before we continue with our extension, we provide some intuition by explaining some examples covered in Armenti *et al* (2015) [7] but run our code as a replication exercise. We start by considering the case $d = 2$ with a bivariate Gaussian distributions with $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$ and $\alpha = 0$. We set the bound to be $c = 1$. To provide some context, we assume that the mean loss of the two sub-portfolios are zero with variance equal to a dollar. By setting the bound to be 1, we would like to then determine how much capital to set aside (a static amount at the beginning of the period) such that the expected performance of the two sub-portfolios remain bounded from above. In Table 5.32, we only report the first optimal capital value because the allocation is symmetric. We provide the analytic derivation of the functions required to solve equations (5.41) with $\alpha = 0$. We can simplify $f_k(x)$ and $g_k(x)$ in analytical form

$$f_k(m) = E[(X_k - m)^+] = \frac{\sigma_k}{\sqrt{2\pi}} e^{\frac{-m^2}{2\sigma_k^2}} - m\Phi\left(\frac{-m}{\sigma_k}\right), \quad (5.43)$$

$$g_k(m) = E[((X_k - m)^+)^2] = (m^2 + \sigma_k^2)\Phi\left(\frac{-m}{\sigma_k}\right) - \frac{m\sigma_k}{\sqrt{2\pi}} e^{\frac{-m^2}{2\sigma_k^2}}. \quad (5.44)$$

To see how we get (5.43), we can solve this step by step. After the first step, we drop the k index with the understanding that this applies up to $k = d$ portfolios

$$\begin{aligned}
f_k(m) &= E[(X_k - m)^+] \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - m)^+ e^{\frac{-x^2}{2\sigma^2}} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_m^{\infty} (x - m) e^{\frac{-x^2}{2\sigma^2}} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_m^{\infty} x e^{\frac{-x^2}{2\sigma^2}} dx - \frac{m}{\sigma\sqrt{2\pi}} \int_m^{\infty} e^{\frac{-x^2}{2\sigma^2}} dx \\
&= \frac{-\sigma^2}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}} \Big|_m^{\infty} - \frac{m}{\sigma\sqrt{2\pi}} \int_m^{\infty} e^{\frac{-x^2}{2\sigma^2}} dx \\
&= \frac{-\sigma^2}{\sigma\sqrt{2\pi}} \left[0 - e^{\frac{-m^2}{2\sigma^2}} \right] - \frac{m}{\sigma\sqrt{2\pi}} \int_m^{\infty} e^{\frac{-x^2}{2\sigma^2}} dx \\
&= \frac{\sigma}{\sqrt{2\pi}} e^{\frac{-m^2}{2\sigma^2}} - m\Phi\left(\frac{-m}{\sigma}\right).
\end{aligned}$$

Note the last term was obtained by:

$$\begin{aligned}
&\frac{1}{\sigma\sqrt{2\pi}} \int_m^{\infty} e^{\frac{-x^2}{2\sigma^2}} dx \\
&\Rightarrow P\{X > m\} \\
&= 1 - P\{X \leq m\} \\
&= 1 - P\left\{\frac{x - 0}{\sigma} \leq \frac{m}{\sigma}\right\} \\
&= 1 - \Phi\left(\frac{m}{\sigma}\right) \quad \text{where } \Phi \text{ denotes the CDF of a standard normal random variable} \\
&= \Phi\left(\frac{-m}{\sigma}\right).
\end{aligned}$$

A similar method is used to derive $g_k(m)$

$$\begin{aligned}
 g_k(m) &= E[((X_k - m)^+)^2] \\
 &= \int_{-\infty}^{\infty} ((x - m)^+)^2 p(x) dx \\
 &= \int_m^{\infty} (x - m)^2 p(x) dx \\
 &= \int_m^{\infty} (x^2 - 2xm + m^2) p(x) dx \\
 &= \int_m^{\infty} x^2 p(x) dx - 2m \int_m^{\infty} x p(x) dx + m^2 \int_m^{\infty} p(x) dx \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \int_m^{\infty} x^2 e^{-\frac{x^2}{2\sigma^2}} dx - 2m \underbrace{\left(\frac{\sigma e^{-\frac{m^2}{2\sigma^2}}}{\sqrt{2\pi}} \right)}_{\text{from before}} + m^2 \underbrace{\Phi\left(\frac{-m}{\sigma}\right)}_{\text{from before}}.
 \end{aligned}$$

Note

$$\begin{aligned}
 \frac{1}{\sigma\sqrt{2\pi}} \int_m^{\infty} x^2 e^{-\frac{x^2}{2\sigma^2}} dx &= \frac{1}{\sigma\sqrt{2\pi}} \int_m^{\infty} x \cdot x e^{-\frac{x^2}{2\sigma^2}} dx \\
 u = x, du = dx, \quad dv = x e^{-\frac{x^2}{2\sigma^2}}, v = -\sigma^2 e^{-\frac{x^2}{2\sigma^2}} & \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \left[uv - \int v du \right] \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \left[-x\sigma^2 e^{-\frac{x^2}{2\sigma^2}} \Big|_m^{\infty} + \int_m^{\infty} \sigma^2 e^{-\frac{x^2}{2\sigma^2}} dx \right] \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \left[0 + m\sigma^2 e^{-\frac{m^2}{2\sigma^2}} + \sigma^2 \int_m^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx \right] \\
 &= \frac{m\sigma e^{-\frac{m^2}{2\sigma^2}}}{\sqrt{2\pi}} + \sigma^2 \Phi\left(\frac{-m}{\sigma}\right).
 \end{aligned}$$

Sim. (M)	100,000	200,000	500,000	1,000,000	2,000,000	5,000,000	10,000,000
Mean	-0.1737	-0.1729	-0.1736	-0.1730	-0.1731	-0.1731	-0.1731
5 th %	-0.1819	-0.177	-0.1762	-0.1743	-0.1745	-0.1739	-0.1738
95 th %	-0.1693	-0.1681	-0.1704	-0.1716	-0.1720	-0.1720	-0.1725
Std. Dev.	0.0036	0.0024	0.0018	0.0011	0.0008	0.0006	0.0004
CV (%)	2.0	1.4	1.0	0.6	0.4	0.3	0.2
Time (sec)	4.84	9.91	24.09	47.16	93.06	225.69	461.01

Table 5.33: Bivariate case with systemic risk $\alpha = 0$ averaged over $N = 30$ runs.

Continuing we have

$$\begin{aligned}
g_k(m) &= \frac{m\sigma e^{-\frac{m^2}{2\sigma^2}}}{\sqrt{2\pi}} + \sigma^2 \Phi\left(\frac{-m}{\sigma}\right) - 2m\left(\frac{\sigma e^{-\frac{m^2}{2\sigma^2}}}{\sqrt{2\pi}}\right) + m^2 \Phi\left(\frac{-m}{\sigma}\right) \\
&= (m^2 + \sigma_k^2) \Phi\left(\frac{-m}{\sigma_k}\right) - \frac{m\sigma_k}{\sqrt{2\pi}} e^{-\frac{m^2}{2\sigma_k^2}}.
\end{aligned}$$

This form of analytical $f_k(m)$ and $g_k(m)$ checks out in MATLAB with Monte Carlo simulations. We also see with approximation methods such as Monte Carlo, we can easily seek convergence to the analytical form by increasing the number of simulations. To provide some more detail on convergence and confidence intervals, we summarize some statistics in Table 5.33. We notice that running $M = 2,000,000$ simulations and averaging over 30 runs, that the mean matches the analytic value. Moreover, even if we select to only run 100,000 simulations, the 5th and 95th contain the optimal value. We also compute the Coefficient of Variance (CV) which is also known as relative standard deviation and is a measure of dispersion. It is computed as the ratio of the standard deviation to the mean and may be reported as a percentage. On a relative basis by running 2,000,000 simulations, we obtain a CV less than 0.5%. It is also advantageous to understand when using only 200,000, that we obtain a wider confidence interval but still manage a CV less than 2%. When we move away from a stylized example and utilize an empirical data set with systemic risk and correlated portfolios, our computation time increases substantially. However, we may hope to achieve a similar CV that yields results that are still have relatively low dispersion.

Referring back to Table 5.32, we interpret a negative m_1^* to mean there is slack in the allocation. That is, there is excess resource allocated over the minimum necessary to accomplish the task. Said differently, there is a savings of sort that need not be used and could be used elsewhere throughout an organization. We see how this changes directly by successively lowering c as shown in Table 5.34 and the linear relationship as shown in Figure 5.6.

Bound	Method	m_1^*
$c = 1$	Analytic	-0.1731
$c = 0.8$	Analytic	-0.1052
$c = 0.6$	Analytic	-0.0355
$c = 0.5$	Analytic	0
$c = 0.4$	Analytic	0.0360
$c = 0.2$	Analytic	0.1093

Table 5.34: Bivariate case with systemic risk $\alpha = 0$ with varying bounds c .

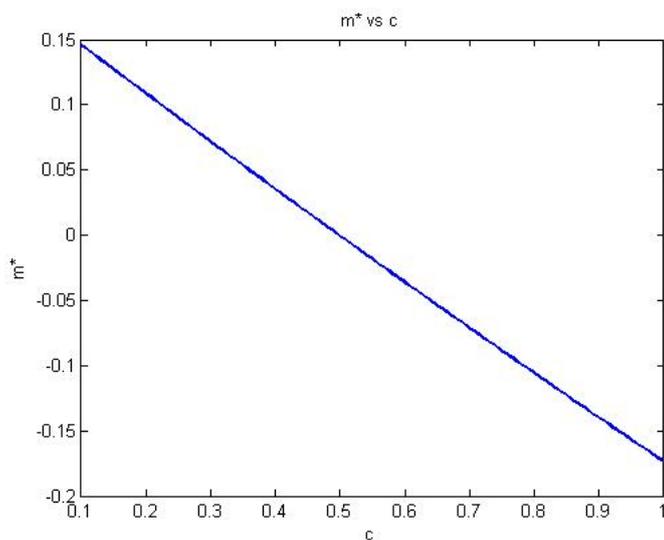


Figure 5.6: Linear relationship for m^* vs. c .

What we are observing when we decrease bound c is that our risk manager has set more punitive restraints and hence has become more risk adverse. Thus we see the sign change of m^* from negative to positive which then in turns means that capital must be set aside to cover expected losses and can not be used elsewhere.

The next attribute to investigate would be to see the effect of the systemic factor. In this case, we may set $\alpha = 1$. Moreover, we may add complexity by varying the correlation ρ . Note that to simulate correlated variates from our Monte Carlo routine, we simulate from the following covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \quad (5.45)$$

The results of this procedure are detailed in Table 5.35. To keep inline with the replication exercise, we limit ourselves to 2×10^6 simulations as was done in the paper. When comparing the case $\alpha = 0$ from Table 5.32 and $\alpha = 1$ with corresponding $\rho = 0$ from Table 5.35, we have values -0.1724 and -0.103 respectively. We interpret this to mean that with a systemic risk component, there is less slack in the allocation in the latter case. When we allow for interconnected risk through the α weight, we have different directional behaviour as we vary the correlation value. In this simple two sub-portfolio case, it is easy to see that if we have anti-correlated values (negative correlation) between two sub-portfolios, then we will have a lower capital requirement (more slack). However if we have positive correlation between for two sub-portfolios, then there is more capital is required. As we will see later on in an empirical example using operational risk data, we observe only positive correlation. Hence when we toggle between $\alpha = 0$ and $\alpha = 1$, we see an increase in capital requirements.

