

Essays on Capital Calculation in Insurance

THÈSE N° 7269 (2016)

PRÉSENTÉE LE 18 NOVEMBRE 2016

AU COLLÈGE DU MANAGEMENT DE LA TECHNOLOGIE
CHAIRE SWISSQUOTE EN FINANCE QUANTITATIVE
PROGRAMME DOCTORAL EN MATHÉMATIQUES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Mathieu Jacques David CAMBOU

acceptée sur proposition du jury:

Prof. T. Mountford, président du jury
Prof. D. Filipovic, Prof. A. C. Davison, directeurs de thèse
Prof. M. Hofert, rapporteur
Dr M. Dacorogna, rapporteur
Prof. S. Morgenthaler, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2016

Acknowledgements

I would like to address my first thanks to my thesis director Damir Filipović. I felt like a chance to be advised and supported with such a strong dedication, while diving into mathematical concepts, bridging the gap with the industry and trying to explore ideas from an innovative angle. This is the base of the knowledge I have acquired during this thesis. I also truly thank my thesis co-director Anthony Davison who has provided me with constant support and encouragements since my undergraduate studies at EPFL, even sometimes wishing my birthday before my own parents.

I am indebted to SCOR for funding this thesis and getting me involved in practical projects, with a constant desire to share knowledge and experience. It was such a pleasure to work in a research driven company with a good sense of pragmatism.

Many thanks to the individuals I have had the chance to meet at SCOR, who always shared reinsurance knowledge with passion and patience, especially: Philipp Arbenz, Andrea Biancheri, Davide Canestraro, Michel Dacorogna, Eric Dal Moro, Alessandro Ferriero, William Guevara, Stephan Küttel, Tony Neghaiwi, Patrick Tapernoux, Nigel Riley and Jessika Walter. I also thank Christoph Hummel and Raffaele Dell'Amore from Secquaero Advisors.

I had the chance to visit other universities to complete research projects. In particular, I thank RiskLab at ETHZ and Paul Embrechts for hosting me during the research project on importance sampling, and Philippe Charmoy at Oxford. I also thank Marius Hofert, Chritiane Lemieux and the members of the Department of Statistics and Actuarial Science at University of Waterloo for welcoming me during the research project on quasi-random copula, probably the best academic visiting experience I had. Many thanks to Sophie at EPFL for her true support, also and most importantly for the non-administrative part of this thesis.

While I am just putting this research experience behind me, a great adventure is waiting at the corner, of which I am a modest and admiring contributor. Hence, I would like to thank all my colleagues from EdgeLab for their enthusiasm and boundless energy during this transition.

I am truly thankful to the precious friends met at different stages: in Lausanne, Joachim, Mikael, Orane, Timothée and Vlad, at Cambridge, Anaël, Alexandre, Emilie, Jean-Jacques and Rémy, in London, Bertrand, Christoph, Fabrice, Guillaume, Morgane and Romain. My friends from home, Benoit, Jonathan, Julien, Maxime and Sylvain, with who it remains as if I never left.

Thanks to Julie whose love and patience were necessary to create a stable equilibrium.

Finally, I thank my parents whose unbounded love and pride will always remain a main driver. After being nearly asked everyday it is such a relief to be able to say... "Voilà papa, j'ai fini ma thèse!".

Abstract

In order to be able to bear the risk they are taking, insurance companies have to set aside a certain amount of cushion that can guarantee the payment of liabilities, up to a defined probability, and thus to remain solvent in case of bad events. This amount is named capital. The calculation of capital is a complex problem. To be sustainable, capital must consider all possible risk sources that may lead to losses among assets and liabilities of the insurance company, and it must account for the likelihood and the effect of these bad (and usually extreme) events that could occur to the risk sources. Insurance companies build models and tools in order to perform this capital calculation. For that, they have to collect data, build statistical evidence, build mathematical models and tools in order to efficiently and accurately derive capital.

The papers exposed in this thesis deal with three major difficulties. First, the uncertainty behind the choice of a specific model and the quantification of this uncertainty in terms of additional capital. The use of external scenarios, i.e. opinions on the likelihood of some events happening, allows to build a coherent methodology that make the cushion more robust against wrong model specification. Second, the computational complexity in using these models in an industrialized environment, and numerical methods available for increasing their computational efficiency. Most of these models cannot provide an analytical formula of capital. Consequently, one has to approximate it via simulation methods. Considering the high number of risk sources and the complexity of insurance contracts, these methods can be slow to run before providing a reasonable accuracy. This often makes these models unusable in practical cases. Enhancements of classical simulation methods are presented in the aim of making these approximations faster to run for the same level of accuracy. Third, the lack of reliable data and the high complexity of problems with long time horizons, and statistical methods for identifying and building reliable proxies in such cases. A typical example is life-insurance contracts that imply being exposed to multiple risks sources over a long horizon. Such contracts can in fact be approximated wisely by proxies that can capture the risk over time.

Key words: model uncertainty, scenario aggregation, statistical divergence, importance sampling, quasi-random numbers, copula, dependence models, conditional distribution method, risk measure, expected shortfall, value-at-risk, tail event, replicating portfolio, Swiss Solvency Test, Solvency capital, insurance, asset-liability portfolio

Résumé

Afin de supporter le risque qu'elles prennent, les companies d'assurance doivent conserver un coussin de sécurité qui a pour but de garantir, jusqu'à une certaine probabilité, le remboursement de son passif d'assurance et donc de rester solvable en cas d'évènements adverses. Ce coussin de sécurité est appelé le capital. Le calcul de ce capital est un problème complexe. Afin qu'il soit approprié, il doit être calculé en prenant en compte toutes les sources de risque parmi l'actif et le passif de l'assurance, qui pourraient provoquer des pertes. De plus, il doit prendre en compte la probabilité et l'effet de ces évènements adverses (souvent extrêmes). Pour cela, les compagnies d'assurance construisent des modèles et des outils afin de calculer ce capital. Plus précisément, elles collectent des données, les analysent et en déduisent des évidences statistiques, construisent les modèles sur ces évidences afin de calculer de manière efficiente et précise le capital requis.

Les papiers de recherche présentés dans cette thèse se concentrent sur trois difficultés majeures. Premièrement, l'incertitude dans le choix d'un modèle spécifique et la quantification de cette incertitude en terme de capital supplémentaire. Des opinions exogènes sur la probabilité de certains évènements nous permettent de construire une méthode afin de rendre ce coussin de sécurité plus robuste contre de possibles erreurs faites dans la spécification de modèle. Deuxièmement, la complexité de l'utilisation de ces modèles dans une réalité de production industrielle et les méthodes numériques visant à accroître l'efficacité de ces outils. La plupart de ces modèles ne permettent pas de calculer le capital de manière analytique et exacte. L'utilisation de méthodes de simulations numériques est donc nécessaire. Au vue du grand nombre de sources de risque et de la complexité de certains contrats d'assurance, ces méthodes ne peuvent souvent fournir un résultat d'une précision acceptable avant un temps de calcul important, de telle sorte que leur utilisation pratique est parfois remise en cause. Des améliorations de ces méthodes numériques sont présentées afin de les accélérer de manière importante tout en conservant un niveau de précision élevé. Finalement, le manque de données utilisables et la complexité élevée de certains problèmes avec un horizon de temps important, et les méthodes statistiques afin d'identifier et de construire des approximations suffisamment précises pour de tels cas. Un exemple typique serait un contrat d'assurance vie qui expose la compagnie d'assurance à divers risques durant un horizon de temps important. De tels contrats peuvent en fait être approximés par des proxies qui capturent l'évolution du risque au cours du temps.

Contents

Abstract	2
Introduction	6
Model Uncertainty and Scenario Aggregation	8
Variance reduction methods for copula models	9
Replicating Portfolio Approach to Capital Calculation	11
A Model Uncertainty and Scenario Aggregation	12
B An Importance Sampling Approach for Copula Models in Insurance	45
C Quasi-Random Numbers for Copula Models	73
D Replicating Portfolio Approach to Capital Calculation	120
Cumulative Bibliography	143
Curriculum Vitae	152

Accompanying papers

- A Mathieu Cambou, Damir Filipović
Model Uncertainty and Scenario Aggregation, [20]
Mathematical Finance, forthcoming.
- B Philipp Arbenz, Mathieu Cambou, Marius Hofert
An Importance Sampling Approach for Copula Models in Insurance, [4]
Submitted
- C Mathieu Cambou, Marius Hofert, Christiane Lemieux
Quasi-Random Numbers for Copula Models, [22]
Published in *Statistics and Computing*
- D Mathieu Cambou, Damir Filipović
Replicating Portfolio Approach to Capital Calculation, [21]
Submitted

Introduction

The role of an insurance company consists in relieving individuals, groups and institutions of damage to property and loss of life caused by abnormal events (Property and Casualty Insurance) and of diseases and death while providing pensions and annuity saving plans (Life Insurance). Risk is the essential sustaining element of this industry. Insurance companies are compensated for taking the risk to face the uncertain outcome of the insurance contracts they are underwriting. This compensation is what the insured parties are paying to (partially) transfer this risk. Also considering the assets that these insurance companies hold, the nature of the risks that they bear on their balance sheets can potentially be highly heterogeneous. Typically, it depends on the type of the underlying event that drives the result of the outcome (e.g. car accidents, loss reserves, storms, flooding, interest rates, credit spreads,...). It also depends on the type of contract in which the firm is engaged (e.g. derivatives, reinsurance contracts, alternative risk transfers,...). The understanding and the quantification of these risks are crucial for, e.g., underwriting and valuing insurance contracts, renewing and expanding the insurance exposure in a cost-efficient manner, deriving regulatory capital cushion in order to fulfil minimum solvency requirements, managing diversification and the impact of rare events,... These questions usually require a particular approach to modelling and their answers depend on a particular quantification of risk. However, they all have the same characteristic: the risks of these exposures arise from a finite collection of sources that are common to all exposures. For example, change in interest rates may simultaneously affect life insurance contracts and fixed income assets while storms would typically impact several insurance and (re)insurance contracts on the balance sheet of the same company.

These risk sources do not change in an isolated and independent manner. For example credit worthiness of debtors tend to decrease when interest rate increase while it is sometimes observed that car accidents increase following storms or flooding. In consequence, we cannot assume these risks sources to be independent. The modelling of risk in insurance therefore stands on an important probabilistic and statistical field: *the joint modelling of random events*, i.e. the joint modelling of the risk sources. A good overview of the applications of this field in finance and insurance can be found in [89]. It requires modelling the marginal behaviour of these risk sources, i.e. how they behave individually isolated from the other sources, and the dependency structure between them. When setting up a joint modelling framework in place, companies adopt a statistical approach, i.e.

- collecting, cleaning and analysing the data,
- building, comparing, selecting and possibly fitting models using these data,
- doing inference, predictions and estimations based on these models.

Ultimately, the derivations made by these models can serve different purposes. The focus of this thesis will be on the derivation of *capital*.

Capital is a risk-adjusted value requirement on the current balance sheet of an insurance company to be economically or regulatory acceptable. Capital is calculated by applying a risk measure to the profit and loss distribution of the asset-liability portfolio over a time horizon that is typically one year. Risk measures are quantifications of outcomes when facing future and uncertain events. In general, it is meant that the outcomes are the ones with adverse effects. In the insurance and financial industries, these adverse effects are typically losses incurred by insurance and financial exposures. [5] is an early reference for the goal of risk measures in these industries and the fundamental properties they should have. Risk measures are typically used for deriving capital, managing risk and evaluating fair compensations for underwriting insurance contracts.

Capital can be *economic capital* or *regulatory capital*. Economic capital, in a nutshell, is an amount that can be invested in risky but liquid assets an insurance company has to put aside, as a cushion against future adverse events. These events are typically tail scenarios in the distribution of balance sheet over specific time horizons, typically 1 year. Hence the use of risk measures. The calculation of economic capital is mostly guided by risk assessment and methodologies that are internal to the insurance company, with an internal view of risk exposures. The aim of this capital cushion is to remain solvent. This cushion has a cost. Hence companies derive metrics such as return on risk-adjusted capital (RORAC) or capital contribution in order to assess the profitability of business lines, contracts, projects, . . . Regulatory capital's primary goal is to protect policyholders against events that may affect the solvency of the insurance company. The methodology is typically guided and reviewed by the regulators. It is shaped from the view of the regulators that typically take a more global view of risk exposures and have access to more industry data in order to account for the systemic aspect of financial and insurance risks. Examples of regulatory capital can be found in e.g. [24], [52].

Several challenges arise in deriving capital. The literature of mathematics for insurance and finance is now providing a rich content for addressing these challenges. We are drawing attention to the following ones:

- (i) Although a model is chosen with all the available information and the best knowledge of the modellers, it may be misspecified. Due to the dimensionality of the problems that are being solved, it may happen that models are inappropriate although they show good fit to empirical data or attractive intuitive explanation. The derivation of risk measures based on statistical models may therefore not account for this uncertainty and therefore gives an underestimation of the true capital needed for maintaining solvency.
- (ii) Even if risk measures and capital are well defined, their derivations are often cumbersome. Large obstacles are
 - the complexity of the risk to quantify (e.g. the risk of highly non-linear insurance contracts that depend on several risk sources, the distribution of risk of these insurance contracts over time, . . .),
 - the complex and highly non-linear dependence between the various risk sources,
 - the computing resources demanded by high-dimensional modelling frameworks and the rarity of the events for which they intend to capture the impact.

A commonly used approach for deriving estimates of a quantity that cannot be derived analytically is the Monte-Carlo approach. The main drawback of these computational methods

is that their convergence rate to the true result may be poor, i.e. the runtime for obtaining an estimate with a reasonable error could be extremely long. As a result, insurance companies sometimes run simulations for several hours, or days, and consume high computing resources in order to obtain reasonably accurate results for problems that are typically difficult to implement in parallel.

These two subjects are addressed in this thesis across four papers. More precisely:

- Paper **A** gives an introduction to joint modelling of risk in insurance, to capital calculation and proposes an approach to bring robustness against model uncertainty.
- Papers **B** and **C** introduce variance reduction techniques for capital calculation in a particular joint modelling framework and study two specific methods, namely importance sampling and quasi-random sampling.
- Paper **D** proposes an alternative method to a highly involved nested Monte–Carlo problem. The Replicating Portfolio approach is not a new concept, however it is presented with an extensive improvement.

While Paper **A** is a generic improvement of modelling frameworks, Papers **B–C** give a contribution to the numerical methods that can be used in order to derive capital from models. Paper **D** applies to a specific context of capital calculation. The following gives a first introduction to these papers.

Model Uncertainty and Scenario Aggregation

The design of a model for deriving capital usually raises questions of appropriateness at many levels. The choices of the marginal models in order to model events individually or of the dependency structure in between these events are of typical concerns and are challenged by regulators, auditors or model validation teams. The inherent ambiguity in these choices is usually referred as *model uncertainty*. Paper **A** provides a coherent method for incorporating external views on scenarios into a model. The method is called *scenario aggregation*. It originally aims at supporting regulatory purposes, such as stress testing. It can in fact serve as a generic device in order to address model uncertainty.

A model in insurance can be formalized as a loss random variable L for a given time horizon defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which assigns to any possible state of the world $\omega \in \Omega$ a loss $L(\omega)$. The probability measure \mathbb{P} and the random variable L are specific for each insurance institution. While Ω can be viewed as a universal object, one can argue that the σ -algebra \mathcal{F} is not, as different institutions may not access the same information. We assume that \mathcal{F} is universal, taking \mathcal{F} to be the σ -algebra generated by all institution-specific σ -algebras and an appropriate extension of \mathbb{P} if needed. The aim of the scenario aggregation method is to reduce the uncertainty in the specification of \mathbb{P} . Section 2 of Paper **A** gives a deeper introduction to notion of model for capital calculation, in an insurance context.

Scenario are events, that are often given as narrative descriptions of state of the world. Mathematically, this event is a set $S \in \mathcal{F}$. A view on such scenario S is a requirement of the form $\mathbb{Q}[S] \geq c$, for some target probability c , to be satisfied by any probability measure \mathbb{Q} . If a model \mathbb{P} does not satisfy this view, the method would require to replace it by some alternative model \mathbb{Q} which satisfies it. Under the premise that models are designed with the best knowledge of the

modeller, the objective is then to choose the alternative model \mathbb{Q} as close as possible to \mathbb{P} . The distance of \mathbb{Q} to \mathbb{P} is measured through a statistical tool, called divergence and noted $d(\mathbb{Q}, \mathbb{P})$. Aggregating the views on scenarios can now be defined as finding a solution of the optimization problem

$$\begin{aligned} & \text{minimize} && d(\mathbb{Q}, \mathbb{P}) \\ & \text{subject to} && \text{views.} \end{aligned}$$

The paper details the coherent aspect of the method and puts a strong emphasis on the fact that it satisfies criteria that are particularly relevant from a regulatory point of view. In particular, it boils down to a finite-dimensional convex optimization problem which is a highly tractable numerical exercise. The alternative model resulting from the minimum divergence scenario aggregation is an interpolation between the original model and the views within minimal distance from the original model. In a regulatory framework, it gives remote control to the regulators over the capital requirements of insurance companies as he can tune the trade off between idiosyncrasy (companies own model) and standardization (regulator's views) via increasing the number of scenarios.

Variance reduction methods for copula models

For capital calculation, the quantity of interest can often be written $\mathbb{E}[\Psi_0(\mathbf{X})]$, where $\mathbf{X} = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ is a random vector with distribution function H on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\Psi_0 : \mathbb{R}^d \rightarrow \mathbb{R}$ is some measurable function. Note that, as per the above notation, \mathbf{X} could be the losses of each business lines of an insurance company and thus $L = \mathbf{1}^T \mathbf{X}$ could be the loss of the entire company. Since components of \mathbf{X} are typically dependent, a way to account for this dependence is to model the distribution of \mathbf{X} as

$$H(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d)), \quad \mathbf{x} \in \mathbb{R}^d,$$

where $F_j(x) = \mathbb{P}(X_j \leq x)$, $j \in \{1, \dots, d\}$, are the marginal distribution functions of H and $C : [0, 1]^d \rightarrow [0, 1]$ is a *copula*, i.e., a distribution function with standard uniform univariate margins. Such a dependence model allows one to separate the dependence structure from the marginal distributions.

The estimate of interest can be rewritten $\mathbb{E}[\Psi_0(\mathbf{X})] = \mathbb{E}[\Psi(\mathbf{U})]$ where $\mathbf{U} = (U_1, \dots, U_d) : \Omega \rightarrow \mathbb{R}^d$ is a random vector with distribution function C , $\Psi : [0, 1]^d \rightarrow \mathbb{R}$ is defined as $\Psi(u_1, \dots, u_d) = \Psi_0(F_1^-(u_1), \dots, F_d^-(u_d))$, and the $F_j^-(p)$'s are the marginal quantile functions. If C and the F_j 's are known, we can use Monte Carlo simulation to estimate $\mathbb{E}[\Psi(\mathbf{U})]$. For a (pseudo-)random sample $\{\mathbf{U}_i : i = 1, \dots, n\}$ from C , the classical Monte Carlo estimator of $\mathbb{E}[\Psi(\mathbf{U})]$ is given by

$$\frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{U}_i) \approx \mathbb{E}[\Psi(\mathbf{U})].$$

The accuracy of this estimator can be given in terms of the estimation error. For some error norm $\|\cdot\|_e$, Monte Carlo analysis have to be run in a setting such that

$$\left\| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{U}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right\|_e$$

is sufficiently small. In a classical approach this essentially means having a reasonably high sampling size n . Computation resource (CPU, memory, ...) and time are increasing functions of n . In some practical situations n is such that the computation time counts in days for reaching a given error acceptance level, n is in fact itself an increasing function of the problem dimension d , which is often high in insurance related problems.

In Paper **B**, we propose an *importance sampling* method in order to reduce the estimation error of the Monte-Carlo method for a given sampling size n . As typical for capital calculation, we focus on problems where Ψ_0 mostly depends on the tail of the distribution of \mathbf{X} . This is the case where Ψ is large only when at least one of the components of \mathbf{X} is large. These are rare-event problems, which correspond to a region in the sampling space of \mathbf{X} that has a small probability of occurrence. In such problems, the estimation error has a rate of convergence that is even slower than most problems. The idea of importance sampling is to find a random variable $\mathbf{V} : \Omega \rightarrow [0, 1]^d$ such that $\mathbb{E}[\Psi(\mathbf{U})] = \mathbb{E}[\Psi(\mathbf{V})w(\mathbf{V})]$ for some weight function $w : [0, 1]^d \rightarrow [0, \infty)$, and such that \mathbf{V} concentrates its samples in the region of interest for our specific problem. The paper details how to build the distribution $F_{\mathbf{V}}$ of this alternative random variable, as a mixture of copula distribution, and how to efficiently sample it (including closed-form sampling scheme for specific copula families for C). We measure the reduction in estimation error that allows this approach through case studies that illustrate typical insurance problems in realistic settings.

In Paper **C**, we apply Quasi Monte-Carlo (QMC) methods to our copula setting. The underlying idea of QMC methods is to replace a (pseudo) random sample $\{\mathbf{U}_i : i = 1, \dots, n\}$ by point $\{\mathbf{v}_i : i = 1, \dots, n\}$ from a deterministic sequence of points in the unit hypercube. In fact, it can be observed that random samples $\{\mathbf{U}_i : i = 1, \dots, n\}$ will inevitably show regions of $[0, 1]^d$ which are lacking points, and other areas which contain more samples than expected by the uniform distribution. To reduce this problem, quasi random number generators (QRNGs) do not aim at mimicking i.i.d samples but instead at producing a homogeneous coverage of $[0, 1]^d$. The homogeneity of a sequence of points over the unit hypercube can be measured by its discrepancy¹. The estimation error is thus also deterministic, and it can be shown that it satisfies

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{v}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq V(\Psi)D^*(\{\mathbf{v}_i : i = 1, \dots, n\}), \quad (1)$$

where $V(\Psi)$ denotes the total variation of the function Ψ and $D^*(\{\mathbf{v}_i : i = 1, \dots, n\})$ the discrepancy of the set. This inequality shows that the efficiency of a QMC method may actually strongly depend on certain properties of Ψ , that is not a feature of standard Monte-Carlo method. This paper is of particular relevance since QMC methods have been widely used for problems where the random variables of interest were independent. In such cases, the transformation of the uniform sample points into observations from the desired model can be easily obtained by transforming each component of the uniform sample points using the inverse transform method, which is deemed to work well with QMC in part because of its monotonicity, and also because it corresponds to an overall one-to-one transformation from $[0, 1]^d$ to \mathbb{R}^d . In our copula setting, the conditional distribution method (which is the inverse of the copula-based version of the Rosenblatt transform) appears to be a good choice to use with quasi random numbers, as it is the direct multivariate extension of the inverse transform. The paper details several of these approaches, with a particular emphasis on the CDM approach. In addition, the paper derives for these sampling methods smoothness

¹This relates to the error incurred by representing the Lebesgue-measure of subsets of the unit hypercube by the fraction of points in these subsets.

conditions on the function Ψ in order to satisfy the above bound for the error. The superiority of the QRNGs applied to our copula setting is over the pseudo random approach is also illustrated with several examples, including simulations addressing an application in the realm of insurance.

Replicating Portfolio Approach to Capital Calculation

Paper **D** deals with a challenging problem that arises mainly in life insurance. In particular for the calculation of regulatory capital, insurance companies need to model the profit and loss distribution of the asset-liability portfolio on a one-year time horizon while most of liabilities, such as life insurance contracts, have uncertain cash flows that run up to 40 years and beyond. The issue is therefore that no observable data would be available for deriving the change in value of these contracts over a one-year time horizon. For that, one could think of nesting Monte-Carlo samples of these cash-flow values over the years until contract termination. Since contract terms are typically long, nested simulations are highly computationally costly and in most cases not feasible. The replicating portfolio (RP) approach consists in projecting the discounted losses of the asset-liability portfolio, at terminal time T , onto a set of factors, usually generated by a set of financial assets, whose discounted profit and loss processes can be efficiently simulated.

Writing Z for the terminal discounted loss of the liability and V_t for the discounted value process of replicating assets, we propose to use

$$L_1 = V_1, \text{ and } L_2 = V_1 + Z - V_T,$$

as two possible approximations for the one-year discounted loss of the asset-liability portfolio.

When estimating capital requirements with such proxies, two main sources of error could materially deteriorate the accuracy of our estimate: the approximation error from the approximation of liability losses with financial instruments, and the Monte-Carlo error from the use of finite-sample estimates in order to derive capital that cannot be obtained in closed form. Paper **D** shows that these proxies are consistent in an asymptotic setting. In addition, the materialities of these two sources of error are compared. The current standard RP approach adopted by the industry is static, in the sense that the proxies are derived only from fixed time value. The insurance market is incomplete under static replication with financial assets, for two reasons. First, there are more factors driving insurance cash flows than there are traded financial instruments for their replication. Second, insurance cash flows are nonlinear in the underlying financial instruments. The paper shows that a dynamic and path-dependent RP significantly outperforms the industry standard static RP.

Statement of Originality

I hereby certify that the four papers of this thesis are the results of my own work, where some parts are the result of collaborations with my thesis supervisors Prof. Damir Filipović, as well as my co-authors Dr. Philipp Arbenz, Dr. Marius Hofert and Prof. Christiane Lemieux. No other person's work has been used without due acknowledgement.

Paper A

Model Uncertainty and Scenario Aggregation

Model Uncertainty and Scenario Aggregation*

Mathieu Cambou[†] Damir Filipović[‡]

May 15, 2015

Abstract

This paper provides a coherent method for scenario aggregation addressing model uncertainty. It is based on divergence minimization from a reference probability measure subject to scenario constraints. An example from regulatory practice motivates the definition of five fundamental criteria that serve as a basis for our method. Standard risk measures, such as value-at-risk and expected shortfall, are shown to be robust with respect to minimum divergence scenario aggregation. Various examples illustrate the tractability of our method.

Key words: model uncertainty, scenario aggregation, expected shortfall, value-at-risk, statistical divergence, Swiss Solvency Test

1 Introduction

The last decades have seen strong developments in the statistical measurement of risk. The quantitative methods used by banks and insurance companies for risk management serve many purposes such as enterprise risk management, pricing, capital allocation or reporting to regulators. The latter have required regulated institutions to implement and document internal models that should be used to report their amount of required capital which is bearing the risk and to show that they would remain solvent in case of extreme scenarios. Although the risk modeling methodology of an institution is reported and subject to approval from the regulators or internal model validation, uncertainty on the validity of the model remains inherent and should therefore be challenged. The risk of inappropriate modeling can usually be raised at many levels. For example, in a factor model, one could question a specific choice of risk factors, the marginal distribution of these risk factors or

*We thank Matthias Aellig, David Babel, Alexis Baily, Rama Cont, Freddy Delbaen, Kabir Dutta, Paul Embrechts, Jan Friedrich, Hansjörg Furrer, Jean-Francois Guérin, Andreas Haier, Stefan Jaschke, Thorsten Pfeiffer, Alexander Schied, Michael Schmutz, Ruodu Wang, three anonymous referees, the editor (Jerome Detemple), and participants at the FINMA Workshop on Scenario Aggregation 2012 in Bern, the 2013 ASTIN Colloquium in The Hague, the 2013 Workshop on Indices of Riskiness and New Risk Measures at ETH Zurich, the 2013 IMA Conference on Mathematics in Finance in Edinburgh, the De Finetti Risk Seminar in Milano, Vienna Graduate School of Finance, the 6th General AMaMeF and Banach Center Conference in Warsaw, RiskMinds Insurance 2014 in Amsterdam, the expert forum on Risk Measures and Regulation in Insurance 2014 in Zurich, the 8th World Congress of the Bachelor Finance Society 2014 in Brussels, the Swiss Re Dependence Day 2014 in Zurich, the Journées Actuarielles de Strasbourg 2014 for comments. As SCOR fellow, Mathieu Cambou thanks SCOR for financial support.

[†]EPFL, Institute of Mathematics, Station 8, EPFL, 1015 Lausanne, Switzerland
email: mathieucambou@gmail.com

[‡]EPFL and Swiss Finance Institute, Quartier UNIL-Dorigny, Extranef 218, 1015 Lausanne, Switzerland
email: damir.filipovic@epfl.ch

the dependence structure between them (see, e.g., [41, 43] for the last aspect). In many examples of risk management processes, the model will be used to estimate risk measures that greatly depend on the tail of the loss distribution and one should therefore check that this part of distribution is appropriately modeled.

This paper provides a coherent method for incorporating external views on scenarios into an internal model. This scenario aggregation method aims at supporting regulatory purposes, such as stress testing, and serves as a device to address internal model uncertainty.

A clear distinction has been made in the literature, in the footsteps¹ of [75], between the notions of *risk* and *uncertainty*. The former relates to the unknowns with respect to future events for which probabilities are known with certainty, while the latter notion happens whenever these probabilities are unknown. Although the wording *model risk* is broadly used in both the academic and industry literatures, it is usually meant to account for uncertainty and *model uncertainty* should be equivalently used. For an exhaustive review of the literature on quantifying model uncertainty, see [28, Section 2]. Our concern in this paper is the model uncertainty and the way it can be challenged using views on scenarios.

In the nineties, the method of stress testing was introduced by (for) risk managers to provide more understanding on the effect of (extreme) stress scenarios on portfolios of insurance or financial contracts. The definition of stress testing was originally discussed in [84, 121, 79]. These stress tests were not meant to quantify any model uncertainty as they were initially thought as a tool to identify scenarios on the risk factors state space that could lead the institution to insolvency and to evaluate their impact. It is only later, see, e.g, [11, 93, 113, 94, 81, 38], that probabilities were assigned to the stress test events and were folded into the reference model in order to account for uncertainty.

Although scenario aggregation can be used by any risk manager, or potentially any person that would like to challenge a model, it is a technique that regulators can leverage to formulate universal (i.e., non entity-specific) requirements on internal models. Currently, each regulator has its own view on how stress scenarios should be defined and whether they should be aggregated or not. On the banking side, the Basel Committee on Banking Supervision, see e.g, [8], requires banks to evaluate stress scenarios which can then be used to set associated capital charges, see e.g., [9]. On the insurance side, although Solvency II requires to evaluate stress scenarios, see [24], it does not require the aggregation step. On the other hand, the scenario aggregation is a predominant point of the Swiss Solvency Test (SST) implementation, see [52].

This paper provides a new scenario aggregation framework, with a particular emphasis on regulatory purposes. The mathematical framework is presented in Section 2. A practical example of an existing scenario aggregation method, the SST, is exposed in Section 3, which serves as a basis to shape important criteria that such a method should satisfy. In Section 4, we formally define the minimum divergence scenario aggregation method. We show in Section 5 that it is equivalent to solving a finite-dimensional convex optimization problem. We elaborate on the robustness of the standard risk measures with respect to this method in Section 6. In Section 7, we show how to solve the convex optimization problem associated and we give examples of some explicit solving. In Section 8 we study the asymptotic properties of the minimum divergence aggregation when the

¹“*The practical difference between the two categories, risk and uncertainty, is that in the former the distribution of the outcome in a group of instances is known (either through calculation a priori or from statistics of past experience), while in the case of uncertainty this is not true, the reason being in general that it is impossible to form a group of instances, because the situation dealt with is in a high degree unique.*”, see Chapter VIII in [75].

number of scenarios increases. An alternative method is presented in Section 9. For conciseness of the paper, all figures are reported in Appendix A and proofs in Appendix B.

2 Internal models

We place under the umbrella of *internal model* any model developed by an institution that is employed for risk management purposes. An internal model is formalized as a loss random variable L for a given time horizon defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which assigns to any possible state of the world $\omega \in \Omega$ a loss $L(\omega)$. The probability measure \mathbb{P} and the random variable L are institution specific. While Ω can be viewed as a universal object, one can argue that the σ -algebra \mathcal{F} is not, as different institutions may not access the same information. We assume that \mathcal{F} is universal, taking \mathcal{F} to be the σ -algebra generated by all institution specific σ -algebras and an appropriate extension of \mathbb{P} if needed. Typically, for solvency models in insurance and financial institutions, the time horizon is fixed to 1 year. Scenario aggregation will potentially result in modifying the probability measure \mathbb{P} , or alternatively, the loss distribution function $F_L(x) = \mathbb{P}[L \leq x]$.

We assume that the loss mapping $L : \Omega \rightarrow \mathbb{R}$ is not subject to model uncertainty. As L is industry specific, from a regulatory perspective, it would be more difficult to design unified and industry-wise supervision on L . Note however that the specification of L is an important issue as it depends on the underlying accounting standards, such as actuarial accounting standards for insurance products or mark-to-market accounting for financial securities. The financial crisis 2007–08 has shown that mark-to-market accounting has its limits in times of market illiquidity, see e.g. [14].

Remark 2.1. *The description of internal models is general enough so that it can fit both insurance and banks. Although this paper mainly tends to focus on the former, the model uncertainty which comes with the specification of \mathbb{P} arises in both industries.*

Once the internal model is specified, the notions of quantiles and risk measures are necessary for evaluating the risk of losses that the institution is exposed to. For a random variable X defined on $(\Omega, \mathcal{F}, \mathbb{P})$, denote by

$$q_\alpha^-(X) = \inf \{x : \mathbb{P}[X \leq x] \geq \alpha\} \leq \inf \{x : \mathbb{P}[X \leq x] > \alpha\} = q_\alpha^+(X),$$

the left and right quantiles, respectively, at level α . In addition, we recall the definition of the risk measures used for the required capital calculation. For more details and references on risk measures, we refer to [51, Chapter 4].