Moving on to a trivariate example, we return to the loss function (5.39) and write out the terms of the series when $d = 3$ and hence by doing so, show the equivalence of (5.21) and (5.22). We do this as it is not immediately obvious of the terms involved

ρ	m_1^*
-0.9	-0.167
-0.5	-0.143
-0.2	-0.120
0	-0.103
0.2	-0.085
0.5	-0.056
0.9	-0.012

Table 5.35: Bivariate case with $\alpha = 1$ using Monte Carlo with M =2 million.

in the first order equations. Hence the ability to replicate the values in Armenti *et al* (2015) [7] lays the foundation and allows for comparison for an extension. From (5.39) we have

$$\ell(x) = x_1 + x_2 + x_3 + \frac{1}{2}((x_1^+)^2 + (x_2^+)^2 + (x_3^+)^2) + \alpha(x_1^+x_2^+ + x_1^+x_3^+ + x_2^+x_3^+). \quad (5.46)$$

The claim is that this is equivalent to (5.22) and hence when $d = 3$ we have

$$\ell(x) = x_1 + x_2 + x_3 + \frac{1-\alpha}{2}((x_1^+)^2 + (x_2^+)^2 + (x_3^+)^2) + \frac{\alpha}{2}(x_1^+ + x_2^+ + x_3^+)^2. \quad (5.47)$$

Working with the right side of (5.47) we have

$$\begin{aligned} \ell(x) &= x_1 + x_2 + x_3 + \frac{1-\alpha}{2}((x_1^+)^2 + (x_2^+)^2 + (x_3^+)^2) + \frac{\alpha}{2}(x_1^+ + x_2^+ + x_3^+)^2 \\ &= x_1 + x_2 + x_3 + \frac{1}{2}((x_1^+)^2 + (x_2^+)^2 + (x_3^+)^2) - \frac{\alpha}{2}((x_1^+)^2 + (x_2^+)^2 + (x_3^+)^2) \\ &\quad + \frac{\alpha}{2}((x_1^+)^2 + (x_2^+)^2 + (x_3^+)^2 + 2x_1^+x_2^+ + 2x_1^+x_3^+ + 2x_2^+x_3^+) \\ &= x_1 + x_2 + x_3 + \frac{1}{2}((x_1^+)^2 + (x_2^+)^2 + (x_3^+)^2) + \alpha(x_1^+x_2^+ + x_1^+x_3^+ + x_2^+x_3^+). \end{aligned}$$

Having written down the loss function for the three portfolio case, the first order conditions follow from simple differentiation. This then allows for more interesting cases to be investigated. The study then migrates from the study of three risk factors where two sub-portfolios have the same risk profile, are correlated and compared

alongside a third stand-alone sub-portfolio. The portfolio losses still have mean zero however have the following variance-covariance matrix

$$\Sigma = \begin{bmatrix} 0.5 & 0.5\rho & 0 \\ 0.5\rho & 0.5 & 0 \\ 0 & 0 & 0.6 \end{bmatrix} \quad (5.48)$$

and ρ varied as before. The higher marginal risk as indicated by 0.6 compared to the other two sub-portfolios imparted interesting dynamics on capital allocations. When $\alpha = 0$, it is expected that the sub-portfolio should attract a higher capital requirement simply due to its increased volatility. However, when ρ is allowed to vary and $\alpha = 1$, the interconnectedness of the first two sub-portfolios are expected to attract a more extreme capital requirement based on the sign and magnitude of ρ . As before, we start with the case where $\alpha = 0$. The results are summarized in Table 5.36. The next case would be to set $\alpha = 1$ and let ρ vary. We see that despite the first

$m_1^* = m_2^*$	m_3^*
-0.165	-0.120

Table 5.36: Trivariate case with $\alpha = 0$ using Monte Carlo with M =2 million.

two sub-portfolios having lower variances, the result of their interdependency causes for increased capital requirements for certain values of ρ . Thus the interdependency relationship trumps the smaller individual variances. This is shown in Table 5.37.

Up until this point we do not claim any originality and the only value-add has been a sanity check on the replication of results. However having seen a simple build-up, we may nicely transition to an application of the methodology using an empirical data set. This would give a sense of the utility of computing multivariate shortfall optimal risk allocations in addition overlaying systemic risk. As well, we may offer an extension in the form of using various copula models. What we have covered thus far are normally-distributed P&Ls of sub-portfolios with linear dependence. However

ρ	$m_1^* = m_2^*$	m_3^*
-0.9	-0.190	0.094
-0.5	-0.133	0.018
-0.2	-0.100	-0.031
0	-0.076	-0.059
0.2	-0.055	-0.088
0.5	-0.023	-0.128
0.9	0.026	-0.173

Table 5.37: Trivariate case with $\alpha = 1$ using Monte Carlo with M =2 million.

without debate we constantly hear of non-normality of financial returns/performance and also of dependency structures falling apart in stress scenarios. Taking the concept of systemic risk from Armenti *et al* (2015) [7] and rather than building in fixed dependencies through α , we may uncover natural, inherent dependencies through the P&Ls themselves. This opens up the discussion on such copula models such as the Student t which models joint fat tail events. This could be viewed as a stress testing tool as well where capital allocation could be assessed under a range of conditions. A further extension may look at Archimedean copulas such as the asymmetric Gumbel copula which exhibits greater dependence in the positive tail rather than in the negative.

5.5 Application to Operational Risk Losses

An application of the previous methodology is transferrable to the management of operational risk as well. As mentioned before, modeling is conducted on a sub-portfolio basis which in this case may be viewed as an arbitrary cut of an operational loss database. Usually we are concerned with modeling on a UoM level which is at the intersection of a unique business line and unique event type. However, technically if we are interested on an aggregate basis, we could also collapse data cells and look at losses on an aggregate business line (any one of eight) or any event type (any one of seven). The caveat would obviously be that losses are not homogenous but

rather from a management perspective, either a certain line of business or event type is of greater interest than the contributing elements. Nonetheless, we will constrain ourselves to modeling on any one of the unique 56 UoMs.

The analogy is then as such. In the market risk space, an application arises in the context of a trading floor with a risk manager wanting to use the shortfall risk allocation optimization technique to determine optimal expected capital offsets based on the P&L distribution of a trading desk. In the credit risk space, a CCP is concerned with optimal margin requirements to ensure the solvency of the default fund of its contributing members. To apply this in the operational risk space, the lens for which we view operational risk losses would need to be bifurcated. A common segregation of an operational risk loss database for initial exploratory data analysis is the split between cells which are either low frequency or high severity. For example, credit card fraud which occurs on a daily time horizon is easily seen as a high frequency loss. One may even say losses are to be expected. In Canada, the Canadian Bankers Association provides credit card fraud and Interac debit card statistics which confirms the high frequency of losses. Others cells however experience infrequent losses. For example, any event type in the agency services business line which is akin to the corporate trust or custodial duties tend to be low frequency events. That is, the business where a bank acts has as a trust or in the lending of securities experiences low frequency losses. Rather than base this purely on speculation, we may confirm this by turning to the SAS OpRisk Global Data.

We make note that the SAS OpRisk Global Database contains losses in financial banking, insurance and even non-financial businesses in 20 such areas from accommodation and foodservices, agriculture, construction, health care, real estate rental and leasing, to utilities and wholesale trade. The data records are publicly reported and collected over a threshold of USD \$100,000. The losses are also provided in a separate column with an annual adjustment based on the consumer price index (CPI)

obtained from the US Department of Labor - Bureau of Labor Statistics. Hence losses are brought up to their current loss value. Since we are interested in a Basel prescribed 56 cell structure, we filter and report this level of granularity. We summarize the number of losses in Table 5.38. We have at our disposable record level data and have only disclosed high level descriptions of the data. We now build the case that

BL/ET	IF	EF	EPWS	CPBP	DPA	BDSF	EDPM	Total
CF	50	24	38	540	0	0	8	660
TS	177	25	41	902	3	10	197	1,355
RB	1,867	1,946	103	1,493	60	35	169	5,673
CB	610	995	31	498	20	5	62	2,221
PS	31	26	1	142	1	15	46	262
AS	28	37	0	103	1	0	20	189
RBR	275	45	63	828	1	2	137	1,351
AM	129	50	13	484	1	2	41	720
Total	3167	3,148	290	4,990	87	69	680	12,431

Table 5.38: Number of losses in SAS OpRisk Global Database up to January 2016 release date for Basel 56 cell structure.

BL/ET	IF	EF	EPWS	CPBP	DPA	BDSF	EDPM	Total (\$mill)
CF	3,328	3,447	2,366	101,155	0	0	214	110,510
TS	30,808	5,235	627	170,230	56	1,227	8,412	216,594
RB	39,402	19,851	3,000	251,670	1,327	1,012	1,703	317,965
CB	81,472	25,877	416	89,119	3,951	336	803	201,975
PS	761	469	1	20,589	2	400	458	22,680
AS	683	5,012	0	12,643	2	0	1,170	19,510
RBR	4,154	264	2,310	20,707	4	3	278	27,719
AM	6,169	18,472	196	66,532	97	3	1,133	92,602
Total (\$mill)	166,776	78,627	8,916	732,645	5,439	2,980	14,171	1,009,555

Table 5.39: Aggregate loss amounts (USD million) CPI adjusted in SAS OpRisk Global Database up to January 2016 release date for Basel 56 cell structure.

aside from regulatory capital purposes which is the use of the VaR risk measure at

the 1 year 99.9 percentile which is usually reported to regulators and for public disclosure on a quarterly basis, there is value from a risk management perspective to get a handle on losses that occur at high frequency and hence could benefit from dynamic risk oversight. Hence moving from a VaR (high quantile risk measure over a one year horizon) and focusing on the expectation brings the focus to a more central and active form of risk management. Fine tuning the amount of capital required to offset expected losses may serve as a way to optimize capital. Moreover, it is possible to analyze in-depth new marketing campaign strategies in retail banking or trading and sales strategies from an operational risk loss perspective. From a credit card perspective, perhaps a new offering may have an anticipated increased business exposure but coincide with a higher amount of internal or external fraud. In that sense, having a feedback mechanism on the expectation of losses may be beneficial. Also from a market risk perspective, we do see an interesting take on operational risk of option hedging as covered by Mitra (2013) [91]. In this context, we see a model injecting an operational cost in rebalancing a replicating portfolio for option hedging. A proxy is used whereby operational costs reflect the scale of operational activity and risks. For instance, as the number of shares traded increase, this would increase operational costs (settlement costs) which in turn would be associated with increased operational activities (e.g. checking trades have been executed) and ultimately increased operational risks (e.g. failed trades). For if a trading strategy calls for an increased volume of trades, perhaps then activities involved such as data entry, checking new market data, accounting reconciliation all increase and may pose more operational risk.

Thus our application of the allocation method will focus on high frequency loss cells. Instead of using the entire database, we use a subset of the data. We affirm this assumption by pointing out that Badescu *et al* (2015) [13] has addressed a different problem but still used high frequency loss data. There we learn the authors test a correlated frequency model which requires sufficient data for model fitting. They use four Basel business lines (Trading and Sales, Commercial Banking, Retail Brokerage

BL/ET	EDPM	EF
TS	12,738	546
CS	4,620	3,501
CB	1,176	9,433
RBR	2,494	342

Table 5.40: Anonymous North American Bank operational loss count data from April 2007 to March 2012 from Badescu *et al* (2015).

and specifically Card Services, which is an sub category level 2 of Retail Banking) across two event types (Execution, Delivery and Process Management and External Fraud). The count of data points from the paper are summarized in Table 5.40. The data was collected by a North American financial institution from April 2007 until March 2012, including recognition data and net loss. The recording threshold was USD \$30,000. For confidentiality, the authours modified the original data by resampling from it while still keeping the same sample size. By doing so the main characteristics of the original data were preserved. From Table 5.40, it is apparent that internal loss data can be very rich in its loss experience when compared to vendor external loss data sets which are discovered through the public domain. Also, this supports our analysis by only using a subset of the available operational risk data in order to test out next methodologies. Comparing Table 5.40 with our available data set in Table 5.38, we see that we may also investigate four business lines: TS, RB (Retail Banking which will serve as a proxy for Card Services), CB and RBR. However, we adapt our event types to IF and CPBP as there are more prevalent losses. Hence our UoM structure is detailed in Table 5.41 with the frequency count shown. A numeric itemizing of each UoM is detailed in Table 5.42. We make note that we are not comparing methodologies in any sense but rather simply building the argument to work with a subset of operational loss data and provide us with a sanity check on the types of business lines that undergo frequent losses.

Having narrowed down the cells to focus on, the next crucial step would be to se-

BL/ET	IF	CPBP
TS	177	902
RB	1,867	1493
CB	610	498
RBR	275	828

Table 5.41: High frequency operational loss count data from March 1971 - December 2015 from SAS OpRisk Global Data.