Remark 2.2. *In this paper, we follow the convention that the argument X refers to the loss and $-X$ to the profit, which is opposite in sign to [51].*

Definition 2.3. *The value-at-risk at level $\alpha \in (0, 1)$, $\text{VaR}_\alpha(X)$, is defined as the left quantile at level α of X , that is $\text{VaR}_\alpha(X) = q_\alpha^-(X)$.*

Remark 2.4. *The interpretation one must have of VaR_α is the level that the loss will not exceed with a probability α . In practice, if the loss distribution is sampled, then the (empirical) $\text{VaR}_{99\%}$ one will compute from, say, 100 realizations is the 99th largest observation. In that case, the (empirical) right quantile would be the 100th largest.*

Definition 2.5. *The expected shortfall at level $\alpha \in (0, 1)$, $\text{ES}_\alpha(X)$, is defined as*

$$\text{ES}_\alpha(X) = \frac{1}{1-\alpha} \mathbb{E}[(X - q_\alpha)^+] + q_\alpha,$$

for $q_\alpha \in [q_\alpha^-(X), q_\alpha^+(X)]$, any α -quantile of X .

We note that Definition 2.5 is independent of the choice of the α -quantile $q_\alpha \in [q_\alpha^-(X), q_\alpha^+(X)]$. The expected shortfall can also be defined as an integral on $(0, 1)$ of the quantile function. As it can be shown that this is equivalent to ours, see [51, Lemma 4.51], we will work with Definition 2.5 for convenience.

The scope of risk management within insurance and other financial institutions is large, see [89] as a quantitative reference. For the sake of conciseness and coherence with many regulatory obligations, we will discuss the calculation of capital that the institution must hold as a buffer against possible extreme losses over a fixed time horizon. This amount of capital can, for example, be computed via risk measures such as the value-at-risk at level α , $\text{VaR}_\alpha(L)$, or the expected shortfall at level α , $\text{ES}_\alpha(L)$. When this capital calculation is required by the regulators, it comes with a specific risk measure and level of confidence. For example, the *required capital* asked to be maintained by the SST is given by $\text{ES}_\alpha(L)$ with $\alpha = 99\%$, see [52], while the Solvency II directives advocate the use of $\text{VaR}_\alpha(L)$ with $\alpha = 99.5\%$. We will avoid the discussion of differences between the different notions of required capital defined by the different regulators, and we will speak in full generality of required capital as a function of a risk measure applied to L .

In this paper, we focus on the uncertainty in the choice of \mathbb{P} , through scenario aggregation and we discuss the impact of changing the internal model on the capital calculation. An example that has already been put in place by the Swiss Financial Market Supervisory Authority (FINMA) is discussed in the next section.

3 A practical example: Swiss Solvency Test

A practical example will help to shape the main criteria we shall use for designing a scenario aggregation method. Within the prudential framework set by the FINMA for the supervision of insurance companies lies the Swiss Solvency Test (SST), see [52] for technical documentation. One SST requirement is to evaluate a given list of d scenarios that have a small probability of occurrence, and that would have a significant effect on the annual loss L . Each scenario is a narrative description of a potential, and typically extreme, event. This includes, market crashes, pandemic or natural catastrophes, etc. If d scenarios are prescribed, scenario i comes with an auxiliary probability weight c_i set subjectively by the regulator, for $i = 1, \dots, d$. Each scenario causes an extra-ordinary loss $z_i = \mathbb{E}[L \mid \text{Scenario } i] - \mathbb{E}[L]$ that has to be calculated by the institution's actuary. The stressed loss distribution function conditional on scenario i is then set to be $F_L(x - z_i)$. The scenario aggregation is done via mixing of the shifted distributions, which leads to the following alternative loss distribution

$$F_L^{\text{SST}}(x) = c_0 F_L(x) + \sum_{i=1}^d c_i F_L(x - z_i),$$

where $c_0 = 1 - \sum_{i=1}^d c_i$ is the implied probability weight for a “normal” unstressed year. We note that the scenario aggregation is on the level of the loss distribution F_L , not on \mathbb{P} . In some specific cases, the effect on \mathbb{P} is tractable. For example, when L is a linear function of a set of risk factors,

mixing shifted loss distributions is equivalent to mixing shifted risk factor distributions, see [57, Lemma 1].

By the law of total probability it follows that $F_L^{\text{SST}}(x)$ is the distribution function of $L + Z$, where Z is a discrete random variable, independent of L , that takes values $z_0 = 0, z_1, \dots, z_d$ with probabilities $c_i, i = 0, \dots, d$, respectively. Using this representation, the following lemma gives bounds on $\text{ES}_\alpha(L + Z)$.

Lemma 3.1. *Let X be an integrable random variable independent of L . Then we have*

$$\text{ES}_\alpha(L) + \mathbb{E}[X] \leq \text{ES}_\alpha(L + X) \leq \text{ES}_\alpha(L) + \text{ES}_\alpha(X).$$

The method described above on the incorporation of scenarios in the loss distribution raises a few questions and motivates the following criteria that would be meaningful from a regulatory point of view. These criteria tend to follow an overall regulatory rationale and they should be naturally encoded in the mathematical method for scenario aggregation:

- (1) *No penalty for conservative internal models.* The SST scenario aggregation method presented in this section will tend to penalize any model to which it is applied, including the most conservative ones. Indeed, we have from Lemma 3.1 that $\text{ES}_\alpha(L + Z) > \text{ES}_\alpha(L)$ as long as $\mathbb{E}[Z] > 0$, so that the required capital is increased even if scenarios are not tail loss events for the insurer. The intent of a regulatory prudential framework is rather to control and validate internal models, and to apply penalties only if it is not judged satisfactory according to the regulator.
- (2) *Focus on tail loss events.* The scenario impact in the SST method is modeled as a shift of the loss distribution, regardless whether or not the scenario is a tail loss event for the insurer. As capital requirements are calculated according to (tail-)risk measures, see Section 2, the tail of the loss distribution should require more scrutiny than its mean.
- (3) *Control over distance from internal model.* In order to be able to give weight to the internal model, one needs to control how far the alternative internal model will be from the original one. In the SST method, there is no control, a priori, on how far $F_L^{\text{SST}}(x)$ is from $F_L(x)$.
- (4) *Robustness of capital requirements.* In order to ensure the stability of the industry, capital requirements should be robust with respect to modifications of the internal model by scenario aggregation. In the SST method, there is no control, a priori, on the robustness of capital requirements with respect to scenario aggregation.
- (5) *Tractability.* It is understood that any scenario aggregation method to be implemented in practice should be computationally tractable, such as is the case for the SST method.

In the next section, we propose a new method for scenario aggregation that incorporates the above criteria.

4 Scenarios, views and aggregation

The point of departure is a moment's reflection on stress tests. A stress test is based on a selected generic state of the world $\hat{\omega} \in \Omega$, which results in a specific loss of $\ell = L(\hat{\omega})$. Interpreted as

null hypothesis, the internal model \mathbb{P} is said to pass the stress test if it is not rejected based on the “observed” loss ℓ for a pre-specified significance level of c , e.g. $c = 1\%$. In terms of the tail loss scenario $S = \{L \geq \ell\}$, the internal model is accepted whenever it satisfies $\ell \leq q_{1-c}^+(L)$, or equivalently $\mathbb{P}[S] \geq c$.²

As ℓ could possibly be unobserved, one can interpret this requirement as an expert *view* on the probability of occurrence of scenario S . It means that the internal model must be conservative enough that the probability assigned to the event of a loss larger than or equal to ℓ be at least c . If \mathbb{P} is rejected, it needs to be replaced by some alternative probability measure \mathbb{Q} which satisfies the view, $\mathbb{Q}[S] \geq c$. This leads to the following formalization, which is in the spirit of [57, Definition 3].

Definition 4.1. *A view on a scenario $S \in \mathcal{F}$ is a requirement of the form $\mathbb{Q}[S] \geq c$, for some target probability $c \geq 0$, to be satisfied by a $\mathbb{Q} \in \mathcal{M}$, where \mathcal{M} denotes the set of all probability measures on (Ω, \mathcal{F}) .*

We shall make a distinction between generic scenarios $S \in \mathcal{F}$ and insurer specific tail loss scenarios which are of the form $S = \{L \geq L(\hat{\omega})\}$ for some generic sample point $\hat{\omega} \in \Omega$. Although the input $\hat{\omega}$ of a tail loss scenario is generic, its definition is specific to the insurer as it involves the evaluation of the loss mapping L . The aforementioned stress test boils down to the view $\mathbb{P}[L \geq L(\hat{\omega})] \geq c$, for the target probability c equal to the significance level of the test, to be satisfied by the internal model \mathbb{P} . Whenever the internal model \mathbb{P} satisfies the view for a target probability $c \geq 1 - \alpha$, and is such that $q_{\alpha}^-(L) = q_{\alpha}^+(L)$, then

$$\text{VaR}_{\alpha}(L) \geq L(\hat{\omega}).$$

Hence the internal model passes the stress test only if the required capital computed as value-at-risk $\text{VaR}_{\alpha}(L)$ can absorb the loss $L(\hat{\omega})$ resulting from the sample point $\hat{\omega}$.³

The following lemma gives an analogous interpretation of a view in terms of expected shortfall.

Lemma 4.2. *Let $S \in \mathcal{F}$ be a scenario with target probability $c \geq 1 - \alpha$. If \mathbb{P} satisfies the view $\mathbb{P}[S] \geq c$, then*

$$\text{ES}_{\alpha}(L) \geq \mathbb{E}[L|S].$$

Hence the required capital computed as expected shortfall $\text{ES}_{\alpha}(L)$ can absorb the conditional expected loss $\mathbb{E}[L|S]$ given scenario S .

From a regulator point of view, the motivation for a scenario aggregation requirement is to discipline regulated institutions and to add a supervisory check on their internal models. In consequence, it is important to treat equally all regulated institutions. So it is reasonable to require that the scenarios to evaluate should be specified by universal inputs, without ambiguity. Views on such scenarios can thus be reproduced and verified by any external third party. As views on scenarios can be calibrated to exogenous expert opinion, they provide a natural device for addressing model uncertainty. For example, experts may predict joint extreme events that would lead to higher joint-tail probabilities than originally calibrated by the model. Typical examples of scenarios for factor

²For more details on hypothesis testing, see, e.g., [33, Section 7.3].

³Bank stress tests performed in most countries do indeed focus on capital adequacy. The adverse scenarios $\hat{\omega}$ are typically provided without an explicit assessment of probabilities. Nonetheless, the criterion for passing a bank stress test is based on capital adequacy, which implies implicitly assumed tail probabilities.

models are quadrants of the form $S = \{\mathbf{x} \in \mathbb{R}^n \mid x_j \geq y_j \text{ for some } j\}$, for some threshold $\mathbf{y} \in \mathbb{R}^n$. These are joint events and thus address the uncertainty of modeling the dependence of the risk factors. Some recent articles have focused on more systematic approaches in choosing scenarios, see, e.g., [17, 91, 16].

We now propose a new scenario aggregation method. Given are d possibly non-disjoint scenarios $S_1, \dots, S_d \in \mathcal{F}$ along with target probabilities $c_1, \dots, c_d \geq 0$. In line with Definition 4.1 this induces the views

$$\mathbb{Q}[S_i] \geq c_i, \quad i = 1, \dots, d, \quad (1)$$

to be satisfied by a $\mathbb{Q} \in \mathcal{M}$. If the internal model \mathbb{P} does not satisfy the views (1), it needs to be replaced by some alternative model $\mathbb{Q} \in \mathcal{M}$ which satisfies the views (1). The objective is to choose \mathbb{Q} as close as possible to \mathbb{P} in \mathcal{M} . This applies under the premise that the internal model, formalised through \mathbb{P} , has been designed to the best of the risk management's knowledge, see Remark 5.6.

In order to measure how far \mathbb{Q} is from \mathbb{P} , we use the concept of statistical divergence which has been well established in statistics and particularly information theory for more than six decades.

Definition 4.3. *Let $\varphi(t)$ be a real-valued continuous convex function on $[0, \infty)$ which is not linear in any neighborhood of $t = 1$ and with $\varphi(1) = 0$. The φ -divergence of \mathbb{Q} from \mathbb{P} is then defined by*

$$d(\mathbb{Q}, \mathbb{P}) = \begin{cases} \mathbb{E} \left[\varphi \left(\frac{d\mathbb{Q}}{d\mathbb{P}} \right) \right] & \text{if } \mathbb{Q} \ll \mathbb{P}, \\ +\infty & \text{else.} \end{cases}$$

Examples of divergence functions $\varphi(t)$ and their associated φ -divergences include

$$\varphi(t) = \begin{cases} t \log t & \text{for the relative entropy } d_E(\mathbb{Q}, \mathbb{P}), \\ (\sqrt{t} - 1)^2 & \text{for the Hellinger divergence } d_H(\mathbb{Q}, \mathbb{P}), \\ |t - 1|^p & \text{for the } L^p \text{ divergence } d_{L^p}(\mathbb{Q}, \mathbb{P}), p \geq 1. \end{cases} \quad (2)$$

Note that $d(\mathbb{Q}, \mathbb{P}) \neq d(\mathbb{P}, \mathbb{Q})$, and hence $d(\mathbb{Q}, \mathbb{P})$ is not a metric, in general. The following lemma provides some other facts about φ -divergences. For more background we refer to [3, 31, 77, 83].

Lemma 4.4. (i) $d(\mathbb{Q}, \mathbb{P}) \geq 0$, and $d(\mathbb{Q}, \mathbb{P}) = 0$ if and only if $\mathbb{Q} = \mathbb{P}$.

(ii) $d(\mathbb{Q}, \mathbb{P})$ is convex in \mathbb{Q} , and strictly convex if $\varphi(t)$ is strictly convex on $(0, \infty)$. Hence all but the L^1 -divergence among the examples in (2) are strictly convex in \mathbb{Q} .

(iii) The following inequalities hold

$$\begin{aligned} d_{L^1}(\mathbb{Q}, \mathbb{P}) &\leq \sqrt{2d_E(\mathbb{Q}, \mathbb{P})}, \\ d_H(\mathbb{Q}, \mathbb{P}) &\leq d_{L^1}(\mathbb{Q}, \mathbb{P}) \leq \sqrt{2d_H(\mathbb{Q}, \mathbb{P})}. \end{aligned}$$

(iv) For $\mathbb{Q} \ll \mathbb{P}$ we have $d_{L^1}(\mathbb{Q}, \mathbb{P}) = 2 d_{TV}(\mathbb{Q}, \mathbb{P})$ where

$$d_{TV}(\mathbb{Q}, \mathbb{P}) = \sup_{B \in \mathcal{F}} |\mathbb{Q}[B] - \mathbb{P}[B]|$$

denotes the total variation distance between \mathbb{Q} and \mathbb{P} , which is a metric on \mathcal{M} .

Aggregating the views on the scenarios S_i with the internal model \mathbb{P} can now be defined as finding a solution of the convex optimization problem

$$\begin{aligned} & \text{minimize} && d(\mathbb{Q}, \mathbb{P}) \\ & \text{subject to} && \text{views (1)} \end{aligned} \tag{3}$$

with domain $\mathbb{Q} \in \mathcal{M}$. This is an infinite-dimensional problem. It is a priori not clear whether the proposed minimum divergence scenario aggregation is actually tractable. In the following sections we show that it can be efficiently solved and that it complies with the criteria listed in Section 3.

5 Minimum divergence problem

In this section we show that the infinite-dimensional problem (3) can be efficiently solved under mild conditions on the target probabilities c_i . Thereto, we denote by $\sigma(S_1, \dots, S_d)$ the σ -algebra generated by the scenarios S_i . Let U_1, \dots, U_n be the \mathbb{P} -atoms of $\sigma(S_1, \dots, S_d)$. That is, U_1, \dots, U_n are mutually disjoint with $U_j \in \sigma(S_1, \dots, S_d)$ and $\mathbb{P}[U_j] > 0$ for all j , and $\cup_{j=1}^n U_j = \Omega$ \mathbb{P} -a.s. Moreover, for every i there exists an index set $J(i) \subset \{1, \dots, n\}$ such that $S_i = \cup_{j \in J(i)} U_j$ \mathbb{P} -a.s. We define the convex set

$$\mathcal{Q} = \left\{ \mathbb{Q} \ll \mathbb{P} \mid \frac{d\mathbb{Q}}{d\mathbb{P}} \text{ is } \sigma(S_1, \dots, S_d)\text{-measurable} \right\} \subset \mathcal{M}.$$

The density $\frac{d\mathbb{Q}}{d\mathbb{P}}$ of any measure $\mathbb{Q} \in \mathcal{Q}$ is constant on the atoms U_j . Hence \mathbb{Q} can be identified with the vector $\mathbf{q} = (q_1, \dots, q_n)^\top$ of probabilities $q_j = \mathbb{Q}[U_j] \geq 0$ through

$$\mathbb{Q}[B] = \sum_{j=1}^n \frac{q_j}{p_j} \mathbb{P}[B \cap U_j], \quad B \in \mathcal{F} \tag{4}$$

where we write $\mathbf{p} = (p_1, \dots, p_n)^\top$ for the vector of probabilities $p_j = \mathbb{P}[U_j] > 0$. In compact notation the views (1) then read $\sum_{j \in J(i)} q_j \geq c_i$, or in matrix form $A\mathbf{q} \geq \mathbf{c}$ for the $d \times n$ -matrix A defined as $A_{ij} = 1$ if $j \in J(i)$ and 0 otherwise.

The following theorem is the main result in this section. It provides necessary and sufficient conditions for the feasibility of the minimum divergence scenario aggregation.

Theorem 5.1. *The following statements are equivalent:*

- (i) *the convex optimization problem (3) admits a solution in \mathcal{M} ;*
- (ii) *the convex optimization problem (3) admits a solution in \mathcal{Q} ;*
- (iii) *there exists some $\mathbb{Q} \in \mathcal{M}$ that satisfies the views (1) and $\mathbb{Q} \ll \mathbb{P}$;*
- (iv) *there exists some $\mathbf{q} \in [0, \infty)^n$ with $\mathbf{1}^\top \mathbf{q} = 1$ and $A\mathbf{q} \geq \mathbf{c}$.*

In either case, there exists a solution \mathbf{q}^* of the convex optimization problem

$$\begin{aligned} & \text{minimize} && \sum_{j=1}^n p_j \varphi \left(\frac{q_j}{p_j} \right) \\ & \text{subject to} && A\mathbf{q} \geq \mathbf{c} \\ & && \mathbf{1}^\top \mathbf{q} = 1 \end{aligned} \tag{5}$$

with domain $\mathbf{q} \in [0, \infty)^n$. Any such solution \mathbf{q}^* determines through (4) a solution $\mathbb{Q}^* \in \mathcal{Q}$ of the convex optimization problem (3). If $\varphi(t)$ is strictly convex on $(0, \infty)$ then \mathbb{Q}^* is the unique solution of (3) in \mathcal{M} .

Theorem 5.1 shows that our scenario aggregation method is efficiently implementable via solving the finite-dimensional convex optimization problem (5). We will further elaborate on this in Section 7. The dimension of problem (5) is bounded in terms of the number of scenarios as $n \leq 2^d$, see (14) in Section 9.

Remark 5.2. *In practice one often encounters simulation based internal models. Formally, a simulation based model consists of a finite number M of simulated sample points $\omega_i \in \Omega$ and associated weights $\mathbb{P}[\omega_i] > 0$, for example obtained from an economic scenario generator. This induces a probability measure on (Ω, \mathcal{F}) through*

$$\mathbb{P}[B] = \sum_{\omega_i \in B} \mathbb{P}[\omega_i], \quad B \in \mathcal{F}.$$

Typically, one would have equally weighted sample points, $\mathbb{P}[\omega_i] = 1/M$. According to (4) and Theorem 5.1, we deduce that minimum divergence scenario aggregation boils down to reweighting the sample points.

We now discuss two extreme cases. First, if the internal model \mathbb{P} itself satisfies the views (1), it is unaffected by the minimum divergence scenario aggregation.

Corollary 5.3. *If \mathbb{P} satisfies the views (1) then $\mathbb{Q}^* = \mathbb{P}$ is the unique solution of (3).*

The other extreme case is if \mathbb{P} is incompatible with the views (1).

Corollary 5.4. *If $\mathbb{P}[S_i] = 0$ for some scenario S_i with positive target probability $c_i > 0$ then there exists no solution of (3).*

In the situation of Corollary 5.4, the internal model \mathbb{P} is to be rejected by the regulator, and to be recalibrated by the risk manager until $\mathbb{P}[S_i] > 0$ for any scenario S_i with positive target probability $c_i > 0$. In the specific case of a simulation based internal model, see Remark 5.2, this would mean producing additional sample points over the state space until all scenarios with positive target probability contain at least one of them. Alternatively, one could adopt a scenario aggregation method that works without the assumption that $\mathbb{P}[S_i] > 0$ for any scenario S_i with positive target probability $c_i > 0$. The point-mass aggregation method exposed in Section 9 has this feature.

Note that the converse of Corollary 5.4 is not true, in general. Indeed, let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ consist of three sample points. Consider the $d = 2$ scenarios $S_1 = \{\omega_1, \omega_2\}$ and $S_2 = \{\omega_2, \omega_3\}$ with target probabilities $c_1 = c_2 = 2/3$. These two scenarios are non-disjoint, $S_1 \cap S_2 = \{\omega_2\}$. We assume that $\mathbb{P}[\omega_j] > 0$ for $j = 1, 3$, and $\mathbb{P}[\omega_2] = 0$. Hence $\mathbb{P}[S_1] > 0$ and $\mathbb{P}[S_2] > 0$, but there exists no solution of (3). Indeed, any $\mathbb{Q} \in \mathcal{M}$ with $\mathbb{Q} \ll \mathbb{P}$ that satisfies the views (1) would imply $\mathbb{Q}[\omega_1] = \mathbb{Q}[\omega_3] = 2/3$, which is absurd. Albeit there exists no solution of (3), defining $\mathbb{Q}[\omega_j] = 1/3$ for all $j = 1, 2, 3$ shows that there exists some $\mathbb{Q} \in \mathcal{M}$ that satisfies the views (1).

The following example shows that the strict convexity condition in Theorem 5.1 cannot be relaxed in general without losing uniqueness of (3) in \mathcal{M} .

Example 5.5. We consider the L^1 divergence, which is not strictly convex on $(0, \infty)$. Assume we are given $d = 1$ scenario S with $\mathbb{P}[S] > 0$ and target probability $\mathbb{P}[S] \leq c \leq 1$. For any $\mathbb{Q} \in \mathcal{Q}$, we use Lemma 4.4(iv) and the characterisation (4) to write

$$\begin{aligned} d_{L^1}(\mathbb{Q}, \mathbb{P}) &= 2 \sup_{B \in \mathcal{F}} |\mathbb{Q}[B] - \mathbb{P}[B]| = 2 \sup_{B \in \mathcal{F}} \left| \frac{q}{\mathbb{P}[S]} \mathbb{P}[B \cap S] + \frac{1-q}{1-\mathbb{P}[S]} \mathbb{P}[B \cap S^c] - \mathbb{P}[B] \right| \\ &\leq 2 \sup_{B \in \mathcal{F}} \left| \frac{q}{\mathbb{P}[S]} \mathbb{P}[B \cap S] - \mathbb{P}[B \cap S] \right| = 2 \left(\frac{q}{\mathbb{P}[S]} - 1 \right) \sup_{B \in \mathcal{F}} \mathbb{P}[B \cap S] \\ &= 2(q - \mathbb{P}[S]). \end{aligned}$$

Noting that the upper bound is attained when $B = S$, we conclude that $d_{L^1}(\mathbb{Q}, \mathbb{P}) = 2(q - \mathbb{P}[S])$. It follows that a solution of (5) is given by $\mathbf{q}^* = (c, 1 - c)^\top$. In consequence, a solution $\mathbb{Q}^* \in \mathcal{Q}$ of (3) is such that

$$\frac{d\mathbb{Q}^*}{d\mathbb{P}} = \begin{cases} c/\mathbb{P}[S] & \text{on } S \\ (1-c)/\mathbb{P}[S^c] & \text{on } S^c. \end{cases}$$

Assume now there exists a measurable partition of $S = B \cup C$ such that $\mathbb{P}[B] = \mathbb{P}[C] = \mathbb{P}[S]/2$. Define $\tilde{\mathbb{Q}}^* \in \mathcal{M}$ through

$$\frac{d\tilde{\mathbb{Q}}^*}{d\mathbb{P}} = \begin{cases} 2c/\mathbb{P}[S] - 1 & \text{on } B \\ 1 & \text{on } C \\ (1-c)/\mathbb{P}[S^c] & \text{on } S^c. \end{cases}$$

Then $\tilde{\mathbb{Q}}^* \notin \mathcal{Q}$ and $d_{L^1}(\tilde{\mathbb{Q}}^*, \mathbb{P}) = 2(c - \mathbb{P}[S])$, which shows that uniqueness of the solution \mathbb{Q}^* of (3) in \mathcal{M} does not hold. \square

Remark 5.6. *Does the minimum divergence scenario aggregation method give incentives for gamming the rules? An institution could modify its internal model on the sole purpose of satisfying the views (1) while reducing its required capital. Suppose an institution with a continuous loss distribution $F_L(x) = \mathbb{P}[L \leq x]$ is given one tail loss scenario $S = \{L \geq q\}$ with target probability $1 - \alpha$, where $q = q_\alpha^+(L)$. By definition of $q_\alpha^+(L)$, the internal model \mathbb{P} already satisfies the view $\mathbb{P}[S] \geq 1 - \alpha$. Hence it is unaffected by the minimum divergence scenario aggregation (Corollary 5.3). The institution could decide to aggressively tune its model towards an alternative model \mathbb{P}' where $\mathbb{P}'[L \leq x]$ coincides with F_L on $(-\infty, q_\alpha^+(L))$ and $\mathbb{P}'[L \leq q_\alpha^+(L)] = 1$. The α -quantile of the loss distribution under \mathbb{P}' is equal to q . Hence \mathbb{P}' still satisfies the view $\mathbb{P}'[S] \geq 1 - \alpha$, and is such that*

$$\text{ES}_\alpha^{\mathbb{P}'}(L) = q < \text{ES}_\alpha(L).$$

The tuned model \mathbb{P}' thus yields a strictly smaller required capital than the initial model \mathbb{P} . However, institutions are not tempted to do so, as they leverage their internal model for many core business functions (such as pricing or enterprise risk management) and internal models are subject to model validation processes (either internally, externally, or both). In consequence, it is reasonable to assume that the institution will build its internal model using its best knowledge and assessment of the risks.

Before solving the finite-dimensional convex optimization problem (5), we elaborate on the robustness of the capital requirement with respect to the proposed scenario aggregation method in the following section.

6 Robustness

The notion of robustness in the specific context of risk management refers to the sensitivity of risk measures to changes in the underlying stochastic model or in the data used for estimation. Robustness has now become a key challenge of risk-based regulation, see, e.g., [23, Section 8] and [42]. In our setting, we shall investigate the impact of minimum divergence modifications of the internal model \mathbb{P} on the required capital.

We consider a sequence of probability measures \mathbb{P}_n in \mathcal{M} . We shall denote by $\text{VaR}_{\alpha}^{\mathbb{P}_n}(X)$ and $\text{ES}_{\alpha}^{\mathbb{P}_n}(X)$ the value-at-risk and expected shortfall of X under \mathbb{P}_n . We start with an auxiliary result that implies the lack of robustness of value-at-risk.

Lemma 6.1. *Let $X \in L^0$. If $d_{L^1}(\mathbb{P}_n, \mathbb{P}) \rightarrow 0$ then the \mathbb{P}_n -distribution functions of X converge in Kolmogorov distance,*

$$\sup_{x \in \mathbb{R}} |\mathbb{P}_n[X \leq x] - \mathbb{P}[X \leq x]| \rightarrow 0, \quad (6)$$

and

$$q_{\alpha}^{-}(X) \leq \liminf_n q_n \leq \limsup_n q_n \leq q_{\alpha}^{+}(X)$$

holds for any sequence q_n of α -quantiles of X with respect to \mathbb{P}_n . Hence, if $q_{\alpha}^{-}(X) = q_{\alpha}^{+}(X)$ then

$$\text{VaR}_{\alpha}^{\mathbb{P}_n}(X) \rightarrow \text{VaR}_{\alpha}(X).$$

Note that Lemma 6.1 boils down to the property of lower (upper) semi-continuity of the left (right)-quantile with respect to the Kolmogorov distance, see also [25]. The following example illustrates that value-at-risk is not robust in general if left and right α -quantiles differ.

Example 6.2. Let us define $X = 0$ or 1 with $\mathbb{P}[X = 0] = \alpha$ and the sequence of probability measures \mathbb{P}_n given by

$$\frac{d\mathbb{P}_n}{d\mathbb{P}} = \begin{cases} 1 + (1 - \alpha)(-1)^n/(\alpha n) & \text{on } \{X = 0\} \\ 1 + (-1)^{n+1}/n & \text{on } \{X = 1\}. \end{cases}$$

It follows by inspection that $d_{L^{\infty}}(\mathbb{P}_n, \mathbb{P}) \rightarrow 0$. However, the sequence of respective value-at-risk's

$$\text{VaR}_{\alpha}^{\mathbb{P}_n}(X) = \begin{cases} 0 = q_{\alpha}^{-}(X), & \text{for } n \text{ even} \\ 1 = q_{\alpha}^{+}(X), & \text{for } n \text{ odd} \end{cases}$$

does not converge to $\text{VaR}_{\alpha}(X) = q_{\alpha}^{-}(X) = 0$. □

This situation is likely to occur in simulation based models, which are commonly used in practice, see Remark 5.2 and [116].

Example 6.3. Consider a simulation based internal model consisting of M simulated sample points $\omega_i \in \Omega$ with equal weights, $\mathbb{P}[\omega_i] = 1/M$, and such that $L(\omega_1) \leq \dots \leq L(\omega_M)$. Assume that αM is integer, so that $\text{VaR}_{\alpha}(L) = q_{\alpha}^{-}(L) = L(\omega_{\alpha M})$. Suppose that a reweighting of the sample points due to scenario aggregation leads to the scaling up of $\mathbb{P}[L \geq L(\omega_{\alpha M+2})]$ to $1 - \alpha$, or equivalently a scaling down of $\mathbb{P}[L \leq L(\omega_{\alpha M+1})]$ to α . The value-at-risk at level α becomes $L(\omega_{\alpha M+1})$, which can be an arbitrarily large increase from $L(\omega_{\alpha M})$, see Figure 1.

In contrast to value-at-risk, expected shortfall is always robust with respect to minimum L^p -divergence modifications of \mathbb{P} .

Theorem 6.4. *Let $p \in [1, \infty)$ and $r \in (1, \infty]$ such that $p^{-1} + r^{-1} = 1$. If $d_{L^p}(\mathbb{P}_n, \mathbb{P}) \rightarrow 0$ then*

$$\text{ES}_\alpha^{\mathbb{P}_n}(X) \rightarrow \text{ES}_\alpha(X), \quad \text{for all } X \in L^r.$$

In view of Lemma 4.4(iii) the following corollary is immediate.

Corollary 6.5. *If $d_E(\mathbb{P}_n, \mathbb{P}) \rightarrow 0$ or $d_H(\mathbb{P}_n, \mathbb{P}) \rightarrow 0$ then*

$$\text{ES}_\alpha^{\mathbb{P}_n}(X) \rightarrow \text{ES}_\alpha(X), \quad \text{for all } X \in L^\infty.$$

To summarize, the capital requirement based on expected shortfall is robust with respect to minimum divergence modifications of the internal model. The same holds for value-at-risk only if the loss distribution satisfies some continuity condition, which may fail in practice.

The robustness properties discussed in this section are weaker than the ones studied in [29, 76]. While we focus on the continuity of the risk measure as a function on the set of probability measures $\mathbb{Q} \ll \mathbb{P}$ with respect to L^p -divergence $d_{L^p}(\mathbb{Q}, \mathbb{P})$, the robustness in the above references boils down to continuity of the (law invariant) risk measure as a function on the set of distributions with respect to the Lévy distance, see [59] and [70, Theorem 2.21]. [29, Proposition 3.5] shows that value-at-risk at level α is continuous with respect to the Lévy distance at any distribution for which the left and right α -quantiles coincide. This is in line with Lemma 6.1, and does not contradict our non-robustness Example 6.2. [70, Theorem 3.7] and [29, Theorem 3.4] show that expected shortfall is not continuous with respect to the Lévy distance at any distribution, see [120, Section 2] for an example. This is compatible with our robustness results as the topology used in Theorem 6.4 and Corollary 6.5 is stronger than the one induced by the Lévy distance. If one splits the specification of a stochastic model in a dependence function (copula) and the marginal distributions, [43] show that expected shortfall is continuous with respect to the Lévy distance whenever the model uncertainty only concerns the dependence function. It has also been shown in [120, 46] that expected shortfall is continuous with respect to the Wasserstein distance.

7 Solving the optimization problem

The finite-dimensional convex optimization problem (5), and thus (3), can be efficiently solved through numerical algorithms. In this section, we derive the Karush–Kuhn–Tucker (KKT) conditions, which are first order necessary conditions for a solution to hold. Numerically solving the optimization problem (5) boils down to numerically solving the KKT system of equations, see [102]. From the KKT conditions we derive closed form solutions of (5) for the case of disjoint scenarios in general, and the case of two non-disjoint scenarios with relative entropy and L^2 -divergence.

7.1 KKT conditions

From now on, in addition to what is stated in Definition 4.3, we make the following standing assumption on the divergence function.

Assumption 1. *$\varphi(t)$ is differentiable on $(0, \infty)$.*

Assumption 1 implies that $\varphi'(t)$ is continuous on $(0, \infty)$, see [115, Corollary 25.5.1]. It is satisfied by all, but the L^1 -divergence, of the examples in (2).

The Lagrangian function corresponding to (5) is

$$L(\mathbf{q}, \boldsymbol{\lambda}, \nu) = \sum_{j=1}^n p_j \varphi\left(\frac{q_j}{p_j}\right) + \boldsymbol{\lambda}^\top (\mathbf{c} - A\mathbf{q}) + \nu (\mathbf{1}^\top \mathbf{q} - 1)$$

for Lagrange multiplier values $\boldsymbol{\lambda} \in [0, \infty)^n$ and $\nu \in \mathbb{R}$. The following lemma provides the KKT conditions. It follows from [115, Corollary 28.2.2 and Theorem 28.3] or [15, Sections 5.2.3 and 5.5.3].