BL/ET	UoM
TS/IF	1
RB/IF	2
CB/IF	3
RBR/IF	4
TS/CPBP	5
RB/CPBP	6
CB/CPBP	7
RBR/CPBP	8

Table 5.42: Mapping BL/ET to UoM sub-portfolios.

lect relevant frequency and severity distribution for which to model the compound loss distribution. For example, Badescu *et al* (2015) [13] in their method model frequency via a multivariate mixed Poisson distribution and assume that severities are mixed Erlang distributed which is a two parameter continuous probability distribution. Goodness of fit was examined by virtue of histograms, probability-probability plots (P-P), quantile-quantile plots (Q-Q plots), empirical vs fitted moments. Other forms of empirical operational risk research on different topics see simple formulations of frequency and severity modeled across an aggregate UoM using Poisson and Lognormal distributions such as in Jiménez-Rodríguez *et al* (2011) [78]. Therein, goodness of fit was assessed purely on the Kolmogorov Smirnov body test. Leveraging off the analysis by De Jongh *et al* (2015) [50], they too use the SAS OpRisk Global database. Over a 23 year horizon, 53 UoMs were tested for best-fit parametric representation. Poisson was used to model frequency and choices for severity covered Burr, Exponential, Gamma, Pareto, Lognormal, Weibull, GPD and Inverse Gaussian.

Each of these distributions were fitted conditional on the left truncation of the data at USD \$100,000. Best-fitting distributions were ranked by three goodness-of-fit test statistics, namely Anderson-Darling (AD), Kolmogorov-Smirnov (KS) and Cramér-von Mises (CVM) in each UoM. The results of the testing showed that the Burr distribution was the best choice overall obtaining a top three position in 83% (KS), 96% (AD) and 92%(CM) among all 53 UoMs. The closest second place contender of the distributions considered was the Lognormal. The three parameter Burr type XII distribution function is given as

$$F_B(x; \eta, \tau, \alpha) = 1 - \left(1 + \left(\frac{x}{\eta}\right)^\tau\right)^{-\alpha} \quad \text{for } x > 0. \quad (5.49)$$

The parameters have the requirement $\eta > 0, \tau > 0, \alpha > 0$ where η is a scale parameter and τ and α are shape parameters. Hence, we adopt the same parametric representations across our eight cell framework but recalibrate new Poisson and Burr distributions.

We too use a 23 year time horizon but starting from January 1993 - December 2015 inclusive. Since our focus is not on an annual time horizon, we select to use the most granular time horizon and calibrate a Poisson process over a monthly time horizon. That is, our intensity parameter represents the mean number of losses to be expected in a given month. Independent of the time component, we fit Burr continuous probability distributions to the size of the losses. For both frequency and severity, we follow De Jongh *et al* (2015) [50] and fit truncated versions of the distributions. As the details were not provided, we referred to Cruz *et al* (2015) [47] to make this adjustment.

We learn in practice that often the missing data below the threshold are completely ignored. This could potentially lead to an underestimation or overestimation of capital. Since we have the luxury of an empirical data set and base these results on it, we should account for this adjustment. Take for instance losses originating from severity

with PDF $f(x|\beta)$ where β are the calibrated parameters of the severity density and frequency with PMF $p(n|\lambda)$ with single parameter λ . If losses come from a data set above a known reporting truncation level L , then the density of the losses above L is a left-truncated density

$$f_L(x|\beta) = \frac{f(x|\beta)}{1 - F(L|\beta)} \quad L \leq x < \infty. \quad (5.50)$$

The events of the losses above L follow the so-called *thinned* Poisson process with intensity

$$\theta(\gamma, L) = \lambda(1 - F(L|\beta)) \quad (5.51)$$

and hence the number of events above the threshold is distributed from $\text{Poisson}(\theta)$. As there is little discussion on how to perform this method, we point out that the method of importance sampling may be used to simulate from the modified distributions.

Upon making this adjustment, we are able to simulate from our truncated distributions. Unlike a trading desk that has a P&L everyday, an operational loss database may not have a loss in any given day/month - even if it is a high frequency cell. This then leads us to still consider the compound loss distribution of frequency and severity but we may alter the time horizon so that we are looking at monthly losses instead of annual. Then we may apply Theorem 5.3.4 where X calibrated is the calculated compound loss distribution for each UoM from our structure in Table 5.41.

The final requirement would be to place a realistic bound c in Theorem 5.3.4. For starters, we may play the role of the risk manager and allow for a loss of 10^9 in each of the eight UoMs. Then for conservatism, double it to arrive have a threshold of $c = 16 \times 10^9$. Later on we will look at sensitivities to the the threshold value to yield insight into an appropriate amount depending on the purpose. Given that we will still use the convex loss function (5.39) which penalizes losses, this may conservative

from a risk management standpoint. At this point, we are in a position to determine the minimum amount of capital required (m^*) for each sub-portfolio such that the expected value of the monthly aggregate loss under the penalty function ℓ is bounded by c dollars. We utilize 200,000 Monte Carlo simulations as the higher dimensions adds much more computation time. Before we show the numerical simulation results, we provide a road map of the type of sensitivities that may be of interest to a risk manager.

1. Systemic risk weight $\alpha = 0, c = 16 \times 10^9$;
 2. $\alpha = 0, c$ varies;
 3. $\alpha = 1, c = 16 \times 10^9$, use empirical correlation matrix from loss simulation;
 4. $\alpha = 1, c$ varies, use empirical correlation matrix from loss simulation;
 5. Repeat 2. but look backwards at how the allocation changes at three lag intervals. Note that the probability distributions and correlation matrix may only be calibrated on data that is available at the quarter end.
1. With no systemic risk component, each sub-portfolio attracts a capital requirement based on its individual risk profile. This is summarized in Table 5.43. We see the

UoM_i	m_i^* in \$million
1	117.5
2	401.9
3	373.2
4	-0.3
5	610.9
6	591.6
7	324.4
8	255.9
Total	2,675.1

Table 5.43: Systemic risk $\alpha = 0$, bound $c = \$16$ billion.

effect of the loss function creating a gap between the aggregate capital requirement and that of the bound. Since we are not simply concerned with expected loss, but expected losses that are penalized by ℓ , this further contributes to the gap. However,

what this does do is give the risk manager some ability to determine what bounds may be better suited to their tolerance. What we also observe is that there is slack in UoM_4 in Table 5.43. That is, there is a capital benefit that may be carried over or spread across to other portfolios. Up to this point, we have proceeded by hypothetical cases (Standard Normal portfolios) to showcase the utility of the algorithm. Depending on the context, we may or may not permit a UoM to observe a negative allocation. From Panjer (2002) [100], we also see a method in the allocation of solvency capital in multi-line financial businesses. The TailVaR risk measure is extended to allocate capital to each business unit. Using real company data, two out of ten business lines observe a negative capital allocation using TailVaR-based proportional allocation. Alternatively, we may envision the case where reporting to senior management or an independent regulator may not permit negative allocation despite the aggregate sum still being positive. This may be principle-based and hence we may also accommodate this requirement in our algorithm. That is, we set a lower bound constraint on all UoMs that they must be non-negative. We may then re-run our algorithm and produce results accordingly. This happens in two instances and we provide a comparison. The first comparison is seen in Table 5.44. Hence comparing Table 5.43 and Table 5.44, we see an aggregate increase of 0.4 or \$400,000 (rounded). Also what was once optimal for each UoM in one construct, has shifted allocations in various UoMs. This is again expected under the objective to minimize the sum of allocations.

2. What we notice from Table 5.45 is that by successively lower the bound c , we are required to set aside more and more capital for each sub-portfolio. What we also notice is that specifically for UoM_4 , the existing slack disappears in the case when the bound is halved. This is useful if we want to meet the requirement of (5.34). What is also of importance is to notice that when our initial bound is multiplied by 10, UoM_5 and UoM_6 still require a positive amount of capital. This makes intuitive sense in that the Trading and Sales business line in the Clients Products & Business Practices event type attract heavy fines and lawsuits. Such notable examples would

UoM_i	m_i^* in \$million
1	125.6
2	398.1
3	370.4
4	2.0
5	608.5
6	589.5
7	323.9
8	257.5
Total	2,675.5

Table 5.44: Systemic risk $\alpha = 0$, bound $c = \$16$ billion, non-negative allocation.

UoM_i	m_i^* $c = \$2B$	m_i^* $c = \$4B$	m_i^* $c = \$8B$	m_i^* $c = \$16B$	m_i^* $c = \$160B$
1	354.0	302.9	223.0	117.5	-52.4
2	600.6	562.3	498.6	401.9	-31.5
3	581.9	541.2	474.2	373.2	-34.6
4	90.3	57.3	21.5	-0.3	-57.5
5	752.0	726.2	681.7	610.9	3.5
6	737.7	710.5	664.2	591.6	4.9
7	540.0	496.9	426.6	324.4	-37.9
8	488.4	441.3	365.7	255.9	-44.1
Total	4,144.9	3,838.6	3,355.5	2,675.1	-249.6

Table 5.45: Systemic risk $\alpha = 0$, varying bound c .

be lawsuits related to mortgage back securities and market manipulation. On the Retail Banking side within the same event type, we see fines and lawsuits settled in relation to tax evasions and failure of banks to keep their customer well informed on particulars of retail products.

Again for completeness, we may enforce a non-negativity requirement for an allocation scheme. Since we observe several negative entries when we increase the threshold c to 10 times the baseline value, we re-run Table 5.45 and get Table 5.46 and replace the last two columns. As expected, since there is so much slack in the allocation corre-

UoM_i	m_i^* $c = \$2B$	m_i^* $c = \$4B$	m_i^* $c = \$8B$	m_i^* $c = \$16B$	m_i^* $c = \$160B$
1	354.0	302.9	223.0	125.6	0
2	600.6	562.3	498.6	398.1	0
3	581.9	541.2	474.2	370.4	0
4	90.3	57.3	21.5	2.0	0
5	752.0	726.2	681.7	608.5	0
6	737.7	710.5	664.2	589.5	0
7	540.0	496.9	426.6	323.9	0
8	488.4	441.3	365.7	257.5	0
Total	4,144.9	3,838.6	3,355.5	2,675.5	0

Table 5.46: Systemic risk $\alpha = 0$, varying bound c , non-negative allocation.

sponding to the last column, we can not realistically put aside capital in this instance.

3. Taking the analysis a step further, we may add in the systemic risk component and see the effect on capital allocation. We recall back from Section 2.4.6 the possible ways to analyze correlations according to Frachot *et al* (2004) [71]. When considering two cells with annual frequencies N_1 and N_2 that are correlated, they may not be considered independent variables. It is reasoned that both frequencies N_1 and N_2 share common dependence with respect to some variables such as gross income, economic cycle, size of business etc. As suggested in Frachot *et al* (2004) [71], empirical frequency correlation may be evidenced by computing historical correlation between past frequencies of operational loss events if a long time series is available. Hence we adopt this approach as well.

We avoid trying to capture severity correlation that would imply a loss X in one cell is high (low) when a loss Y in another cell is also high (low). Hence writing this in a compact way, we mimic Frachot *et al* (2004) [71] and suppose that aggregate loss correlation (between losses L_1 and L_2) is fundamentally conveyed by the underlying

correlation between frequencies

$$\left. \begin{array}{l} \text{corr}(N_1, N_2) \neq 0 \\ \text{corr}(X, Y) = 0 \end{array} \right\} \implies \text{corr}(L_1, L_2) \neq 0. \quad (5.52)$$

It is reasoned this is a simple way to add correlation between aggregate losses, that is the correlation is conveyed through the compounded process by virtue of the frequency correlation. However in Frachot *et al* (2004) [71], it is also stated that $\text{corr}(L_1, L_2) \leq \text{corr}(N_1, N_2)$. This indicates that a strong frequency correlation may be mellowed-out under convolution with the severity distribution and hence would manifest itself in lower loss correlation. Again, this ability to investigate frequency correlation on the impending loss correlation is unique to operational risk loss modeling and is different from the market/credit risk application where we directly model correlation at the P&L level.