Lemma 7.1. *Assume that Slater's condition is satisfied, i.e. there exists a $\mathbf{q} \in (0, \infty)^n$ such that $A\mathbf{q} \geq \mathbf{c}$ and $\mathbf{1}^\top \mathbf{q} = 1$. Then \mathbf{q}^* is a solution of (5) if and only if there exist Lagrange multiplier values $(\boldsymbol{\lambda}^*, \nu^*)$ which, together with \mathbf{q}^* , satisfy the KKT conditions*

$$\boldsymbol{\lambda} \geq \mathbf{0}, \quad A\mathbf{q} \geq \mathbf{c}, \quad \boldsymbol{\lambda}^\top (A\mathbf{q} - \mathbf{c}) = 0 \quad (7)$$

$$\mathbf{1}^\top \mathbf{q} = 1 \quad (8)$$

$$\varphi'\left(\frac{\mathbf{q}}{\mathbf{p}}\right) - A^\top \boldsymbol{\lambda} + \nu \mathbf{1} = \mathbf{0}, \quad (9)$$

where $\varphi'(\mathbf{q}/\mathbf{p})$ denotes the n -vector with components $\varphi'(q_i/p_i)$, $i = 1, \dots, n$.

7.2 Disjoint scenarios

A closed form solution of (3) can be derived from the KKT conditions for the case of disjoint scenarios.

Theorem 7.2. *Assume the scenarios S_i are mutually disjoint with $\mathbb{P}[S_i] > 0$ and $\cup_{i=1}^d S_i = \Omega$ \mathbb{P} -a.s. and such that, after re-ordering the indices if necessary, $c_1/p_1 \geq \dots \geq c_d/p_d$. Then $n = d$ and we can assume that $U_i = S_i$ for all i , so that A is the $d \times d$ identity matrix. A solution of (5) exists if and only if $\sum_{i=1}^d c_i \leq 1$. If $\sum_{i=1}^d c_i = 1$ then $\mathbf{q}^* = \mathbf{c}$ is the unique solution of (5). If $\sum_{i=1}^d c_i < 1$ then a solution \mathbf{q}^* is given by $q_i^* = \max\{c_i, p_i \mu^*\}$ with*

$$\mu^* = \frac{1 - \sum_{i=1}^{k^*-1} c_i}{\sum_{i=k^*}^d p_i}, \quad k^* = \min \left\{ 1 \leq k \leq d \mid \frac{c_k}{p_k} \leq \frac{1 - \sum_{i=1}^{k-1} c_i}{\sum_{i=k}^d p_i} \right\},$$

and where we write $\sum_{i=1}^0 c_i = 0$. In particular, this solution \mathbf{q}^* is independent of the specific choice of the divergence function $\varphi(t)$.

Theorem 7.2 provides a solution of (5), and thus (3), which does not depend on the choice of the divergence function $\varphi(t)$. However, Example 5.5 shows that this solution of (3) may not be unique in \mathcal{M} if $\varphi(t)$ is not strictly convex on $(0, \infty)$. The examples in Sections 7.5 and 7.6 below show that a solution of (3) does depend on the choice of $\varphi(t)$ for the general case of non-disjoint scenarios.

The case of one given scenario follows as a corollary.

Corollary 7.3. *Assume that $d = 1$ scenario $S \in \mathcal{F}$ is given with target probability $0 \leq c \leq 1$. Then a solution $\mathbb{Q}^* \in \mathcal{M}$ of the convex problem (3) is given by $\mathbb{Q}^* \in \mathcal{Q}$ such that*

$$\frac{d\mathbb{Q}^*}{d\mathbb{P}} = \begin{cases} \max\{1, c/\mathbb{P}[S]\} & \text{on } S \\ \min\{1, (1-c)/\mathbb{P}[S^c]\} & \text{on } S^c. \end{cases} \quad (10)$$

A special case of Corollary 7.3 is a stress test as described at the beginning of Section 4.

Corollary 7.4. *Assume that $d = 1$ tail loss scenario $S = \{L \geq \ell\}$ is given with target probability $c = 1 - \alpha$ for some $\alpha \in (0, 1)$ and loss level $\ell \in \mathbb{R}$. Then minimum divergence scenario aggregation results in an expected shortfall at level α of*

$$\text{ES}_\alpha^{\mathbb{Q}^*}(L) = \begin{cases} \text{ES}_\alpha(L), & \text{if } \mathbb{P}[S] \geq 1 - \alpha, \\ \text{ES}_{\mathbb{P}[L < \ell]}(L) = \mathbb{E}[L \mid L \geq \ell], & \text{otherwise.} \end{cases}$$

Corollary 7.4 emphasizes an interesting property. Assume the internal model \mathbb{P} does not pass the stress test defined by the loss ℓ at the significance level $1 - \alpha$, and it is replaced by $\mathbb{Q}^* \in \mathcal{M}$ the solution of the minimum divergence scenario aggregation method. Then the expected shortfall under \mathbb{Q}^* at level α can directly be calculated as expected shortfall under \mathbb{P} with level equal to $\mathbb{P}[L < \ell] > \alpha$. In the following subsection we provide a numerical example.

7.3 Example: Stress test

In order to avoid non-uniqueness issues we assume in this and the following example that the divergence function $\varphi(t)$ is strictly convex on $(0, \infty)$, see Theorem 5.1.

We consider a stress test as in the beginning of Section 4 with a tail loss scenario $S = \{L \geq \ell\}$ and associated target probability $c > 0$. We suppose that the loss random variable L has a Gaussian distribution with mean 0 and variance 3 under \mathbb{P} . Depending on ℓ and c , the minimum divergence aggregation results in an alternative model \mathbb{Q}^* .

The top panel of Figure 2 shows the values of $\text{VaR}_{99\%}^{\mathbb{Q}^*}(L)$ and $\text{ES}_{99\%}^{\mathbb{Q}^*}(L)$ for $c \in [0, 2\%]$ and $\ell \in \{\text{VaR}_\alpha(L) \mid \alpha \in [98\%, 99.9\%]\}$. For $\ell \leq \text{VaR}_{1-c}(L)$ we have $\mathbb{P}[L \geq \ell] = 1 - \alpha \geq c$, and hence $\mathbb{Q}^* = \mathbb{P}$ by Corollary 5.3. In particular, $\text{VaR}_{99\%}^{\mathbb{Q}^*}(L) = \text{VaR}_{99\%}(L)$ and $\text{ES}_{99\%}^{\mathbb{Q}^*}(L) = \text{ES}_{99\%}(L)$. For $\ell > \text{VaR}_{1-c}(L)$ we have $\mathbb{P}[L \geq \ell] = 1 - \alpha < c$, and \mathbb{P} is replaced by an alternative model $\mathbb{Q}^* \in \mathcal{M}$, the solution of the convex problem (3), which verifies $\mathbb{Q}^*[L \geq \ell] = c$ by Corollary 7.3. This implies that $\text{VaR}_{99\%}^{\mathbb{Q}^*}(L) > \text{VaR}_{99\%}(L)$ and $\text{ES}_{99\%}^{\mathbb{Q}^*}(L) > \text{ES}_{99\%}(L)$, see also Corollary 7.4.

The middle panel of Figure 2 shows the respective value-at-risk and expected shortfall values at level 99% for the SST aggregation method. The extra-ordinary losses $z = \mathbb{E}[L \mid L \geq \ell] - \mathbb{E}[L] > 0$ are all positive. As predicted by Lemma 3.1, the required capital is increased for all arguments (c, ℓ) , no matter whether the internal model passes the stress test or not. This contradicts criterion 1 in Section 3.

The bottom panel of Figure 2 shows the difference between capital requirements derived from the SST and the minimum divergence aggregation. The difference is positive for most but not all arguments (c, ℓ) . The comparison of the two aggregation methods in terms of capital requirements is also reported in Table 1 for $c = 0, 0.5\%, 1\%, 2\%$ and $\ell \in \{\text{VaR}_\alpha(L) \mid \alpha = 98\%, 99\%, 99.5\%, 99.9\%\}$.

Figure 3 shows the distribution function of L under \mathbb{Q}^* for $c = 1\%$ and three loss levels $\ell \in \{\text{VaR}_{99\%}(L), \text{VaR}_{99.5\%}(L), \text{VaR}_{99.9\%}(L)\}$. The 99%-quantiles of these distributions coincide with the respective loss level ℓ .

7.4 Example: Tail loss event

In this example the loss random variable is given as non-linear function

$$L = \max(X_1, -1) + \max(\min(X_2, 5), -1)$$

of two risk factors X_1 and X_2 . The loss is monotonic in both factors. The loss is capped in X_2 at 5 and the gain (negative loss) is floored in X_1 and X_2 at -1 . For example, X_1 could be related to an interest rate change and X_2 could be related to CAT events for which the insurer is re-insured. We assume that (X_1, X_2) has a bivariate Gaussian distribution under \mathbb{P} with mean zero, a correlation of -0.5 and variances of 1 and 4, respectively.

We consider two disjoint scenarios $S_1 = \{X_1 \geq 1, X_2 \geq 1\}$ and $S_2 = \{X_1 < -2\}$, with associated target probabilities c_1 and c_2 . As the insurer has uncapped losses in X_1 , scenario S_1 has a nonempty intersection with the tail loss event $W = \{L \geq \text{VaR}_{99\%}(L)\}$. Scenario S_2 and W are disjoint, as $L = \max(\min(X_2, 5), -1) - 1 \leq 4$ on S_2 and $\text{VaR}_{99\%}(L) > 4$. Figure 4 shows these events in the risk factor space.

The extra-ordinary losses for both scenarios are positive, $z_1 = \mathbb{E}[L|S_1] - \mathbb{E}[L] > 0$ and $z_2 = \mathbb{E}[L|S_2] - \mathbb{E}[L] > 0$. The SST aggregation of either scenario, S_1 or S_2 , results in an increase of the required capital, see Lemma 3.1. This is remarkable as scenario S_2 lies outside of the tail loss event W of the insurer. This contradicts criterion 2 in Section 3. In contrast, minimum divergence aggregation of S_2 does not lead to an increase of the required capital.

Figure 5 shows value-at-risk and expected shortfall at level 99% resulting from the minimum divergence aggregation (top), the SST aggregation (middle) and the difference between the two (bottom), for a range of target probabilities $c_1, c_2 \in [0, 4\%]$. The capital requirement for the SST aggregation method is strictly increasing in both arguments, c_1 and c_2 . As predicted by Theorem 7.2, the capital requirement for the minimum divergence aggregation method is constant in c_1 for $c_1 \leq \mathbb{P}[S_1]$ and increasing in c_1 for $c_1 > \mathbb{P}[S_1]$, and constant in c_2 for $c_2 \leq \mathbb{P}[S_2]$ and decreasing in c_2 for $c_2 > \mathbb{P}[S_2]$. The comparison of the two aggregation methods in terms of capital requirements is also reported in Table 2 for $c_1 = 0, p_1$ and 4%, $c_2 = 0, p_1$, and 4%, with $p_1 = 1.2\%$ and $p_2 = 2.3\%$.

7.5 Example: Two non-disjoint scenarios with relative entropy

In this and the following section, we explicitly derive the unique solution of problem (5) for the case of $d = 2$ non-disjoint scenarios S_1 and S_2 with relative entropy and L^2 divergence. We assume that $\mathbb{P}[S_1 \cap S_2] > 0$ and $\sum_{i=1}^2 \mathbb{P}[S_i] < 1$, and let c_1, c_2 be the associated target probabilities. There are $n = 4$ \mathbb{P} -atoms $U_1 = (S_1 \cup S_2)^c$, $U_2 = S_1 \cap S_2$, $U_3 = S_1 \setminus S_2$ and $U_4 = S_2 \setminus S_1$.

We first consider relative entropy, $\varphi(t) = t \log t$. The objective function in the optimization problem (5) is $\sum_{j=1}^4 q_j \log \frac{q_j}{p_j}$. In consequence, the KKT condition (9) becomes

$$\log \mathbf{q} - \log \mathbf{p} + \mathbf{1} - A^\top \boldsymbol{\lambda} + \nu \mathbf{1} = \mathbf{0}. \quad (11)$$

By inspection, we note that the first coordinate of (11) gives

$$q_1 = p_1 e^{-\nu-1},$$

while the third and fourth coordinates give

$$q_3 = \max\{c_1 - q_2, p_3 e^{-\nu-1}\}, \quad q_4 = \max\{c_2 - q_2, p_4 e^{-\nu-1}\},$$

and subsequently the second one gives

$$q_2 = e^{\nu+1} p_2 \max \left\{ \frac{c_1 - q_2}{p_3}, e^{-\nu-1} \right\} \max \left\{ \frac{c_2 - q_2}{p_4}, e^{-\nu-1} \right\}.$$

Using the KKT condition (8), we now solve for the values of $q_i^*, i = 1, \dots, 4$, that give the unique solution of (5). We do so by distinguishing between four different possible cases.

Case 1: $(c_1 - q_2)/p_3 \leq e^{-\nu-1}, (c_2 - q_2)/p_4 \leq e^{-\nu-1}$

We deduce that $\nu^* = -1$ and $q_i^* = p_i, i = 1, \dots, 4$. Hence Case 1 is equivalent to $\mathbf{p} \in I_1$ where

$$I_1 = \left\{ \mathbf{p} \in [0, \infty)^4 : c_1 \leq p_2 + p_3, c_2 \leq p_2 + p_4 \text{ and } \mathbf{1}^\top \mathbf{p} = 1 \right\}.$$

Case 2: $(c_1 - q_2)/p_3 > e^{-\nu-1}, (c_2 - q_2)/p_4 \leq e^{-\nu-1}$

We deduce that $\nu^* = -(1 + \log((1 - c_2)/(p_1 + p_3)))$ and

$$q_1^* = \frac{(1 - c_2)p_1}{p_1 + p_3}, \quad q_2^* = \frac{p_2 c_1}{p_2 + p_3}, \quad q_3^* = \frac{p_3 c_1}{p_2 + p_3}, \quad q_4^* = \frac{(1 - c_2)p_4}{p_1 + p_3}.$$

Hence Case 2 is equivalent to $\mathbf{p} \in I_2$ where

$$I_2 = \left\{ \mathbf{p} \in [0, \infty)^4 : c_1 > \frac{(1 - c_2)(p_2 + p_3)}{p_1 + p_3}, c_2 \leq \frac{c_1 p_2 (p_1 + p_3) + p_4 (p_2 + p_3)}{(1 - p_2)(p_2 + p_3)} \text{ and } \mathbf{1}^\top \mathbf{p} = 1 \right\}.$$

Case 3: $(c_1 - q_2)/p_3 \leq e^{-\nu-1}, (c_2 - q_2)/p_4 > e^{-\nu-1}$

We deduce that $\nu^* = -(1 + \log((1 - c_1)/(p_1 + p_4)))$ and

$$q_1^* = \frac{(1 - c_1)p_1}{p_1 + p_4}, \quad q_2^* = \frac{p_2 c_2}{p_2 + p_4}, \quad q_3^* = \frac{(1 - c_1)p_3}{p_1 + p_4}, \quad q_4^* = \frac{p_4 c_2}{p_2 + p_4}.$$

Hence Case 3 is equivalent to $\mathbf{p} \in I_3$ where

$$I_3 = \left\{ \mathbf{p} \in [0, \infty)^4 : c_1 \leq \frac{c_2 p_2 (p_1 + p_4) + p_3 (p_2 + p_4)}{(1 - p_2)(p_2 + p_4)}, c_2 > \frac{(1 - c_1)(p_2 + p_4)}{p_1 + p_4} \text{ and } \mathbf{1}^\top \mathbf{p} = 1 \right\}.$$

Case 4: $(c_1 - q_2)/p_3 > e^{-\nu-1}, (c_2 - q_2)/p_4 > e^{-\nu-1}$

Denote $\delta = p_1 p_2 - p_3 p_4$. In this case, q_2^* will be solution of the quadratic equation

$$(q_2^*)^2 \delta - q_2^* ((c_1 + c_2) \delta + p_3 p_4) + p_1 p_2 c_1 c_2 = 0.$$

and $q_1^* = 1 - c_1 - c_2 + q_2^*, q_3^* = c_1 - q_2^*, q_4^* = c_2 - q_2^*$. Case 4 is equivalent to $\mathbf{p} \in I_4$ where

$$I_4 = \left\{ \mathbf{p} \in [0, \infty)^4 : \mathbf{1}^\top \mathbf{p} = 1 \right\} \setminus (I_1 \cup I_2 \cup I_3).$$

7.6 Example: Two non-disjoint scenarios with L^2 divergence

We consider the same scenario setup as in the previous section, but now with L^2 divergence, $\varphi(t) = (t-1)^2$. The objective function in the optimization problem (5) is $\sum_{j=1}^4 p_j (q_j/p_j - 1)^2$. In consequence, the KKT condition (9) becomes

$$2 \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} - A^\top \boldsymbol{\lambda} + \nu \mathbf{1} = \mathbf{0}. \quad (12)$$

By inspection, we note that the first coordinate of (12) gives

$$q_1 = p_1 \left(1 - \frac{\nu}{2}\right),$$

while the third and fourth coordinates give

$$q_3 = \max \left\{ c_1 - q_2, p_3 \left(1 - \frac{\nu}{2}\right) \right\}, \quad q_4 = \max \left\{ c_2 - q_2, p_4 \left(1 - \frac{\nu}{2}\right) \right\},$$

and subsequently the second one gives

$$q_2 = p_2 \left(\max \left\{ \frac{c_1 - q_2}{p_3}, 1 - \frac{\nu}{2} \right\} + \max \left\{ \frac{c_2 - q_2}{p_4}, 1 - \frac{\nu}{2} \right\} - 1 + \frac{\nu}{2} \right).$$

Using the KKT condition (8), we now solve for the values of q_i^* , $i = 1, \dots, 4$, that give the unique solution of (5). We do so by distinguishing between four different possible cases.