We provide the pairwise linear correlation coefficient matrix resulting from the empirical data based on frequency counts. Since we are concerned with a monthly time horizon, we focus on the frequency monthly correlation matrix in Table 5.47. The extraction of the correlation matrix is an attempt to investigate the existence of any significant linear correlation. The matrix of p -values tests the hypothesis of no correlation against the alternative that there is a nonzero correlation. Each element in the matrix of p -values corresponds to an element of ρ . If $p(i, j)$ is small, in this case less than 0.05, then $\rho(i, j)$ is significantly different from zero. We produce the associated p -value matrix in Table 5.48 and the significant pairs in Table 5.49. Although we did not focus on an annual time horizon, we provide the annual correlation matrix for completeness. The matrix of annual frequency correlations, p -values, and statistical significant correlations are located in Tables 5.50-5.52. We note that in the monthly matrix, we calibrated the values based on $23 \times 12 = 276$ realizations whereas in the annual matrix, we just had 23 values. We do see a different matrix of values between monthly and annual count of losses.

1	0.37	0.35	0.04	0.08	0.16	0.08	0.12
0.37	1	0.49	0.26	0.41	0.53	0.24	0.28
0.35	0.49	1	0.17	0.29	0.32	0.27	0.17
0.04	0.26	0.17	1	0.11	0.11	0.07	0.12
0.08	0.41	0.29	0.11	1	0.39	0.12	0.16
0.16	0.53	0.32	0.11	0.39	1	0.1	0.31
0.08	0.24	0.27	0.07	0.12	0.1	1	0.08
0.12	0.28	0.17	0.12	0.16	0.31	0.08	1

Table 5.47: Monthly frequency correlation matrix.

1	0.01	0.01	0.61	0.24	0.01	0.2	0.06
0.01	1	0.01	0.01	0.01	0.01	0.01	0.01
0.01	0.01	1	0.01	0.01	0.01	0.01	0.01
0.61	0.01	0.01	1	0.09	0.08	0.27	0.07
0.24	0.01	0.01	0.09	1	0.01	0.07	0.02
0.01	0.01	0.01	0.08	0.01	1	0.14	0.01
0.2	0.01	0.01	0.27	0.07	0.14	1	0.19
0.06	0.01	0.01	0.07	0.02	0.01	0.19	1

Table 5.48: Matrix of p -values (two-tailed, significance = 0.05) corresponding to monthly correlation matrix.

	0.37	0.35			0.16		
0.37		0.49	0.26	0.41	0.53	0.24	0.28
0.35	0.49		0.17	0.29	0.32	0.27	0.17
	0.26	0.17					
	0.41	0.29			0.39		0.16
0.16	0.53	0.32		0.39			0.31
	0.24	0.27					
	0.28	0.17		0.16	0.31		

Table 5.49: Monthly frequency correlation matrix filtering for statistical significance with $p < 0.05$.

1	0.76	0.83	0.25	0.38	0.44	0.66	0.32
0.76	1	0.85	0.54	0.78	0.82	0.5	0.54
0.83	0.85	1	0.57	0.68	0.68	0.6	0.6
0.25	0.54	0.57	1	0.66	0.6	0.24	0.53
0.38	0.78	0.68	0.66	1	0.8	0.26	0.68
0.44	0.82	0.68	0.6	0.8	1	0.36	0.79
0.66	0.5	0.6	0.24	0.26	0.36	1	0.47
0.32	0.54	0.6	0.53	0.68	0.79	0.47	1

Table 5.50: Annual frequency correlation matrix.

1	0.01	0.01	0.26	0.09	0.04	0.01	0.15
0.01	1	0.01	0.01	0.01	0.01	0.02	0.01
0.01	0.01	1	0.01	0.01	0.01	0.01	0.01
0.26	0.01	0.01	1	0.01	0.01	0.29	0.02
0.09	0.01	0.01	0.01	1	0.01	0.24	0.01
0.04	0.01	0.01	0.01	0.01	1	0.1	0.01
0.01	0.02	0.01	0.29	0.24	0.1	1	0.03
0.15	0.01	0.01	0.02	0.01	0.01	0.03	1

Table 5.51: Matrix of p -values (two-tailed, significance = 0.05) corresponding to annual correlation matrix.

	0.76	0.83			0.44	0.66	
0.76		0.85	0.54	0.78	0.82	0.5	0.54
0.83	0.85		0.57	0.68	0.68	0.6	0.6
	0.54	0.57		0.66	0.6		0.53
	0.78	0.68	0.66		0.8		0.68
0.44	0.82	0.68	0.6	0.8			0.79
0.66	0.5	0.6					0.47
	0.54	0.6	0.53	0.68	0.79	0.47	

Table 5.52: Annual frequency correlation matrix filtering for statistical significance with $p < 0.05$.

Referring back to Table 5.49 we notice several statistically significant correlated cells. In both the monthly and annual matrices, business lines 2 and 3 (Retail Banking and Corporate Banking) with event type 1 (Internal Fraud) show positive correlation across all cells. It could be reason that these are large divisions with far reach throughout an organization that could be associated the an increase in losses elsewhere. Retail Banking encompasses (i) retail banking (retail lending and deposits, banking services, trust and estates), (ii) private banking (private lending and deposits, banking services, trust and estates, investment advice) and (iii) card services (merchant/commercial/corporate cards, private labels and retail). Commercial Banking encompasses project finance, real estate, export finance, trade finance, factoring, leasing, lending, guarantees, and bills of exchange. Internal Fraud also covers a wide range of activities that arises from losses intended to defraud, misappropriate property, circumvent regulations or company policy which involves at least one internal party. It may be realistic to expect directional behaviour with Trading and Sales in that the business line covers everything from sales, market making and treasury businesses. Also, Retail Brokerage accounts for both execution and full service capabilities which is again wide-spread throughout the organization. The Clients, Products and Business Practices covers all aspects of losses arising from failure to meet a professional obligation to specific clients or from product flaws. Hence the two events overlap in various forms of unauthorized activities.

Hence since there are many forms of positive correlation present, we should account for this in our frequency simulation. We provide the output due to systemic risk between portfolios in Table 5.53. When comparing Table 5.43 with $\alpha = 0$ and Table 5.53 with $\alpha = 1$, we see a monotonic increase of capital required for all sub-portfolios. This ordering holds in that there are no negative correlations present and only positive correlations. This shows that introducing an interaction between sub-portfolios is a relevant exercise. Although we have introduced this relationship through α and

UoM_i	m_i^* in \$million
1	139.6
2	427.3
3	392.8
4	23.1
5	644.4
6	625.4
7	350.5
8	278.0
Total	2,881.1

Table 5.53: Systemic risk $\alpha = 1$, bound $c = \$16$ billion.

UoM_i	m_i^* $c = \$2B$	m_i^* $c = \$4B$	m_i^* $c = \$8B$	m_i^* $c = \$16B$	m_i^* $c = \$160B$
1	356.9	306.4	231.0	139.6	0.8
2	614.4	577.3	516.3	427.3	67.6
3	580.1	542.4	482.2	392.8	61.3
4	91.6	66.4	40.7	23.1	1.8
5	762.4	740.9	703.3	644.4	147.2
6	752.8	728.3	688.2	625.4	155.6
7	549.3	509.7	442.9	350.5	49.9
8	490.6	445.2	375.1	278.0	23.5
Total	4,198.1	3,916.6	3,479.7	2,881.1	507.7

Table 5.54: Systemic risk $\alpha = 1$, varying bound c .

a full weight of 1, the different combinations of interacting sub-portfolios introduced through the last term in the loss function in equation (5.21) should give some sense of the connectedness of the network and be informative to a risk manager.

4. We may also repeat the same exercise by varying the bound c but keeping $\alpha = 1$. This is located in Table 5.54. As expected, we see a monotonically decreasing requirement as the threshold is expanded for increasing c . Surprisingly, even with a bound of \$160B, we still require capital to be set aside to offset losses and hence there is no slack in the allocation.

UoM_i	$m_i^*(Q4)$	$m_i^*(Q3)$	$m_i^*(Q2)$	$m_i^*(Q1)$
1	139.6	138.0	151.7	142.9
2	427.3	428.6	430.0	437.6
3	392.8	390.9	391.8	390.8
4	23.1	24.2	22.4	23.2
5	644.4	642.0	637.0	612.8
6	625.4	627.9	619.1	616.2
7	350.5	349.9	349.1	346.7
8	278.0	283.7	261.2	249.2
Total	2,881.1	2,885.2	2,862.3	2,819.4

Table 5.55: Systemic risk $\alpha = 1$, bound $c = \$16$ billion, calibrations at quarter end.

5. As a final investigation, we may take the results of Table 5.53 and compare how the allocation changed from previous time horizons. That is, since we used 23 years of data from January 1993 - December 2015, we may see what sort of requirement we would have come up with having rolled back successive quarters (three month intervals rolling back three, six, nine, months prior to the reference date of December 2015). Although we did not have future information say at quarter end Q1, Q2, Q3 we do get a sense of how the allocation changes as new data is brought online up until the final calibration at Q4 end. To be explicit, we would calibrate probability distributions for simulation using all the data available to us at any given time. The purpose of conducting this exercise for high frequency losses would be to provide a dynamic response to the changing environment and potentially manage capital requirements more actively.

Looking at the results in Table 5.55, we see that the total capital requirement is tied in a range of \$65.8 million over 2015 which is approximately 2% of the total capital requirement. This is reassuring given that we would appreciate some certainty around the expected value of the losses for high frequency cells. Also this provides a monitoring mechanism and the frequent updating could serve well if new strategies

are launched and the output on losses examined following quarters.

5.6 Extension using Copulas

Before we continue on the journey to exploring the extension of the multivariate shortfall risk allocation optimization problem using copulas, we must ask ourselves why is this necessary, what do we gain from adding this modeling element, and also how do we implement such a method? Recapping from the start of the chapter, we first covered various risk measures that could be used to quantify the risk of some random process. Then being able calculate the riskiness of a position, we could aggregate the risk (which is a form of required capital) among all contributing members of a portfolio. However, we were then faced with the problem how to allocate capital requirements back to the contributing sub-portfolios in a fair manner. There, we saw five such methods to allocate back to sub-portfolios. We then framed the problem in terms of a shortfall problem whereby we were more concerned with the expectation of the loss profile and the capital allocation methodology used to allocate sufficient capital to sub-portfolios. To aid in the solution of the problem we used convex representations of the loss random variable in order to formulate an optimal minimum amount of capital to set aside. In order to greater emphasize the connectedness of the system, we added a scalar systemic term that controlled the way dependencies between losses among portfolios would contribute to aggregate capital.

However, an intermediate step was sort of glazed over in the pursuit of the goal - that of the dependencies on the return or losses themselves. From Aas (2004) [1] we learn that the linear correlation coefficient which is by far the most used measure to test dependence in the financial community is a measure of linear dependence only. This assumption may hold true if asset returns are well represented by an elliptical distribution such as the multivariate Gaussian or multivariate Student t probability distributions, then the dependence being linear makes sense. However, venturing

outside elliptical distributions, a linear correlation coefficient to measure dependency may lead to misleading conclusions. Thus the ability to decompose a model into the marginal distributions and the dependence structure may provide flexibility.

An example was given in Aas (2004) [1] of modeling the joint distribution of a stock market index and exchange rate. Both data sets could be fit reasonably well to separate univariate Student t distribution but fitting a bivariate Student t distribution had the restriction of imposing the same tail heaviness. Hence the natural conclusion to model with the bivariate Student t distribution was not reasonable. The use of a copula was seen as providing the solution to this modeling nuance. By virtue of this example, we note it was both necessary and had much to gain in terms of flexibility by still being able to fit marginal distributions but leave open the discussion of dependency through a copula. We intentionally focused on Monte Carlo implementation methods that allowed for such copula structures to be computed. For our implementation, the ability to compute estimators from the set of equations seen in (5.42) could be achieved when the underlying random variables are simulated from copulas.