Case 1: $(c_1 - q_2)/p_3 \leq 1 - \nu/2, (c_2 - q_2)/p_4 \leq 1 - \nu/2$

We deduce that $\nu^* = 0$ and that $q_i^* = p_i, i = 1, \dots, 4$. Hence Case 1 is equivalent to $\mathbf{p} \in I_1$ where

$$I_1 = \left\{ \mathbf{p} \in [0, \infty)^4 : c_1 \leq p_2 + p_3, c_2 \leq p_2 + p_4 \text{ and } \mathbf{1}^\top \mathbf{p} = 1 \right\}.$$

Case 2: $(c_1 - q_2)/p_3 > 1 - \nu/2, (c_2 - q_2)/p_4 \leq 1 - \nu/2$

We deduce from this case that $\nu^* = 2(1 - (1 - c_2)/(p_1 + p_3))$ and

$$q_1^* = \frac{(1 - c_2)p_1}{p_1 + p_3}, \quad q_2^* = \frac{p_2 c_1}{p_2 + p_3}, \quad q_3^* = \frac{p_3 c_1}{p_2 + p_3}, \quad q_4^* = \frac{(1 - c_2)p_4}{p_1 + p_3}.$$

Hence Case 2 is equivalent to $\mathbf{p} \in I_2$ where

$$I_2 = \left\{ \mathbf{p} \in [0, \infty)^4 : c_1 > \frac{(1 - c_2)(p_2 + p_3)}{p_1 + p_3}, c_2 \leq \frac{c_1 p_2 (p_1 + p_3) + p_4 (p_2 + p_3)}{(1 - p_2)(p_2 + p_3)} \text{ and } \mathbf{1}^\top \mathbf{p} = 1 \right\}.$$

Case 3: $(c_1 - q_2)/p_3 \leq 1 - \nu/2, (c_2 - q_2)/p_4 > 1 - \nu/2$

We deduce that $\nu^* = 2(1 - (1 - c_1)/(p_1 + p_4))$ and

$$q_1^* = \frac{(1 - c_1)p_1}{p_1 + p_4}, \quad q_2^* = \frac{p_2 c_2}{p_2 + p_4}, \quad q_3^* = \frac{(1 - c_1)p_3}{p_1 + p_4}, \quad q_4^* = \frac{p_4 c_2}{p_2 + p_4}.$$

Hence Case 3 is equivalent to $\mathbf{p} \in I_3$ where

$$I_3 = \left\{ \mathbf{p} \in [0, \infty)^4 : c_1 \leq \frac{c_2 p_2 (p_1 + p_4) + p_3 (p_2 + p_4)}{(1 - p_2)(p_2 + p_4)}, c_2 > \frac{(1 - c_1)(p_2 + p_4)}{p_1 + p_4} \text{ and } \mathbf{1}^\top \mathbf{p} = 1 \right\}.$$

Case 4: $(c_1 - q_2)/p_3 > 1 - \nu/2, (c_2 - q_2)/p_4 > 1 - \nu/2$

We deduce that

$$\nu^* = 2 \left(1 - \frac{1 - \left(c_1 \left(\frac{1}{p_2} + \frac{1}{p_4} \right) + c_2 \left(\frac{1}{p_2} + \frac{1}{p_3} \right) \right) \left(\frac{1}{p_2} + \frac{1}{p_3} + \frac{1}{p_4} \right)^{-1}}{p_1 + \left(\frac{1}{p_2} + \frac{1}{p_3} + \frac{1}{p_4} \right)^{-1}} \right),$$

and

$$q_1^* = p_1 \left(1 - \frac{\nu^*}{2} \right), \quad q_2^* = \frac{\frac{c_1}{p_3} + \frac{c_2}{p_4} - 1 + \frac{\nu^*}{2}}{\frac{1}{p_2} + \frac{1}{p_3} + \frac{1}{p_4}}, \quad q_3^* = c_1 - q_2^*, \quad q_4^* = c_2 - q_2^*.$$

Case 4 is equivalent to $\mathbf{p} \in I_4$ where

$$I_4 = \left\{ \mathbf{p} \in [0, \infty)^4 : \mathbf{1}^\top \mathbf{p} = 1 \right\} \setminus (I_1 \cup I_2 \cup I_3).$$

In both Examples 7.5 and 7.6 we obtain closed form solutions of (3). These solutions are given by case distinction. Inspection shows that the solution for the relative entropy is different from the solution for the L^2 divergence. This shows that a solution of (3) depends on the choice of $\varphi(t)$ in general.

8 Asymptotic scenario aggregation

In this section we study the asymptotic properties of the optimal solution of (3) when the number of scenarios increases, and the views (1) are derived from an auxiliary reference model $\mathbb{A} \in \mathcal{M}$. This model could encode the regulator's views on \mathcal{F} .

Let $\mathcal{C} = \{S_1, \dots, S_d\}$ denote the collection of scenarios. We say that \mathcal{C} is closed under complementation if $S^c \in \mathcal{C}$ for any $S \in \mathcal{C}$, and \mathcal{C} is closed under intersection if $S \cap S' \in \mathcal{C}$ for all $S, S' \in \mathcal{C}$. If \mathcal{C} is closed under complementation and intersection then it forms an algebra. The following lemma, which we state here for its own interest, serves as a preliminary finding in preparation of the asymptotic result.

Lemma 8.1. *Assume the target probabilities are given by $c_i = \mathbb{A}[S_i]$ for $i = 1, \dots, d$. Then the following properties hold.*

- (i) *If \mathcal{C} is closed under complementation then any $\mathbb{Q} \in \mathcal{M}$ satisfying the views (1) satisfies $\mathbb{Q} = \mathbb{A}$ on \mathcal{C} .*
- (ii) *If \mathcal{C} forms an algebra then any $\mathbb{Q} \in \mathcal{M}$ satisfying the views (1) satisfies $\mathbb{Q} = \mathbb{A}$ on $\sigma(\mathcal{C})$.*
- (iii) *If \mathcal{C} forms an algebra and $\mathbb{A} \ll \mathbb{P}$ then the measure $\mathbb{Q}^* \ll \mathbb{P}$ defined by $\frac{d\mathbb{Q}^*}{d\mathbb{P}} = \mathbb{E} \left[\frac{d\mathbb{A}}{d\mathbb{P}} \mid \sigma(\mathcal{C}) \right]$ is a solution of (3).*

Note that the condition in (i) cannot be relaxed in general. For example, consider the case of $d = 1$ scenario, $\mathcal{C} = \{S\}$, with $\mathbb{P}[S] > \mathbb{A}[S]$. Then $\mathbb{Q}^* = \mathbb{P}$ is the unique solution of (3) and does obviously not coincide with \mathbb{A} on \mathcal{C} .

By increasing the number d of scenarios the regulator can interpolate between the institution's internal model \mathbb{P} and his reference model \mathbb{A} . In the following theorem we show that, asymptotically, the regulator has full remote control over the required capital of the institution, as the optimal solutions of (3) converge to \mathbb{A} when d goes to infinity.

Theorem 8.2. Assume $\mathbb{A} \ll \mathbb{P}$, and let $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \dots$ be a sequence of nested algebras of scenarios along with target probabilities given by $\mathbb{A}[S]$ for all $S \in \mathcal{C}_n$, and such that $\mathcal{F} = \bigvee_{n \geq 1} \sigma(\mathcal{C}_n)$. Then the measures $\mathbb{Q}_n^* \ll \mathbb{P}$ defined by $\frac{d\mathbb{Q}_n^*}{d\mathbb{P}} = \mathbb{E} \left[\frac{d\mathbb{A}}{d\mathbb{P}} \mid \sigma(\mathcal{C}_n) \right]$ are solutions of (3) subject to the respective views. These solution measures converge to \mathbb{A} in the sense that $\frac{d\mathbb{Q}_n^*}{d\mathbb{P}} \rightarrow \frac{d\mathbb{A}}{d\mathbb{P}}$ in $L^1(\mathbb{P})$. The respective capital requirements converge as

$$\text{VaR}_\alpha^{\mathbb{Q}_n^*}(X) \rightarrow \text{VaR}_\alpha^{\mathbb{A}}(X), \quad \text{for all } X \in L^0 \text{ with } q_\alpha^{\mathbb{A}^-}(X) = q_\alpha^{\mathbb{A}^+}(X),$$

and

$$\text{ES}_\alpha^{\mathbb{Q}_n^*}(X) \rightarrow \text{ES}_\alpha^{\mathbb{A}}(X), \quad \text{for all } X \in L^\infty(\mathbb{P}).$$

9 Point-mass scenario aggregation

In this section we develop a scenario aggregation method that could serve as an alternative to the minimum divergence method in case where the internal model \mathbb{P} is incompatible with the views (1) as described in Corollary 5.4.⁴ The method developed here is subject to more ad hoc specifications than the minimum divergence scenario aggregation method, as illustrated by Example 9.2 below, which makes it less attractive for regulatory purposes.

As in Section 4 we are given d possibly non-disjoint scenarios $S_1, \dots, S_d \in \mathcal{F}$ along with target probabilities $c_1, \dots, c_d \geq 0$, inducing the views (1) to be satisfied by a $\mathbb{Q} \in \mathcal{M}$. We aggregate these views on the scenarios S_i with the internal model \mathbb{P} by solving a similar convex optimization problem as (3) but with the φ -divergence $d(\mathbb{Q}, \mathbb{P})$ being replaced by the total variation distance,

$$\begin{aligned} & \text{minimize} && d_{TV}(\mathbb{Q}, \mathbb{P}) \\ & \text{subject to} && \text{views (1)}. \end{aligned} \tag{13}$$

Problem (13) with domain $\mathbb{Q} \in \mathcal{M}$ is inherently infinite-dimensional, which makes it hard if not impossible to solve it. We thus deliberately restrict the domain to a finite-dimensional simplex. Thereto we consider the σ -algebra $\sigma(S_1, \dots, S_d)$ generated by the scenarios S_i , and we let V_1, \dots, V_N be the sample atoms of $\sigma(S_1, \dots, S_d)$. That is, V_1, \dots, V_N is the unique family of mutually disjoint sets with $\sigma(V_1, \dots, V_N) = \sigma(S_1, \dots, S_d)$ and $\bigcup_{k=1}^N V_k = \Omega$. Consequently, for every i there exists an index set $K(i) \subset \{1, \dots, N\}$ such that $S_i = \bigcup_{k \in K(i)} V_k$. Note that there are at least as many sample atoms as \mathbb{P} -atoms, $n \leq N$, with equality if and only if $\mathbb{P}[V_k] > 0$ for all k . The number of sample atoms is bounded as

$$N \leq 2^d. \tag{14}$$

Indeed, any sample atom V_k corresponds to a multi-index $\alpha \in \{0, 1\}^d$ with $\alpha_i = 1$ if $k \in K(i)$ and $\alpha_i = 0$ otherwise. This correspondence is injective. Hence N is bounded by the cardinality of $\{0, 1\}^d$, which is 2^d as claimed. Equality in (14) holds if and only if $S_i \cap S_j \neq \emptyset$ for all i, j and $\bigcup_{i=1}^d S_i \neq \Omega$.

We now fix N sample points $\omega = (\omega_1, \dots, \omega_N)$, one in each sample atom, $\omega_k \in V_k$, and such that $\mathbb{P}[\omega_k] = 0$ for all k .⁵ We then define the N -dimensional convex set

$$\mathcal{Q}_\omega = \left\{ \mathbb{Q}_r = (1 - \mathbf{1}^\top r) \mathbb{P} + \sum_{k=1}^N r_k \delta_{\omega_k} \mid r \in [0, \infty)^n, \mathbf{1}^\top r \leq 1 \right\} \subset \mathcal{M}$$

⁴It has also been discussed in [57] as an alternative to the current SST method.

⁵We assume here that \mathcal{F} contains all singletons $\{\omega\}$.

where δ_ω denotes the Dirac measure at ω . The measures in \mathcal{Q}_ω and \mathbb{P} are singular, but their total variation distance can be easily computed.

Lemma 9.1. *For any $\mathbb{Q}_r \in \mathcal{Q}_\omega$ we have*

$$d_{TV}(\mathbb{Q}_r, \mathbb{P}) = \sum_{k=1}^N r_k. \quad (15)$$

In compact notation the views (1) for a measure $\mathbb{Q}_r \in \mathcal{Q}_\omega$ read

$$\sum_{k \in K(i)} (1 - \mathbf{1}^\top \mathbf{r}) \mathbb{P}[V_k] + r_k \geq c_i, \quad i = 1, \dots, d,$$

or in matrix form $M\mathbf{r} \geq \mathbf{b}$ for the $d \times N$ -matrix M defined as $M_{ij} = 1 - \sum_{k \in K(i)} \mathbb{P}[V_k]$ if $j \in K(i)$ and $M_{ij} = -\sum_{k \in K(i)} \mathbb{P}[V_k]$ otherwise, and the d -vector \mathbf{b} defined as $b_i = c_i - \sum_{k \in K(i)} \mathbb{P}[V_k]$. The optimization problem (13) with domain \mathcal{Q}_ω boils down to the N -dimensional linear problem

$$\begin{aligned} & \text{minimize} && \mathbf{1}^\top \mathbf{r} \\ & \text{subject to} && M\mathbf{r} \geq \mathbf{b}, \\ & && \mathbf{1}^\top \mathbf{r} \leq 1, \\ & && \mathbf{r} \in [0, \infty)^N. \end{aligned} \quad (16)$$

Solutions of a linear problem such as (16), when they exist, are not unique and not in closed form in general, and numerical methods such as the simplex or the interior point methods are used, see, e.g., [102, Chapters 13 and 14]. Note, while any solution \mathbf{r}^* of (16) does not depend on the specific choice of ω_k in V_k , the corresponding optimal measure $\mathbb{Q}_{\mathbf{r}^*}$ solving problem (13) with domain \mathcal{Q}_ω does. We illustrate this with the following stress test example.

Example 9.2. We consider a stress test as described at the beginning of Section 4. We assume that $d = 1$ tail loss scenario $S = \{L \geq \ell\}$ is given with target probability $c = 1 - \alpha$ for some $\alpha \in (0, 1)$ and loss level $\ell \in \mathbb{R}$. We assume that $0 < \mathbb{P}[S] < 1$. The $N = 2$ sample atoms are then $V_1 = S$ and $V_2 = S^c$. We fix two sample points $\boldsymbol{\omega} = (\omega_1, \omega_2)$ with $\omega_k \in V_k$ and $\mathbb{P}[\omega_k] = 0$ for $k = 1, 2$. The linear problem (16) becomes

$$\begin{aligned} & \text{minimize} && r_1 + r_2 \\ & \text{subject to} && (1 - \mathbb{P}[S])r_1 - \mathbb{P}[S]r_2 \geq c - \mathbb{P}[S], \\ & && r_1 + r_2 \leq 1, \\ & && r_1, r_2 \geq 0. \end{aligned} \quad (17)$$

We verify by inspection that the unique solution \mathbf{r}^* to (17) is given by $r_1^* = (c - \mathbb{P}[S])^+ / (1 - \mathbb{P}[S])$ and $r_2^* = 0$. We obtain $\mathbb{Q}_{\mathbf{r}^*}[S] = \mathbb{P}[S] + (1 - \mathbb{P}[S])r_1^* = \mathbb{P}[S] + (c - \mathbb{P}[S])^+$. If the internal model is rejected, $\mathbb{P}[S] < c$, then $\mathbb{Q}_{\mathbf{r}^*}[S] = c$ and ℓ becomes an α -quantile of L under $\mathbb{Q}_{\mathbf{r}^*}$. The expected shortfall of L is thus

$$\text{ES}_\alpha^{\mathbb{Q}_{\mathbf{r}^*}}(L) = \frac{1}{1 - \alpha} \left((1 - r_1^*) \mathbb{E}[(L - \ell)^+] + r_1^* (L(\omega_1) - \ell) \right) + \ell.$$

As $\omega_1 \in S$ we know that $L(\omega_1) - \ell \geq 0$, and this difference can be arbitrarily large, up to $\|L\|_{L^\infty} - \ell$, depending on the choice of ω_1 in S . This result nicely illustrates that, while \mathbf{r}^* does not depend on the specific choice of ω_1 in S , the corresponding optimal measure $\mathbb{Q}_{\mathbf{r}^*}$ solving problem (13) with domain \mathcal{Q}_ω does. Consequently, so does the expected shortfall under $\mathbb{Q}_{\mathbf{r}^*}$. This feature of the point-mass scenario aggregation is potentially not desirable for regulatory purposes.

We also note that the value-at-risk does not depend on the choice of ω_1 in S . It satisfies $\text{VaR}_\alpha^{\mathbb{Q}_{\mathbf{r}^*}}(L) \leq \ell$, with equality if ℓ is the left-quantile at level α .

10 Conclusion

Scenario aggregation is an important part of risk-based solvency regulation. It serves as a device to address model uncertainty. The current SST method put in place by FINMA is subject to a critical review. We provide a novel coherent scenario aggregation method based on minimum statistical divergence subject to expert views on a given set of scenarios. This method has been designed to satisfy five criteria that are particularly relevant from a regulatory point of view:

- (1) *No penalty for conservative internal models.* According to Corollary 5.3, the minimum divergence aggregation leaves the internal model unchanged whenever it already satisfies the views.
- (2) *Focus on tail loss events.* As illustrated by Examples 7.3 and 7.4, the minimum divergence aggregation allows to accurately focus on insurer specific tail events.
- (3) *Control over distance from internal model.* The minimum divergence aggregation method is designed to minimize the φ -divergence of alternative models from the internal model.
- (4) *Robustness of capital requirements.* Capital requirements based on expected shortfall (value-at-risk) are shown to be robust with respect to minimum divergence aggregation (under some technical conditions).
- (5) *Tractability.* The minimum divergence aggregation can be casted as a finite-dimensional convex optimization problem. Efficient numerical algorithms are available in e.g. [102].

The alternative model resulting from the minimum divergence scenario aggregation is an interpolation between the internal model and the regulator's views within minimal distance from the internal model. The alternative model does not depend on the choice of the divergence function if the scenarios are disjoint, but in general it does. The regulator has remote control over the capital requirements as he can tune the trade off between idiosyncrasy (internal model) and standardization (regulator's views) via increasing the number of scenarios. An asymptotic result is available in case where the scenarios form an algebra. We also provide a point-mass scenario aggregation method that can serve as alternative in case where the minimum divergence method does not apply. One could further elaborate on combining these two methods, and on quantifying the impact of the choice of the divergence function.

A Figures

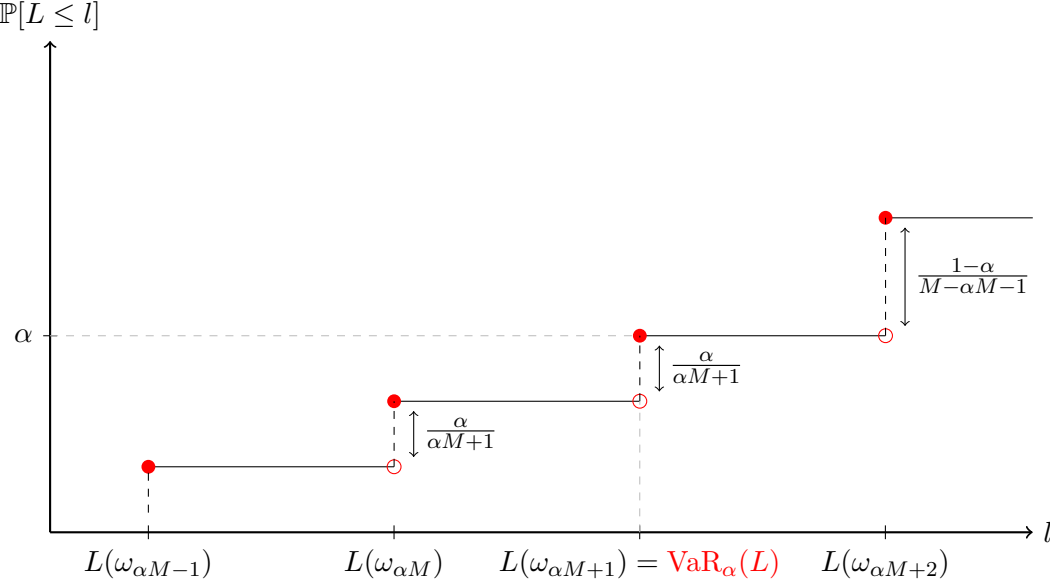
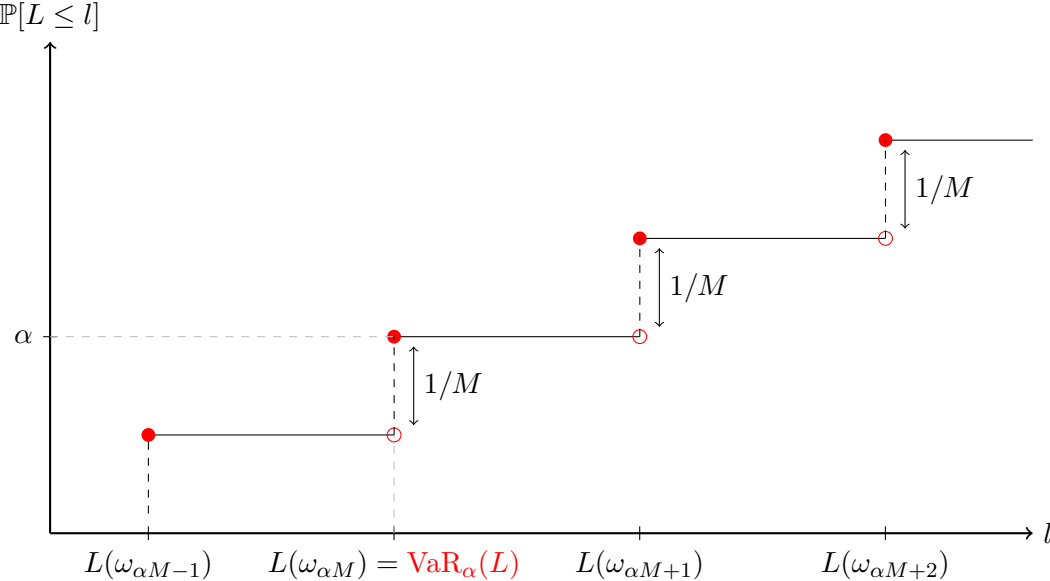


Figure 1: Cumulative distribution function of L in a simulation based model, before scenario aggregation (top) and after scenario aggregation (bottom), from Example 6.3.

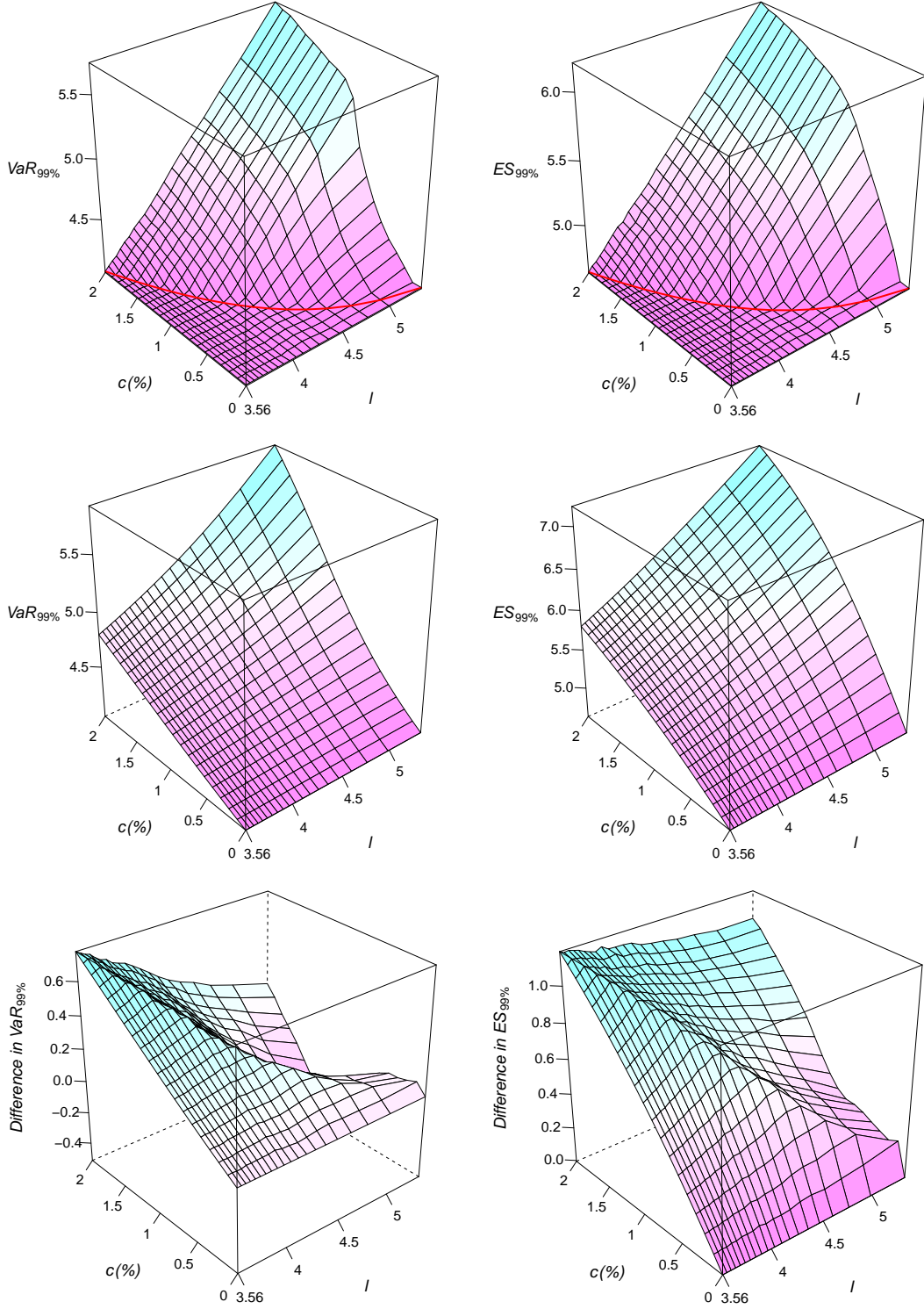


Figure 2: Impact on $\text{VaR}_{99\%}^{\mathbb{Q}^*}$ (left) and $\text{ES}_{99\%}^{\mathbb{Q}^*}$ (right), of the minimum divergence scenario aggregation method (top), the SST method (middle), and the difference between the SST and the minimum divergence aggregation method (bottom) for $c \in [0, 2\%]$ and $l \in \{\text{VaR}_\alpha(L) \mid \alpha \in [98\%, 99.9\%]\}$. The difference between the value-at-risk numbers becomes negative for some (c, l) . The red line on the top plots indicates the curve $(c, l = \text{VaR}_{1-c}(L))$. See Example 7.3.

		ℓ			
		3.56	4.03	4.47	5.35
$c(\%)$	0	4.03	4.03	4.03	4.03
		4.03	4.03	4.03	4.03
		(0)	(0)	(0)	(0)
	0.5	4.03	4.03	4.03	4.37
		4.21	4.25	4.28	4.35
		(0.18)	(0.22)	(0.25)	(-0.02)
	1	4.03	4.03	4.47	5.35
		4.40	4.49	4.58	4.80
		(0.37)	(0.46)	(0.11)	(-0.55)
	2	4.03	4.47	4.85	5.73
		4.80	5.02	5.26	5.91
		(0.77)	(0.55)	(0.41)	(0.18)

		ℓ			
		3.56	4.03	4.47	5.35
$c(\%)$	0	4.62	4.62	4.62	4.62
		4.62	4.62	4.62	4.62
		(0)	(0)	(0)	(0)
	0.5	4.62	4.62	4.62	5.30
		4.98	5.09	5.22	5.52
		(0.36)	(0.47)	(0.60)	(0.22)
	1	4.62	4.62	5.01	5.88
		5.30	5.50	5.72	6.27
		(0.68)	(0.88)	(1.10)	(0.39)
	2	4.62	5.01	5.37	6.20
		5.80	6.13	6.46	7.23
		(1.18)	(1.12)	(1.09)	(1.03)

Table 1: Impact on $\text{VaR}_{99\%}^{\mathbb{Q}^*}$ (top table) and $\text{ES}_{99\%}^{\mathbb{Q}^*}$ (bottom table), of the minimum divergence scenario aggregation method (top numbers), the SST method (middle numbers), and the difference between the SST and the minimum divergence aggregation method (bottom numbers in brackets) for $c \in \{0.1\%, 0.5\%, 1\%, 2\%\}$ and $\ell \in \{\text{VaR}_\alpha(L) \mid \alpha = 98\%, 99\%, 99.5\%, 99.9\%\} = \{3.56, 4.03, 4.47\}$. The difference between the value-at-risk numbers is negative for $c = 0.5\%, 1\%$ and $\ell = 5.35$. See Example 7.3.

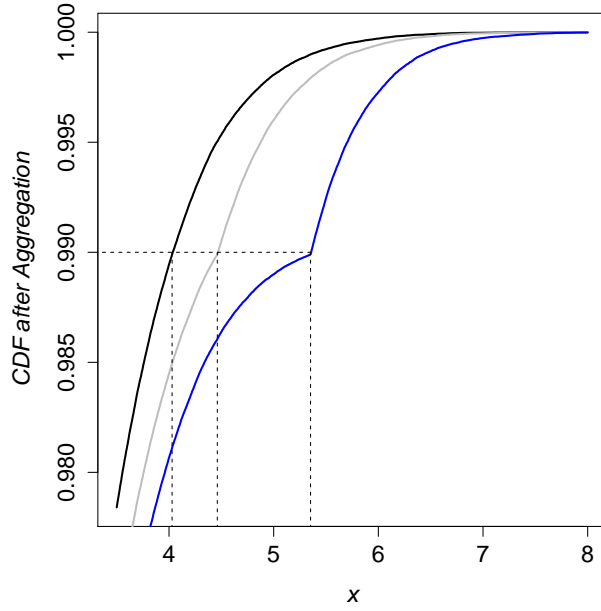


Figure 3: Distribution function of L under \mathbb{Q}^* , $x \mapsto \mathbb{Q}^*[L \leq x]$, resulting from minimum divergence scenario aggregation for $c = 1\%$ for different largest loss levels $\ell = \text{VaR}_{99\%}(L)$ (black), $\text{VaR}_{99.5\%}(L)$ (grey) and $\text{VaR}_{99.9\%}(L)$ (blue). The vertical dashed lines draw the different levels of ℓ , which correspond to the 99% quantile of L under \mathbb{Q}^* . See Example 7.3.

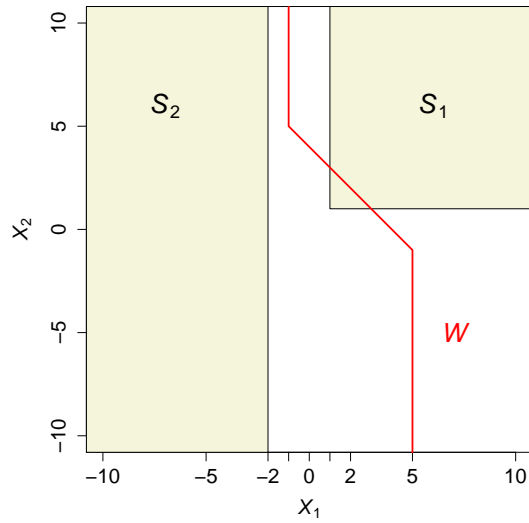


Figure 4: Representation of the two scenario sets $S_1 = \{X_1 \geq 1, X_2 \geq 1\}$ and $S_2 = \{X_1 < -2\}$, and of the tail loss event $W = \{L \geq \text{VaR}_{99\%}(L)\}$, where $L = \max(X_1, -1) + \max(\min(X_2, 5), -1)$, on the state space of the risk factors vector $(X_1, X_2)^\top$. See Example 7.4.