We must also ask ourselves before we start introducing which copulas to investigate and performing simulations, does our optimization method work under such conditions? Referring back at Theorem 5.3.4, our focus is not to prove the extension rather to point out factors that support the extension. Most notably, differentiation is with respect to m_i and hence there is no requirement for differentiation with respect to any random portfolio X_i nor copula incorporating X_i . The original proof put restrictions on the loss function by virtue of Definition 5.3.5 and since we are using the same unbiased loss function, we have not introduced this uncertainty. To give a sense of the proof, it starts with looking at a set $C \subseteq \mathbb{R}^d$ and defining $m \mapsto \delta(m, C)$ as 0 if $m \in C$ and ∞ otherwise. The function $f(m) = \sum m_k + \delta(m, \mathcal{A}(X))$ (where $\mathcal{A}(X)$ is the set of monetary allocations) is increasing, convex, lower-semi-continuous, proper

and such that $R(X) = \inf\{f(m) : m \in \mathbb{R}^d\}$. In order to prove the existence of a risk allocation, it suffices to show that f is constant along its direction of recession. This is precisely where the unbiasedness of ℓ is used to show the existence of a risk allocation m^* . The uniqueness follows from the convexity property. Hence since we have the dual problem formalization, we focused on the smoothness property of f and hence our extension using copulas is applicable. Again for proof of the original theorem, we refer the reader to Armenti *et al* (2015) [7].

The natural question arises when using copulas is which one should you use? It has been stated in Embrechts (2009) [59] that this question has no obvious answer. Moreover, even Blum *et al* (2002) [32] draws the parallel that selecting the correct copula is as difficult as estimating the joint distribution F in the first place. In a multi-dimension framework, Aas (2004) [1] does cover some formal and informal goodness of fit tests for copula selection. However what Aas (2004) [1] offers and in essence forms the outline for our investigation is the introduction of two parametric families of copulas: the copulas of normal mixture distributions and Archimedean copulas. What we will see is that if we want to implement a simple model with linear dependence, then the Gaussian copula is sufficient. To emphasize extreme co-movements, then one should use a Student t copula. If we envision asymmetries in returns, then Archimedean copulas can be used. The Clayton copula shows greater dependence in the negative tail and the Gumbel shows greater dependence in the positive tail.

We turn to Aas (2004) [1] for a necessary introduction for the copulas that we will investigate. However, we believe the Student t and Gumbel copula are the relevant choices in our shortfall risk allocation simulation study and just highlight the Clayton copula for comparison purposes. The reason being, when we draw realizations from our CDFs, we think in terms of simulating losses where a negative value denotes a profit. Hence with this convention, we are more concerned with positive tail dependence which coincides with joint losses.

We introduced copulas in Section 2.4.6 in a generic setting. From McNeil *et al* (2005) [87], we learn that copulas may be divided into three categories: fundamental, implicit and explicit. Fundamental copulas represent a number of important special dependence structures. We will not cover these but such examples are independence, comonotonicity, and countermonotonicity copulas. Implicit copulas are extracted from well-known multivariate distributions using Sklar's Theorem but do not necessarily possess simple closed-form expressions. We recall that Sklar's Theorem states that for any joint distribution function, there is a unique copula that satisfies (2.35). Explicit copulas have simple closed-form expressions and follow general mathematical constructions known to yield copulas. Hence we will study the implicit Student t copula and the explicit Clayton and Gumbel copulas. We also introduce the implicit Gaussian copula just for reference.

In a two dimensional setting, we define the following copulas.

Definition 5.6.1. (*Gaussian copula*) *The Gaussian copula (central) is given by*

$$C_\rho = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}} dx dy \quad (5.53)$$

where ρ is the correlation parameter and $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution function.

Relative to the Gaussian copula, the Student t copula allows for joint fat tail events and hence a greater number of joint extreme events.

Definition 5.6.2. (*Student t copula*) The central Student t cumulative distribution function is given as

$$t_\nu(x) = \int_{-\infty}^x \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\frac{\nu}{2}} \left(1 + \frac{s^2}{\nu}\right)^{-\frac{\nu+1}{2}} ds, \quad (5.54)$$

where Γ is the gamma function, and ν represents the degrees of freedom. The bivariate distribution corresponding to t_ν is given by

$$t_{\rho,\nu}(x, y) = \int_{-\infty}^x \int_{-\infty}^y \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \left(1 + \frac{s^2 + t^2 - 2\rho st}{\nu(1-\rho)^2}\right)^{-\frac{\nu+2}{2}} ds dt. \quad (5.55)$$

Then the Student t copula is given as

$$C_{t_{\rho,\nu}}(u, v) = t_{\rho,\nu}(t_\nu^{-1}(u), t_\nu^{-1}(v)). \quad (5.56)$$

Increasing ν decreases the extreme co-movements and hence decreasing ν increases the realization of joint extreme events in both positive and negative directions.

Turning to explicit copulas, we achieve simple closed forms.

Definition 5.6.3. (*Clayton copula*) The Clayton copula is given by

$$C_\delta(u, v) = \frac{1}{(u^{-\delta} + v^{-\delta} - 1)^{\frac{1}{\delta}}} \quad (5.57)$$

where $0 < \delta < \infty$ is a parameter controlling dependency. Perfect dependence is achieved if $\delta \rightarrow \infty$ and independence is achieved if $\delta \rightarrow 0$. This asymmetric copula has greater dependence in the negative tail.

Definition 5.6.4. (*Gumbel copula*) The Gumbel copula is given by

$$C_\delta(u, v) = e^{-\left((-\ln u)^\delta + (-\ln v)^\delta\right)^{\frac{1}{\delta}}} \quad (5.58)$$

where $0 < \delta \leq 1$ is a parameter controlling dependency. Perfect dependence is achieved if $\delta \rightarrow 0$ and independence is achieved if $\delta = 1$. This asymmetric copula has greater dependence in the positive tail.

When using copulas, we have alternatives to Pearson linear correlation to summarize dependence measures. According to McNeil *et al* (2005) [87], rank correlations are simple scalar measures of dependence that depend only on the copula of a bivariate distribution and not on the marginal distributions, unlike linear correlation, which depends on both. When working with data, the standard empirical estimators of rank correlation may be calculated by looking at the ranks of the data alone, hence the name. That is, we only need to know the ordering of the sample for each variable of interest and not the actual numerical values. We will exploit this characteristic when looking at rank correlations in order to calibrate copulas to empirical data. There are two types of rank correlations: Kendall's tau and Spearman's rho for which we will focus on the former.

Kendall's tau rank correlations can be understood as a measure of concordance. That is if we have a set of observations $(x_1, y_1), (x_2, y_2) \dots, (x_n, y_n)$ of joint random variables X and Y , then any pair of observations (x_i, y_i) and (x_j, y_j) for $i \neq j$ are said to be concordant if the ranks for both elements agree: that is, if both $x_i > x_j$ and $y_i > y_j$ or if both $x_i < x_j$ and $y_i < y_j$. They are said to be discordant, if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant nor discordant. Thus Kendall's tau rank correlation is given by

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\frac{1}{2}n(n-1)}. \quad (5.59)$$

As we will see, since the copula of a multivariate distribution characterizes its dependence structure, it would be logical to use measures of dependence that are copula-based. From McNeil *et al* (2005) [87], we recall the following.

Proposition 5.6.1. *Suppose X and Y have continuous marginal distributions and a unique copula C . Then Kendall's tau rank correlations is given by*

$$\rho_\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1. \quad (5.60)$$

Now for Gaussian and Student t copulas, the relationship between the linear coefficient and Kendall's tau is given by

$$\text{corr}(X, Y) = \sin\left(\frac{\pi}{2}\rho_\tau\right), \quad (5.61)$$

where $\text{corr}(X, Y)$ is the linear correlation coefficient. For the Clayton copula, if we want to associate the single copula parameter from (5.57) with the Kendall tau parameter, we have

$$\rho_\tau(X, Y) = \frac{\delta}{\delta + 2}. \quad (5.62)$$

For the Gumbel copula, if we want to associate the single copula parameter from (5.58) with the Kendall's tau parameter, we have

$$\rho_\tau(X, Y) = 1 - \frac{1}{\delta}. \quad (5.63)$$

Now here in lies the benefit of (5.61) - (5.63); we have all the necessary ingredients to calibrate our copulas. Using empirical data, we may infer the correlation coefficient and also produce the necessary degrees of freedom for the implicit copulas. If we would rather focus on rank correlation, we may do that as well. In that case, we have necessary relationships to calibrate the single parameter for the explicit copulas.

In Section 5.4, we covered a case from Armenti *et al* (2015) [7] where we looked at two sub-portfolios each normally distributed with mean zero and covariance matrix stated in (5.45). We allowed ρ to vary between -0.9 to 0.9 and kept the systemic weight at one. Since we believe copulas help paint a different dependence picture, we

choose to analyze various copulas alongside this baseline. We do this to not model the best fitting dependence structure, but rather show directionally if using different copulas directionally the anticipated output. Under the various copula schemes, we compute optimal m^* and compare and contrast.

What we propose to do is start by simulating M random P&Ls from $X = (X_1, X_2) \sim N(0, \Sigma)$ for a fixed diagonal ρ . That is, we may use the simulated data as given and from it calibrate all the required parameters necessary to calibrate our copulas and solve our optimization problem. Although we know precisely what the dependence structure is, we may choose to investigate the situations where we believe there to be more joint fat tailed events or even asymmetric events. Why - one reason may be that this is a method of stress testing. That is, if we believe that there is an onset of macroeconomic stress, then we may want to stress the underlying dependence structure. What we should be able to do is rank the baseline, Student t , Clayton and Gumbel outputs based on increased capital requirements. Alternatively, we could have simulated from any one copula first and used it as a baseline to rank other outputs. We also point out that for the Student t copula, we can infer a suitable degree of freedom parameter from the data itself. However, we may also override this parameter and set the value at a conservative value of $\nu = 3$. This is a common practice when imposing a conservative dependence structure.

Our copula extension takes Table 5.35 and adds various copula dependency structures for comparison. This is summarized in Table 5.56. We may analyze Table 5.56 by going down each column. We notice that using the Student t copula with a high negative or positive correlation parameter is close to the baseline value. This is intuitive in that we expect both portfolios to display this behaviour if they are highly correlated/anti-correlated. The departure occurs when there is low correlation and the Student t copula forces a higher joint extreme movement. Hence we see smaller m_1^* near the middle correlation values relative to the baseline values. We interpret

ρ	ρ_τ	m_1^*			
		Baseline(Gaussian)	Student t ($\nu = 3$)	Clayton	Gumbel
-0.9	-0.713	-0.167	-0.166	N/A ($\delta > 0$)	N/A ($\delta > 0$)
-0.5	-0.333	-0.143	-0.137	N/A ($\delta > 0$)	N/A ($\delta > 0$)
-0.2	-0.123	-0.120	-0.113	N/A ($\delta > 0$)	N/A ($\delta > 0$)
0	0	-0.103	-0.096	-0.104	-0.104
0.2	0.123	-0.085	-0.080	-0.094	-0.078
0.5	0.333	-0.056	-0.053	-0.075	-0.045
0.9	0.713	-0.012	-0.012	-0.032	-0.009

Table 5.56: Bivariate case with $\alpha = 1$, Monte Carlo with $M = 2$ million, various copula dependencies.

this to mean there is less slack in the allocation (smaller negative numbers). The case for $\rho = 0$ being lower under the Student t copula is because this was forced for $\nu = 3$ as we did not empirically calculate the inferred value from the data. For the Clayton copula, we have the restriction from (5.57) of the non-negativity of the parameter hence we could not fit the copula. Again, we see the expected trajectory in that we have more slack in the optimal allocation for all non-negative correlation cases relative to the baseline. Actually, the Clayton copula has the greatest slack compared to all cases for a given correlation case. This is expected since the Clayton copula is known to show greater dependency in the negative tail. That is because our loss function penalizes losses more in the positive tail. Again, the Gumbel has the same restriction from (5.58). Also, since we are simulating losses for our sub-portfolios and a positive value indicates a loss, this in turn means that the Gumbel copula exacerbates this penalty on losses and hence we have the least slack.

5.7 Summary

In this chapter we were motivated to explore the problem of disseminating an aggregate amount of capital fairly to sub-portfolios. This led to the exploration of a related problem of calculating shortfall risk whereby an optimization problem was

formulated. The convex minimization problem was positioned to determine the minimum amount of capital required to be applied towards individuals sub-portfolios distributed according to unique loss distributions such that the expected loss of the aggregate remained bounded. The same formulation was applied in an operational risk context to high frequency cells whereby the ability to manage capital dynamically would be beneficial. An extension was provided alongside the original framework of multivariate shortfall risk allocation with systemic risk but accounting for advanced dependency structures through copulas. We confirmed under stylized cases that the Student t and Gumbel copula produced less slack in an allocation network compared to the Gaussian and the Clayton copula.

6 Conclusion and Future Work

We have explored problems of relevance to the regulation of operational risk models used within financial institutions. We took an essay style approach for each topic whereby there was a natural flow from the first essay to the last. At each step, a new problem revealed itself through an introspective analysis of the issue at hand.

Our first essay started with the intent to dissect operational risk and focus on the area of scenario analysis. With little guidance on how to collect and incorporate such input into a loss distribution approach model, we proposed a new methodology of incorporating business foresight into the model. Using frequency and severity estimates obtained from business line experts, we proposed to build an aggregate loss distribution. We then turned to an area of signal processing and recycled the convolution operator as a method to blend two probability distributions. We adopted the same view and used the data-derived loss distribution to smooth out the loss distribution obtained from human judgement. The resulting convolved loss distribution then characterized a model incorporating both modeling elements which could then be used to set capital requirements.

Our second essay revisited the idea of scenarios and looked at the type of scenarios that would be derived if a catastrophic risk occurred at a bank. Rather than isolate the line of thinking to specific business experts working in a particular business line for which particular type of losses may occur, we sought to quantify risks that effect the bank at the enterprise level. Still drawing upon expert input, we divided

the quantification of certain parameters to different experts. One such example of an earthquake study was taken to show how disaster planning experts could provide input into the scenario design process. Thereafter, a generalized framework based on a flexible number of scenarios, intensities and loss profiles was formulated to outline a catastrophic scenario algorithm. The output was also shown to be useable in the convolution integration method seen previously. When viewed in aggregate terms, we allocated capital using the marginal contribution method to associate a portion of capital to each contributing UoM.

Our third essay applied an established framework of multivariate shortfall risk capital allocation with systemic risk in an operational risk setting. Using high frequency loss data, an analysis was undertaken to solve the minimal capital optimization problem among contributing UoMs. The framework was then extended in a generalized setting to include systemic risk on a secondary level using copulas. Joint fat-tailed events and asymmetries between sub-portfolios were used to see the effect on capital requirements.

The intent of each essay was to investigate a new area of research while leaving room for further analysis for the next generation. For the first essay on scenario integration, the method to obtain a smooth continuous distribution from responses is open to debate. Bucketing responses based on answers to probability of losses exceeding a certain thresholds is just a single textbook technique used to build a piece-wise smooth curve. We also see in the industry proprietary methods to map scenario responses to a suitable candidate of distributions based on a line of questioning. This line of research has been mentioned through the CIRANO research institute. By collecting the percentage of losses expected to occur relative to multiples of a median loss, a decision-tree could be built to filter out certain continuous candidate severity distributions and help determine a suitable fitting distribution. Also once a scenario curve is obtained, the coefficient used for convolving both distributions could be sub-

ject to much scrutiny. We feel that by using actual bank data and actual scenario responses for a particular bank, various types of averages/weighting functions could be explored to combine both sets of information.

In the second essay, a survey of various probabilistic models could be analyzed for various types of catastrophic events. We looked at one such example through California earthquakes. In addition, the ability to produce a loss curve based on expert judgement from essay one could be reused in this context as well.

Lastly, we covered a single type of convex loss function when various types could have been used. For the application to operational risk, again empirical internal loss data which is confidential and not publicly disclosed would help quantify relevant capital requirements for certain UoMs. We also looked at an extension of dependency in the form of copula models. We studied only a subset of possible dependency structures available for which more options could be investigated. We wish the next generation the best of luck in their research aspirations!

7 References

- [1] Aas, K. (2004). *Modelling the dependence structure of financial assets: A survey of four copulas*. Research report SAMBA/22/04, Norwegian Computing Centre. Retrieved from <https://www.nr.no/files/samba/bff/SAMBA2204.pdf>
- [2] Agostini, A., Talamo, P., & Vecchione, V. (2010). Combining operational loss data with expert opinions through advanced credibility theory. *The Journal of Operational Risk*, 5(1), 3-28.
- [3] Ames, M., Schuermann, T., & Scott, H. S. (2015). Bank Capital for Operational Risk: A Tale of Fragility and Instability. *Journal of Risk Management in Financial Institutions*, 8(3), 227-243.
- [4] Anders, U., & Sandstedt, M. (2003). An operational risk scorecard approach. *Risk*, 39-42.
- [5] Andersen, T. G., Bollerslev, T., Christoffersen, P., & Diebold, F. X. (2007). Practical volatility and correlation modeling for financial market risk management. *National Bureau of Economic Research*, 513-548.
- [6] Anderson, T. W., & Darling, D. A. (1952). Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. *The Annals of Mathematical Statistics*, 23(2), 193-212.
- [7] Armenti, Y., Crépey, S., Drapeau, S., & Papapantoleon, A. (2015). Multivariate Shortfall Risk Allocation and Systemic Risk. Retrieved from <http://arxiv.org/abs/1507.05351v2>

- [8] Aroda, P., Guergachi, A., & Huang, H. (2015). Application of the convolution operator for scenario integration with loss data in operational risk modeling. *Journal of Operational Risk*, 10(4), 23-44.
- [9] Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203-228.
- [10] Aue, F., & Kalkbrener, M. (2006). LDA at work: Deutsche Bank's approach to quantifying operational risk. *Journal of Operational Risk*, 1(4), 49-93.
- [11] Aumann, R. J. & Shapley, L. S. (1974). *Values of Non-Atomic Games*. Princeton, NJ: Princeton University Press.
- [12] Babbel, D. G. (2010). A note on scenario analysis in the measurement of operational risk capital: a change of measure approach. Retrieved from <http://fic.wharton.upenn.edu/fic/papers/10/10.htm>
- [13] Badescu, A. L., Gong, L., Lin, X. S., & Tang, D. (2015). Modeling correlated frequencies with application in operational risk management. *Journal of Operational Risk*, 10(1), 1-43.
- [14] Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *The Journal of Economic Perspectives*, 27(1), 173-195.
- [15] Basel Committee on Banking Supervision. (1988). *International Convergence of Capital Measurement and Capital Standards*. Retrieved from <http://www.bis.org/publ/bcbsc111.pdf>
- [16] Basel Committee on Banking Supervision. (1996). *Amendment to the Capital Accord to Incorporate Market Risks*. Retrieved from <http://www.bis.org/publ/bcbs24.pdf>

- [17] Basel Committee on Banking Supervision. (2003). *Quantitative Impact Study 3 - Overview of Global Results*. Retrieved from <http://www.bis.org/bcbs/qis/qis3results.pdf>
- [18] Basel Committee on Banking Supervision. (2005). *An Explanatory Note on the Basel II IRB Risk Weight Functions*. Retrieved from <http://www.bis.org/bcbs/irbriskweight.pdf>
- [19] Basel Committee on Banking Supervision. (2006). *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. Retrieved from <http://www.bis.org/publ/bcbs128.pdf>
- [20] Basel Committee on Banking Supervision. (2009). *Principles for sound stress testing practices and supervision*. Retrieved from <https://www.bis.org/publ/bcbs147.pdf>
- [21] Basel Committee on Banking Supervision. (2009). *Revisions to the Basel II market risk framework*. Retrieved from <https://www.bis.org/publ/bcbs158.pdf>
- [22] Basel Committee on Banking Supervision. (2011). *Operational Risk Supervisory Guidelines for the Advanced Measurement Approaches*. Retrieved from <http://www.bis.org/publ/bcbs196.pdf>
- [23] Basel Committee on Banking Supervision. (2011). *Principles for the Sound Management of Operational Risk*. Retrieved from <http://www.bis.org/publ/bcbs195.pdf>
- [24] Basel Committee on Banking Supervision. (2011). *Basel III: A global regulatory framework for more resilient banks and banking systems*. Retrieved from <http://www.bis.org/publ/bcbs189.pdf>
- [25] Basel Committee on Banking Supervision. (2012). *Composition of capital disclosure requirements*. Retrieved from <http://www.bis.org/publ/bcbs221.pdf>

- [26] Basel Committee on Banking Supervision. (2013). *Charter*. Retrieved from <http://www.bis.org/bcbs/charter.pdf>
- [27] Basel Committee on Banking Supervision. (2013). *Basel III: The Liquidity Coverage Ratio and liquidity risk monitoring tools*. Retrieved from <http://www.bis.org/publ/bcbs238.pdf>
- [28] Basel Committee on Banking Supervision. (2014). *Operational risk - Revisions to the simpler approaches*. Retrieved from <http://www.bis.org/publ/bcbs291.pdf>
- [29] Basel Committee on Banking Supervision. (2014). *Basel III: the net stable funding ratio*. Retrieved from <http://www.bis.org/bcbs/publ/d295.pdf>
- [30] Berger, A. N., Herring, R. J., & Szegö, G. P. (1995). The role of capital in financial institutions. *Journal of Banking & Finance*, 19(3), 393-430.
- [31] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York, NY: Springer-Verlag.
- [32] Blum, P., Dias, A., & Embrechts, P. (2002). The ART of dependence modelling: the latest advances in correlation analysis. *Alternative risk strategies*, 339-356.
- [33] Board of Governors of the Federal Reserve System. (2011). *Supervisory Guidance on Implementation Issues Related to the Advanced Measurement Approaches for Operational Risk*. Retrieved from <https://www.federalreserve.gov/bankinfo/srletters/sr1108.pdf>
- [34] Board of Governors of the Federal Reserve System. (2014). *Supervisory Guidance for Data, Modeling, and Model Risk Management Under the Operational Risk Advanced Measurement Approaches*. Retrieved from <https://www.federalreserve.gov/bankinfo/basel/files/bcc1401.pdf>
- [35] Böcker, K., & Klüppelberg, C. (2005). Operational VaR: a Closed-Form Approximation. *Risk*, 18(12), 90-93.

- [36] Böcker, K., & Sprittulla, J. (2006). Operational VaR: meaningful means. *Risk*, 19(12), 96-98.
- [37] Bühlmann, H., Shevchenko, P. V., & Wüthrich, M. V. (2007). A Toy Model for Operational Risk Quantification using Credibility Theory. *Journal of Operational Risk*, 2(1), 2-19.
- [38] Campbell, B. (2011). *Modeling Risk with Limited Internal Data*. Retrieved from <http://www.northinfo.com/documents/455.pdf>
- [39] Chapelle, A. (2014). Modern Scenario Analysis. *Operational Risk & Regulation*, 15(9), 37.
- [40] Chaudhury, M. (2010). A Review of the Key Issues in Operational Risk Capital Modeling. *Journal of Operational Risk*, 5(3), 37-66.
- [41] Cheridito, P., Delbaen, F., & Kupper, M. (2006). Dynamic Monetary Risk Measures for Bounded Discrete-Time Processes. *Electronic Journal of Probability*, 11(3), 57-106.
- [42] Cheridito, P., & Li, T. (2009). Risk measures on Orlicz hearts. *Mathematical Finance*, 19(2), 189-214.
- [43] Cherny, A. S. (2009). Capital Allocation and Risk Contribution with Discrete-Time Coherent Risk. *Mathematical Finance*, 19(1), 13-40.
- [44] Colombo, A., & Desando, S. (2008). *Developing and Implementing Scenario Analysis Models to Measure Operational Risk at Intesa Sanpaolo*. Retrieved from http://www.mathworks.com/tagteam/52614_91606v00_intesa_upd.pdf
- [45] Cope, E. W. (2012). Combining scenario analysis with loss data in operational risk quantification. *Journal of Operational Risk*, 7(1), 39-56.

- [46] Cope, E., Mignola, G., Antonini, G., & Ugoccioni, R. (2009). Challenges and pitfalls in measuring operational risk from loss data. *Journal of Operational Risk*, 4(4), 3-27.
- [47] Cruz, M. G., Peters, G. W., & Shevchenko, P. V. (2015). *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*. Hoboken, NJ: Wiley.
- [48] Danielsson, J. (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking & Finance*, 26(7), 1273-1296.
- [49] Davies, J., Finlay, M., McLenaghan, T., & Wilson, D. (2006). Key Risk Indicators - Their Role in Operational Risk Management and Measurement. *ARM and RiskBusiness International*, 1-32.
- [50] De Jongh, P. J., De Wet, T., Raubenheimer, H., & Venter, J. H. (2015). Combining scenario and historical data in the loss distribution approach: a new procedure that incorporates measures of agreement between scenarios and historical data. *Journal of Operational Risk*, 10(1), 45-76.
- [51] Deelstra, G., Dhaene, J., & Vanmaele, M. (2011). *An overview of comonotonicity and its applications in finance and insurance. In Advanced mathematical methods for finance*. Berlin - Heidelberg: Springer-Verlag.
- [52] Degen, M. (2010). The calculation of minimum regulatory capital using single-loss approximation. *Journal of Operational Risk*, 5(4), 3-17.
- [53] Denault, M. (2001). Coherent allocation of risk capital. *Journal of Risk*, 4(1), 1-34.
- [54] Derman, E. (2011). Metaphors, Models & Theories. *The Quarterly Journal of Finance*, 1(1), 109-126.

- [55] Dhaene, J., Goovaerts, M. J., & Kaas, R. (2003). Economic capital allocation derived from risk measures. *North American Actuarial Journal*, 7(2), 44-56.
- [56] Doff, R. (2008). Defining and measuring business risk in an economic-capital framework. *The Journal of Risk Finance*, 9(4), 317-333.
- [57] Dutta, K. K., & Babbel, D. F. (2013). Scenario Analysis in the Measurement of Operational Risk Capital: A Change of Measure Approach. *The Journal of Risk and Insurance*, 81(2), 303-334.
- [58] Dutta, K., & Perry, J. (2006). A Tale of Tails: An Empirical Analysis of Loss Distributions Models for Estimating Operational Risk Capital. FRB of Boston. Working Paper No. 06-13. Retrieved from <http://www.bostonfed.org/economic/wp/wp2006/wp0613.pdf>
- [59] Embrechts, P. (2009). Copulas: A personal view. *Journal of Risk and Insurance*, 76(3), 639-650.
- [60] Embrechts, P. (2012). Four Theorems and a Financial Crisis, Fields Quantitative Finance Seminar, Fields Institute, Toronto, ON.
- [61] Embrechts, P., & Hofert, M. (2011). Practices and Issues in Operational Risk Modeling Under Basel II. *Lithuanian Mathematical Journal*, 51(2), 180-193.
- [62] Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., & Beleraj, A. (2014). An Academic Response to Basel 3.5. *Risks*, 2(1), 25-48.
- [63] Epstein, C. L. (2008). *Introduction to the Mathematics of Medical Imaging*. Philadelphia, PA: Society for Industrial & Applied Mathematics.
- [64] Ergashev, B. A. (2012). A Theoretical Framework for Incorporating Scenarios into Operational Risk Modeling. *Journal of Financial Services Research*, 41(3), 145-161.

- [65] Ergashev, B., Mittnik, S., & Sekeris, E. (2013). A Bayesian approach to extreme value estimation in operational risk modeling. *Journal of Operational Risk*, 8(4), 55-80.
- [66] European Parliament and Council of the European Union. (2009). Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the Taking-Up and Pursuit of the Business of Insurance and Reinsurance (Solvency II). Official Journal of the European Union 52 (L 335).
- [67] Federal Reserve Bank of New York. (2003). *Scenario-based AMA*. Retrieved from <http://www.newyorkfed.org/newsevents/events/banking/2003/con0529c.pdf>
- [68] Field, E. H., Dawson, T. E., Felzer, K. R., Frankel, A. D., Gupta, V., Jordan, T. H., ... & Wills, C. J. (2009). Uniform California Earthquake Rupture Forecast, Version 2 (UCERF 2). *Bulletin of the Seismological Society of America*, 99(4), 2053-2107.
- [69] Föllmer, H., & Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and stochastics*, 6(4), 429-447.
- [70] Frachot, A., Moudoulaud, O., & Roncalli, T. (2003). *Loss distribution approach in practice*. Retrieved from http://www.maths-fi.com/malliavin/Loss_Distribution_Approach_in_Practice_05_02_2003.pdf
- [71] Frachot, A., Roncalli, T., & Salomon, E. (2004). *The Correlation Problem in Operational Risk*. Retrieved from <http://www.thierry-roncalli.com/download/lda-correlations.pdf>
- [72] Gordy, M. B. (2003). A Risk-factor Model Foundation for Ratings-based Bank Capital Rules. *Journal of Financial Intermediation*, 12(3), 199 - 232.
- [73] Haackert, M., & Regling, S. (2012). Picture this. *Operational Risk & Regulation*, 13(2), 34-35.

- [74] Hora, M., & Klassen R. (2013). Learning from others' misfortune: Factors influencing knowledge acquisition to reduce operational risk. *Journal of Operations Management*, 31(1), 52-61.
- [75] Horbenko, N., Ruckdeschel, P., & Bae, T. (2011). Robust estimation of operational risk. *Journal of Operational Risk*, 6(2), 3-30.
- [76] Inanoglu, H. & Ulman, S. (2009). Revisiting Copula Dependency Modeling: A Case for Conservatism. Retrieved from http://www.fdic.gov/bank/analytical/cfr/2009/apr/CFR_ulman_04.08.09.pdf
- [77] James, C. M. (1996). *RAROC based capital budgeting and performance evaluation: a case study of bank capital allocation*. Research report, Wharton Financial Institutions Center. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1000
- [78] Jiménez-Rodríguez, E. J., Feria-Domínguez, J. M., Martín-Marín, J. L. (2011). The regulatory loss cut-off level: Does it undervalue the operational capital at risk? *The Spanish Review of Financial Economics*, 9, 49-54.
- [79] Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-292.
- [80] Kalkbrener, M. (2005). An axiomatic approach to capital allocation. *Mathematical Finance*, 15(3), 425-437.
- [81] Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). *Loss Models From Data To Decisions 4th ed.* Hoboken, NJ: John Wiley and Sons Inc.
- [82] Lambrigger, D. D., Shevchenko, P. V., & Wüthrich, M. V. (2007). The quantification of operational risk using internal data, relevant external data and expert opinion. *Journal of Operational Risk*, 2(3), 3-27.

- [83] Levy, H. (2006). *Stochastic dominance: investment decision making under uncertainty*. New York, NY: Springer Science+Business Media.
- [84] McConnell, P., & Blacker, K. (2013). Systemic operational risk: does it exist and, if so, how do we regulate it? *Journal of Operational Risk*, 8(1), 59-99.
- [85] McDermott, P., O'Brien, P. & Swenson, K. (2012). A model approach - examining operational risk capital modelling. *Operational Risk & Regulation*, 13(2), 26-29.
- [86] McKenna, B. (2014, March 28). In 'new reality,' Poloz changes the game plan. *The Globe and Mail*. Retrieved from <http://www.theglobeandmail.com/report-on-business/bank-of-canada-adapting-economic-models-to-new-reality/article17721912/>
- [87] McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, NJ: Princeton University Press.
- [88] Meek, J. (2012). Why we quantify: profile Steve Wilson. *Operational Risk and Regulation*, 13(2), 14-17.
- [89] Merton, R. C. (1974) On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29(2), 449 - 470.
- [90] Merton, R. C., & Perold, A. F. (1993). Theory of risk capital in financial firms. *Journal of Applied Corporate Finance*, 6, 16-32.
- [91] Mitra, S. (2013). Operational risk of option hedging. *Economic Modelling*, 33, 194-203.
- [92] Moosa, I. (2011). Operational risk as a function of the state of the economy. *Economic Modelling*, 28(5), 2137-2142.
- [93] Myers, S. C., & Read Jr, J. A. (2001). Capital allocation for insurance companies. *Journal of Risk and Insurance*, 68(4), 545-580.

- [94] Neslehova, J., Embrechts, P., & Chavez-Demoulin, V. (2006). Infinite-mean models and the LDA for operational risk. *Journal of Operational Risk*, 1(1), 3-25.
- [95] Office of the Comptroller of the Currency. (2012). Operational Risk for Foreign Banking Supervisors - 4 day workshop. Washington, DC.
- [96] Operational Risk and Regulation. (2009). *Data*, 10(2), 42.
- [97] Operational Risk and Regulation. (2011). *Data*, 12(6), 40-41.
- [98] Operational Risk and Regulation. (2013). *Data*, 14(3), 35-36.
- [99] Osgood, B. (2007). Lecture Notes for EE 261 The Fourier Transform and its Applications. Stanford University. Retrieved from <http://see.stanford.edu/materials/lsoftae261/book-fall-07.pdf>
- [100] Panjer, H. H. (2002). *Measurement of Risk, Solvency Requirements and Allocation of Capital Within Financial Conglomerates*. Research report, Institute of Insurance and Pension Research. Retrieved from https://www.soa.org/Files/Research/Projects/measurement_risk.pdf
- [101] Panjer, H. H. (2006). *Operational Risk Modeling Analytics*. Hoboken, NJ: John Wiley and Sons Inc.
- [102] Robertson, J. (1992). The computation of aggregate loss distributions. *Proceedings of the Casualty Actuarial Society*, 79(150), 57-133.
- [103] Ross, S. (2002). *A First Course in Probability*. Upper Saddle River, NJ: Prentice Hall.
- [104] Shafer, M., & Yildirim, Y. (2013). Operational risk and equity prices. *Finance Research Letters*, 10(4), 157-168.
- [105] Shapley, L. S. (1953). *A Value for n-Person Games*. Retrieved from <http://www.dtic.mil/dtic/tr/fulltext/u2/604084.pdf>

- [106] Shevchenko, P. V., & Wüthrich, M. V. (2006). The structural modeling of operational risk via Bayesian inference: combining loss data with expert opinions. *Journal of Operational Risk*, 1(3), 3-26.
- [107] Spronk, J., & Hallerbach, W. (1997). Financial modelling: Where to go? With an illustration for portfolio management. *European Journal of Operational Research*, 99(1), 113-125.
- [108] Tasche, D. (2008). *Capital allocation to business units and sub-portfolios: the Euler principle*. Retrieved from <http://arxiv.org/pdf/0708.2542.pdf>
- [109] Tozer-Pennington, V. (2011). Japan shows value of multivariate scenarios. *Operational Risk & Regulation*, 12(4), 11.
- [110] Urban, M., Dittrich, J., Klüppelberg, C., & Stölting, R. (2004). Allocation of Risk Capital to Insurance Portfolios. *Blätter der DGVM*, 26, 389-406.
- [111] Vasicek, O. (2002). Loan portfolio value. *RISK*, 15(12), 160 - 162.

8 Appendix A: Probability Distributions

8.1 Frequency Distributions

The following definitions for all probability distributions were taken from Klugman *et al* (2012) [81].

Definition 8.1.1. (*Poisson*) A Poisson distribution is a discrete distribution. The probability density function p_k denotes the probability that exactly k events (in this case losses) occur. Let N be a random variable representing the number of such events. Then

$$p_k = Pr(N = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots \quad (8.1)$$

The mean and variance are equal and given as $E(N) = Var(N) = \lambda$.

Definition 8.1.2. (*Negative Binomial*) A Negative Binomial distribution is a discrete distribution. The probability density function is given as

$$p_k = \binom{k+r-1}{k} \left(\frac{1}{1+\beta} \right)^r \left(\frac{\beta}{1+\beta} \right)^k, \quad k = 0, 1, 2, \dots, \quad r > 0, \quad \beta > 0. \quad (8.2)$$

The mean is given by $E(N) = r\beta$ and the variance given by $Var(N) = r\beta(1 + \beta)$.

8.2 Severity Probability Distributions

Definition 8.2.1. (*Lognormal*) The Lognormal distribution is a continuous two-parameter distribution (μ, σ) . The probability density function is given as

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad z = \frac{\ln x - \mu}{\sigma}. \quad (8.3)$$

The mean is given by $E[X] = e^{\mu + \frac{1}{2}\sigma^2}$ and the variance given by $\text{Var}[X] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$, or in general the moments given by $E[X^k] = e^{k\mu + \frac{1}{2}k^2\sigma^2}$.

Definition 8.2.2. (*Log-Gamma*) The Log-Gamma distribution is a continuous two-parameter distribution (α, θ) . We state the probability density and moments for $X \sim \text{Gamma}(\alpha, \theta)$ where is understood that $X = \log Y \sim \text{Gamma}$, then $e^X = Y \sim \text{Log-Gamma}$

$$f(x) = \frac{\left(\frac{x}{\theta}\right)^\alpha e^{-\frac{x}{\theta}}}{x\Gamma(\alpha)}. \quad (8.4)$$

The moments are given by $E[X^k] = \frac{\theta^k \Gamma(\alpha + k)}{\Gamma(\alpha)}$, $k > -\alpha$.

Definition 8.2.3. (*Weibull*) The Weibull distribution is a continuous two-parameter distribution (α, τ) . The probability density function and moments are given as

$$f(x) = \frac{\tau \left(\frac{x}{\theta}\right)^{\tau-1} e^{-\left(\frac{x}{\theta}\right)^\tau}}{x} \quad (8.5)$$

and $E[X^k] = \theta^k \Gamma\left(1 + \frac{k}{\tau}\right)$ $k > -\tau$.

Definition 8.2.4. (*Generalized Pareto*) The Generalized Pareto distribution is a continuous three-parameter distribution (α, θ, τ) - beta of the second kind. The probability density function and moments are given as

$$f(x) = \frac{\Gamma(\alpha + \tau)}{\Gamma(\alpha)\Gamma(\tau)} \frac{\theta^\alpha x^{\tau-1}}{(x + \theta)^{\alpha+\tau}} \quad (8.6)$$

$$\text{and } E[X^k] = \frac{\theta^k \Gamma(\tau+k) \Gamma(\alpha-k)}{\Gamma(\alpha) \Gamma(\tau)}, \quad -\tau < k < \alpha.$$

Definition 8.2.5. (Burr) The Burr Type XII distribution is a continuous three-parameter distribution (α, θ, γ) . The probability density function and moments are given as

$$f(x) = \frac{\alpha \gamma \left(\frac{x}{\theta}\right)^\gamma}{x \left[1 + \left(\frac{x}{\theta}\right)^\gamma\right]^{\alpha+1}} \quad (8.7)$$

$$\text{and } E[X^k] = \frac{\theta^k \Gamma(1+\frac{k}{\gamma}) \Gamma(\alpha-\frac{k}{\gamma})}{\Gamma(\alpha)}, \quad -\gamma < k < \alpha\gamma.$$

9 Appendix B: Loss Aggregation

9.1 Single Loss Approximation

Böcker and Klüppelberg (2005) [35] show that when loss data is heavy-tailed, a simple closed-form approximation for operational VaR can be obtained. A key result needed is that for a standard LDA model with subexponential severities, one has under weak regularity conditions, for every fixed $t > 0$

$$\bar{G}_t(x) \sim EN(t)\bar{F}(x), \quad x \rightarrow \infty, \quad (9.1)$$

where $EN(t)$ is the expected frequency and $\bar{F}(\cdot) = 1 - F(\cdot)$ and $\bar{G}_t(\cdot) = 1 - G_t(\cdot)$ are the tail distributions of severity and aggregate loss, respectively. The symbol \sim indicates that the quotient of right hand and left hand side tends to 1; i.e. $\lim_{x \rightarrow \infty} \frac{\bar{G}_t(x)}{\bar{F}(x)} = EN(t)$ for every fixed $t > 0$. Using this result, we have the following asymptotic formula.

Theorem 9.1.1. *Consider the standard LDA model for fixed $t > 0$ and a subexponential severity with distribution function F . Assume, moreover, that the tail estimate (9.1) holds. Then, the $VaR_t(\kappa)$ satisfies the approximation*

$$VaR_t(\kappa) = F^{\leftarrow} \left(1 - \frac{1 - \kappa}{EN(t)} (1 + o(1)) \right), \quad \kappa \rightarrow 1, \quad (9.2)$$

where F^{\leftarrow} denotes the generalized inverse of F and $o(1)$ stands for a function which

tends to 0 if its arguments tends to a boundary (in our case if $\kappa \rightarrow 1$ or $x \rightarrow \infty$).

It is noteworthy to point out that a slight improvement to (9.2) is seen in Böcker and Sprittulla (2006) [35] and referred to as the 'mean-corrected' version. Thus the modified finite mean single loss approximation is given as

$$VaR_t(\kappa) = F^{\leftarrow} \left(1 - \frac{1 - \kappa}{\lambda} (1 + o(1)) \right) + (\lambda - 1)\mu, \quad \kappa \rightarrow 1, \quad (9.3)$$

where $\lambda = EN(t)$ denotes frequency expectation and the severity distribution F has finite expectation $\mu = E(X_i)$. This ensures that the aggregate operational loss has finite expectation given by $ES = \lambda\mu$. For an analytical justification of (9.3), we refer the reader to Degen (2010) [52]. Moreover, (9.3) has even been adopted as the preferred method to compute capital in the BCBS's benchmark operational capital model (2014) [28].

9.2 Panjer Recursion

Alternative measures of loss aggregation are nicely covered in Panjer (2006) [101] and adapted here. One such method improves upon the direct method to compute an integral by using a recursive method. When attempting to evaluate the convolution of the severity distribution

$$F_X^{*k}(x) = \int_{0^-}^x F_X^{*(k-1)}(x-y) dF_X(y), \quad (9.4)$$

numerical integration methods must be used. A simplification may be used by replacing the severity distribution by a discrete distribution defined at multiples $0, 1, 2, \dots$, of some monetary unit such as \$1,000. This simplifies (9.4) as

$$F_X^{*k}(x) = \sum_{y=0}^x F_X^{*(k-1)}(x-y) f_X(y), \quad (9.5)$$

with PDF

$$f_X^{*k}(x) = \sum_{y=0}^x f_X^{*(k-1)}(x-y)f_X(y). \quad (9.6)$$

Even when the severity distribution is defined on nonnegative integers $0, 1, 2, \dots$, calculating $f_X^{*k}(x)$ requires $x + 1$ multiplications. Then carrying out this operation for all possible values of k and x up to m requires multiplications on the order of m^3 or $\mathcal{O}(m^3)$. The introduction of Panjer recursion (1981) reduces the number of computations to $\mathcal{O}(m^2)$

The method requires that the severity distribution is discretized according to some monetary unit and that the frequency distribution, p_k is a member of the $(a, b, 1)$ class which implies that

$$p_k = \left(a + \frac{b}{k} \right) p_{k-1}, \quad k = 2, 3, 4, \dots \quad (9.7)$$

Hence this recursion starts at p_1 rather than p_0 which has the same shape as a $(a, b, 0)$ class with remaining probability at $k = 0$. Then the following result holds.

Theorem 9.2.1. (*Extended Panjer recursion*) For the $(a, b, 1)$ class, the aggregate loss, f_S , PDF may be calculated as

$$f_S(x) = \frac{[p_1 - (a + b)p_0]f_X(x) + \sum_{y=1}^{\min(x,m)} (a + by/x)f_X(y)f_S(x-y)}{1 - af_X(0)}. \quad (9.8)$$

Corollary 9.2.1. (*Panjer recursion*) For the $(a, b, 0)$ class, (9.8) reduces to

$$f_S(x) = \frac{\sum_{y=1}^{\min(x,m)} (a + by/x)f_X(y)f_S(x-y)}{1 - af_X(0)}. \quad (9.9)$$

As a special case, for the Poisson distribution, (9.9) reduces to

$$f_S(x) = \frac{\lambda}{x} \sum_{y=1}^{\min(x,m)} y f_X(y) f_S(x-y), \quad x = 1, 2, \dots \quad (9.10)$$

In the simplest of interpretations, the right hand side contains terms of $f_S(x-1), \dots, f_S(0)$ and hence defines a recursive formula for $f_S(x)$.

9.3 Fast Fourier Transform

Yet another method to compute the aggregate loss distribution is via an inversion method. The theory is adapted from Panjer (2006) [101] and Robertson (1992) [102] with the latter providing exhaustive detail on the mathematics needed and straight forward examples to provide the mechanics of the algorithm. In the simplest of descriptions, the discrete fourier transform (DFT) is an operation that helps compute the convolution of two vectors: U, V . For brevity we have

$$DFT(U * V) = DFT(U) X DFT(V) \quad (9.11)$$

$$U * V = IDFT\left(DFT(U) X DFT(V)\right), \quad (9.12)$$

where IDFT is the inverse of DFT. Hence to compute the convolution of two vectors, the DFT of each vector is taken and the result multiplied together pointwise. The convolution is recovered by taking the inverse of both sides. The reference to “Fast” (FFT) lies in a computation simplification that reduces the required number of computation and makes the algorithm perform faster.

Providing more of a detailed description, compound distributions lend themselves naturally to inversion methods as their transformations are compound functions and are easily computed when the frequency and severity components are known. For

example, the characteristic function always exists and is unique. It is defined as

$$\varphi_S(z) = E[e^{iSz}] = P_N[\varphi_X(z)]. \quad (9.13)$$

The FFT algorithm can be used for inverting characteristic functions to obtain densities of discrete-random variables.

Definition 9.3.1. (*Fourier Transform*) For any continuous function $f(x)$, the Fourier Transform is the mapping

$$\tilde{f}(z) = \int_{-\infty}^{\infty} f(x)e^{izx} dx, \quad (9.14)$$

where $f(x)$ is the PDF and $\tilde{f}(z)$ is its characteristic function and is complex valued.

The original PDF can be recovered as

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(z)e^{-izx} dz. \quad (9.15)$$

Being able to discretize $f(x)$ is a necessity in order to carry out the computation on the computer. In that sense, we work with the DFT.

Definition 9.3.2. (*Discrete Fourier Transform*) Let f_x denote a function defined for all integer values of x that is periodic with period length n (i.e. $f_{x+n} = f_x$ for all x). For the vector $(f_0, f_1, \dots, f_{n-1})$, the Discrete Fourier Transform is the mapping \tilde{f}_x , $x = \dots, -1, 0, 1, \dots$, defined by

$$\tilde{f}_k = \sum_{j=0}^{n-1} f_j \exp\left(\frac{2\pi i}{n}jk\right), \quad k = \dots, -1, 0, 1, \dots \quad (9.16)$$

The inverse mapping is

$$f_j = \frac{1}{n} \sum_{k=0}^{n-1} \tilde{f}_k \exp\left(\frac{-2\pi i}{n}jk\right), \quad j = \dots, -1, 0, 1, \dots \quad (9.17)$$

From (9.16), it then becomes clear that in order to produce n values of \tilde{f}_k , n^2 terms need to be evaluated, or $\mathcal{O}(n^2)$. The FFT reduces the number of computations to $\mathcal{O}(n \ln_2 n)$. The algorithm uses the property that a DFT of length n can be rewritten as the sum of two discrete transforms, each of length $\frac{n}{2}$, the first consisting of even-numbered points and the second consisting of odd-numbered points:

$$\begin{aligned}
\tilde{f}_k &= \sum_{j=0}^{n-1} f_j \exp\left(\frac{2\pi i}{n} jk\right) \\
&= \sum_{j=0}^{n/2-1} f_{2j} \exp\left(\frac{2\pi i}{n} 2jk\right) + \sum_{j=0}^{n/2-1} f_{2j+1} \exp\left[\frac{2\pi i}{n} (2j+1)k\right] \\
&= \sum_{j=0}^{m-1} f_{2j} \exp\left(\frac{2\pi i}{m} jk\right) + \exp\left(\frac{2\pi i}{n} k\right) \sum_{j=0}^{m-1} f_{2j+1} \exp\left(\frac{2\pi i}{m} jk\right),
\end{aligned}$$

where $m = \frac{n}{2}$. Then

$$\tilde{f}_k = \tilde{f}_k^a + \exp\left(\frac{2\pi i}{n} k\right) \tilde{f}_k^b, \quad (9.18)$$

which in turn can be written as the sum of two transformations of length $\frac{m}{2}$ and continued successively.