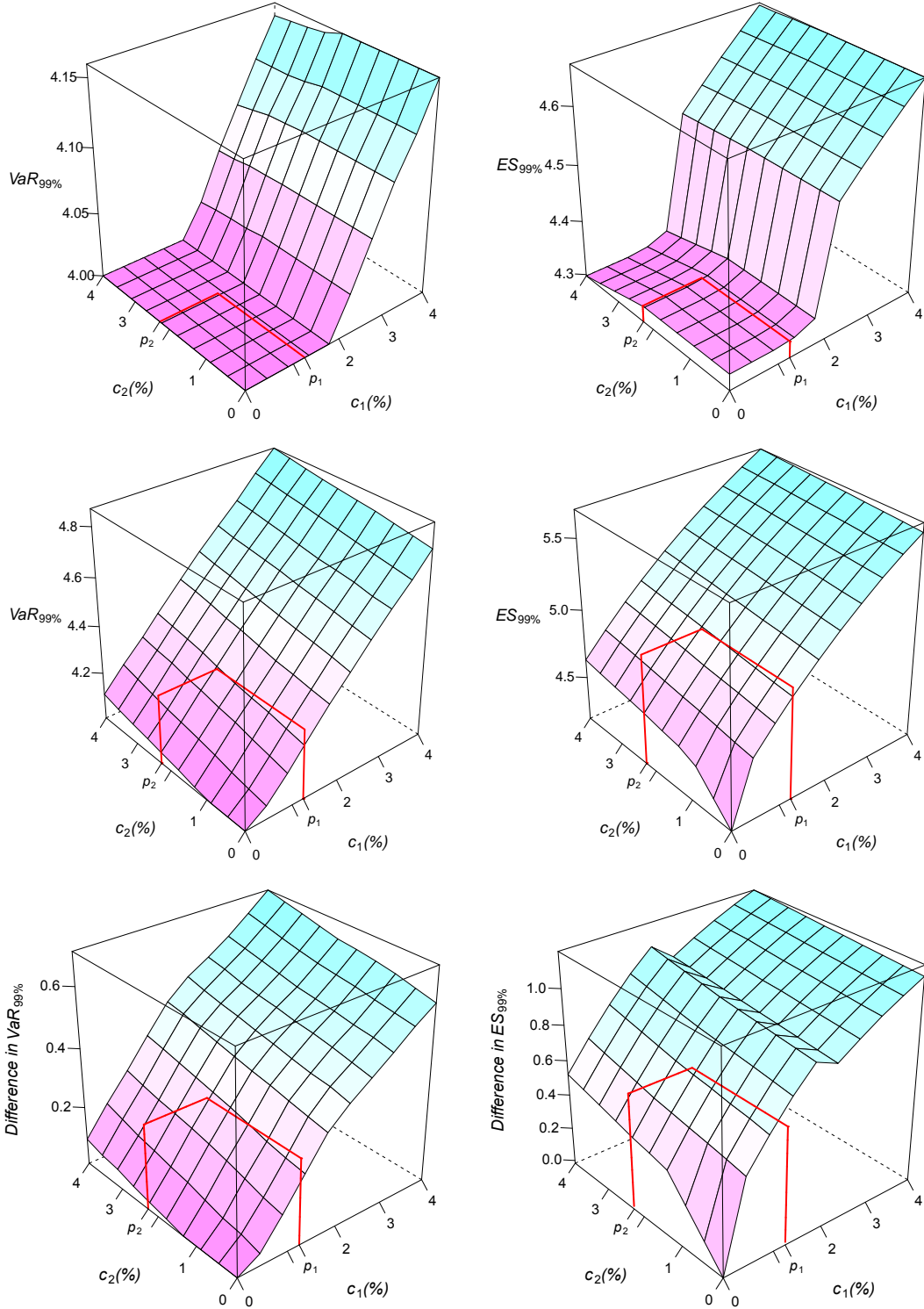


Figure 5: Impact on $VaR_{99\%}^*$ (left) and $ES_{99\%}^*$ (right), of the minimum divergence scenario aggregation method (top), of the SST method (middle), and the difference between the SST and the minimum divergence aggregation method (bottom) for $c_1, c_2 \in [0, 4\%]$. The red lines in the above pictures point to the required capital that would hold for target probabilities $c_1 = p_1, c_2 = p_2$. See Example 7.4.

		c_1					c_1		
		0	p_1	4%			0	p_1	4%
0		4.00	4.00	4.16	0		4.33	4.33	4.67
		4.00	4.22	4.75			4.33	4.90	5.61
		(0)	(0.22)	(0.59)			(0)	(0.57)	(0.94)
c_2	p_2	4.00	4.00	4.16	c_2	p_2	4.33	4.33	4.67
		4.03	4.29	4.81			4.54	4.98	5.64
		(0.03)	(0.29)	(0.65)			(0.21)	(0.65)	(0.97)
4%		4.00	4.00	4.15	4%		4.30	4.30	4.66
		4.09	4.34	4.86			4.62	5.03	5.66
		(0.09)	(0.34)	(0.71)			(0.32)	(0.73)	(1.00)

Table 2: Impact on $\text{VaR}_{99\%}^{\mathbb{Q}^*}$ (left table) and $\text{ES}_{99\%}^{\mathbb{Q}^*}$ (right table), of the minimum divergence scenario aggregation method (top numbers), of the SST method (middle numbers), and the difference between the SST and the minimum divergence aggregation method (bottom numbers) for $c_1 = 0, p_1$ and 4%, $c_2 = 0, p_2$ and 4%, with $p_1 = 1.2\%$ and $p_2 = 2.3\%$. See Example 7.4.

B Proofs

This appendix contains the proofs of all theorems, lemmas, and corollaries from the main text.

Proof of Lemma 3.1⁶

As L and X are independent, we have that $L + \mathbb{E}[X] = \mathbb{E}[L + X \mid L]$. Because expected shortfall is translation invariant and monotonic with respect to convex stochastic order, [51, Corollary 4.65], we obtain the lower bound

$$\text{ES}_\alpha(L) + \mathbb{E}[X] = \text{ES}_\alpha(L + \mathbb{E}[X]) \leq \text{ES}_\alpha(L + X).$$

The upper bound follows from the subadditivity property of the expected shortfall, see, e.g, [51].

Proof of Lemma 4.2

Let $\text{WCE}_\alpha(L) = \sup \{ \mathbb{E}[L \mid B] \mid B \in \mathcal{F}, \mathbb{P}[B] \geq 1 - \alpha \}$ be the worst conditional expectation at level $\alpha \in (0, 1)$ of L . Note that this definition is different from [51, Example 4.38] where the supremum is taken over events $B \in \mathcal{F}$ with $\mathbb{P}[B] > 1 - \alpha$. In any case, note that [51, Corollary 4.54] applies to our definition of WCE_α , that is

$$\text{ES}_\alpha(L) \geq \text{WCE}_\alpha(L) \geq \mathbb{E}[L \mid S],$$

for any event $S \in \mathcal{F}$, which proves the Lemma.

⁶Thanks to Ruodu Wang for suggesting this version of the proof.

Proof of Lemma 4.4

- (i) See, e.g., [31].
- (ii) See, e.g., [31] for the first statement. The second statement follows from the fact that all φ -divergences in (2) but $\varphi(t) = |t - 1|$ are strictly convex on $(0, \infty)$.
- (iii) To prove the lower bound, use the fact that $(\sqrt{t} - 1)^2 \leq |\sqrt{t} - 1||\sqrt{t} + 1|$. To prove the upper bounds, see [74, Theorem 6.1] for the first and [34, Theorem 1.3] for the second.
- (iv) Follows from the fact that

$$d_{L^1}(\mathbb{Q}, \mathbb{P}) = 2 \cdot \mathbb{E} \left[\left(\frac{d\mathbb{Q}}{d\mathbb{P}} - 1 \right) \mathbf{1}_{\left\{ \frac{d\mathbb{Q}}{d\mathbb{P}} \geq 1 \right\}} \right] = 2 \cdot \sup_{B \in \mathcal{F}} \left| \mathbb{E} \left[\left(\frac{d\mathbb{Q}}{d\mathbb{P}} - 1 \right) \mathbf{1}_B \right] \right| = 2 \cdot d_{TV}(\mathbb{Q}, \mathbb{P}).$$

Proof of Theorem 5.1

The proof of Theorem 5.1 relies on the following lemma, which we state here for its own interest.

Lemma B.1. *For every $\mathbb{Q} \in \mathcal{M}$ satisfying the views (1) and $\mathbb{Q} \ll \mathbb{P}$, there exists some $\mathbb{Q}' \in \mathcal{Q}$ satisfying the views (1) and such that $d(\mathbb{Q}', \mathbb{P}) \leq d(\mathbb{Q}, \mathbb{P})$.*

Proof. Let $\mathbb{Q} \in \mathcal{M}$ satisfy the views (1) and $\mathbb{Q} \ll \mathbb{P}$. Define $\mathbb{Q}' \in \mathcal{M}$ by $\frac{d\mathbb{Q}'}{d\mathbb{P}} = \mathbb{E} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} \mid \sigma(S_1, \dots, S_d) \right]$. Hence, $\mathbb{Q}' \in \mathcal{Q}$ and we obtain

$$c_i \leq \mathbb{Q}[S_i] = \mathbb{E} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} \mathbf{1}_{S_i} \right] = \mathbb{E} \left[\frac{d\mathbb{Q}'}{d\mathbb{P}} \mathbf{1}_{S_i} \right] = \mathbb{Q}'[S_i], \quad i = 1, \dots, d,$$

where we have used the definition of conditional expectation in the second equality. In addition, by Jensen's inequality, we have

$$d(\mathbb{Q}, \mathbb{P}) = \mathbb{E} \left[\varphi \left(\frac{d\mathbb{Q}}{d\mathbb{P}} \right) \right] \geq \mathbb{E} \left[\varphi \left(\mathbb{E} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} \mid \sigma(S_1, \dots, S_d) \right] \right) \right] = d(\mathbb{Q}', \mathbb{P}),$$

which concludes the proof. \square

We now proceed to the proof of Theorem 5.1. The equivalence of (i) and (ii) follows from Lemma B.1.

(ii) \Rightarrow (iii): let $\mathbb{Q}^* \in \mathcal{Q}$ be a solution of (3). Then \mathbb{Q}^* satisfies the views (1) and $\mathbb{Q}^* \ll \mathbb{P}$.

(iii) \Rightarrow (iv): let $\mathbb{Q} \ll \mathbb{P}$ satisfy the views (1). Lemma B.1 implies the existence of some $\mathbb{Q}' \in \mathcal{Q}$ satisfying the views (1). Then $\mathbf{q} \in [0, \infty)^n$ given by $q_j = \mathbb{Q}'[U_j]$ satisfies $\mathbf{1}^\top \mathbf{q} = 1$ and $A\mathbf{q} \geq \mathbf{c}$.

(iv) \Rightarrow (ii): let us define the feasibility set

$$\mathcal{C} = \left\{ \mathbf{q} \in [0, \infty)^n \mid A\mathbf{q} \geq \mathbf{c} \text{ and } \mathbf{1}^\top \mathbf{q} = 1 \right\}, \quad (18)$$

and the objective function $\psi(\mathbf{q}) = \sum_{j=1}^n p_j \varphi \left(\frac{q_j}{p_j} \right)$ associated to the convex problem (5). As the feasibility set \mathcal{C} is not empty there exists a solution of (5). Indeed, this follows from continuity of $\psi(\mathbf{q})$ in the non-empty compact feasibility set \mathcal{C} . Using Lemma B.1 and the finite dimensional representation (4), we find that the existence of a solution of (3) in \mathcal{Q} is equivalent to the existence of a solution of (5). This implies (ii).

The uniqueness statement follows from Lemma 4.4(i) stating that the objective function $d(\mathbb{Q}, \mathbb{P})$ of (3) is strictly convex in $\mathbb{Q} \in \mathcal{M}$ if $\varphi(t)$ is strictly convex on $(0, \infty)$.

Proof of Corollary 5.3

If \mathbb{P} satisfies the views (1), then it is in the feasibility set of the convex optimization problem (3). The statement of Lemma 4.4(i) then proves that $\mathbb{Q}^* = \mathbb{P}$ is the unique solution of (3).

Proof of Corollary 5.4

If S_i is a scenario such that $\mathbb{P}[S_i] = 0$, then it doesn't contain any \mathbb{P} -atom of $\sigma(S_1, \dots, S_d)$. In consequence, $J(i)$ is empty. As the i -th inequality constraint of the feasibility set \mathcal{C} given in (18) is $\sum_{j \in J(i)} q_j \geq c_i > 0$ and $\sum_{j \in J(i)} q_j = 0$ when $J(i) = \emptyset$, it cannot be satisfied. In consequence, \mathcal{C} is empty and thus the convex problem (3) has no solution.

Proof of Lemma 6.1

Denote $F_n(x) = \mathbb{P}_n[X \leq x]$, $n \in \mathbb{N}$ and $F(x) = \mathbb{P}[X \leq x]$. In addition, there should be no confusion between $F(q-)$, the left limit of F at q and $F(q^-)$, the evaluation of F at the left quantile q^- .

Property (6) follows from Lemma 4.4(iv). For the second statement, assume that for some $\epsilon > 0$ there exists a subsequence $(q_{n_k})_{k \in \mathbb{N}}$ such that $q_{n_k} \leq q_\alpha^-(X) - \epsilon$ for all $k \in \mathbb{N}$. Therefore,

$$F_{n_k}(q_{n_k}) - F(q_{n_k}) \geq \alpha - F(q_\alpha^-(X) - \epsilon) > 0 \quad \text{for all } k \in \mathbb{N},$$

which is in contradiction to (6). Similarly, assume that from some $\epsilon > 0$, there exists a subsequence $(q_{n_k})_{k \in \mathbb{N}}$ such that $q_{n_k} \geq q_\alpha^+(X) + \epsilon$ for all $k \in \mathbb{N}$. Therefore,

$$F(q_{n_k}-) - F_{n_k}(q_{n_k}-) \geq F((q_\alpha^+(X) + \epsilon)-) - \alpha > 0 \quad \text{for all } k \in \mathbb{N},$$

which is again in contradiction to (6), hence allowing us to conclude the proof.

Proof of Theorem 6.4

Let us assume that for some $\epsilon > 0$, there exists a subsequence $(n_k)_k$ such that

$$\left| \text{ES}_\alpha^{\mathbb{P}^{n_k}}(X) - \text{ES}_\alpha(X) \right| \geq \epsilon \quad \text{for all } k \in \mathbb{N}. \quad (19)$$

From Lemma 6.1, we know that there exists a subsequence of $(n_k)_k$ (still denoted $(n_k)_k$ for simplicity) such that $\lim_{k \rightarrow \infty} q_{n_k} = q$, for some accumulation point $q \in [q_\alpha^-(X), q_\alpha^+(X)]$. Defining $Z_{n_k} = d\mathbb{P}_{n_k}/d\mathbb{P}$, we thus obtain

$$\begin{aligned} & (1 - \alpha) \left| \text{ES}_\alpha^{\mathbb{P}^{n_k}}(X) - \text{ES}_\alpha(X) \right| \\ &= \left| \mathbb{E} [Z_{n_k}(X - q_{n_k})^+] - \mathbb{E} [(X - q)^+] + (1 - \alpha)(q_{n_k} - q) \right| \\ &\leq \left| \mathbb{E} [(Z_{n_k} - 1)(X - q_{n_k})^+] + \mathbb{E} [(X - q_{n_k})^+ - (X - q)^+] \right| + (1 - \alpha) |q_{n_k} - q| \\ &\leq \|Z_{n_k} - 1\|_p \|X\|_r + |q_{n_k} - q| + (1 - \alpha) |q_{n_k} - q| \longrightarrow 0, \end{aligned}$$

where we used the fact that $|(X - q_{n_k})^+ - (X - q)^+| \leq |q_{n_k} - q|$. This result is in contradiction with (19), hence proving the theorem.

Proof of Corollary 6.5

Follows from Lemma 4.4(iii) and Theorem 6.4.

Proof of Theorem 7.2

The assumption of mutually disjoint scenarios S_i with $\mathbb{P}[S_i] > 0$ and $\cup_{i=1}^d S_i = \Omega$ \mathbb{P} -a.s implies that $A = I_d$, the $d \times d$ identity matrix. In consequence the feasibility set \mathcal{C} of (5) given in (18) is non-empty only if $\sum_{i=1}^d c_i \leq 1$, which is thus necessary for the existence of a solution.

To prove the sufficiency assume first that $\sum_{i=1}^d c_i = 1$. Then $\mathcal{C} = \{\mathbf{c}\}$, and thus $\mathbf{q}^* = \mathbf{c}$ is the unique solution of (5). Assume now that $\sum_{i=1}^d c_i < 1$. Define $q_i(\epsilon) = \max\{c_i, \epsilon\}$. Then, by monotonicity and continuity of $\epsilon \mapsto q_i(\epsilon)$, and thus of $\epsilon \mapsto \sum_{i=1}^d q_i(\epsilon)$, there exists some $\epsilon^* > 0$ such that $\sum_{i=1}^d q_i(\epsilon^*) = 1$. In consequence, $\mathbf{q} = (q_i(\epsilon^*)) \in \mathcal{C}$ and is such that $\mathbf{q} > 0$, which satisfies the hypothesis of Lemma 7.1, Slater's condition, that can now be applied. The KKT conditions read here

$$\boldsymbol{\lambda} \geq 0, \quad \mathbf{q} \geq \mathbf{c}, \quad \boldsymbol{\lambda}^\top (\mathbf{q} - \mathbf{c}) = 0 \quad (20)$$

$$\mathbf{1}^\top \mathbf{q} = 1 \quad (21)$$

$$\varphi' \left(\frac{\mathbf{q}}{\mathbf{p}} \right) - \boldsymbol{\lambda} + \nu \mathbf{1} = \mathbf{0}. \quad (22)$$

Note that the inequality condition $\boldsymbol{\lambda} \geq 0$ in (20) and the equality condition (22) can be reduced to one inequality condition, we say that $\boldsymbol{\lambda}$ is a slack variable. The KKT conditions then become

$$\mathbf{1}^\top \mathbf{q} = 1 \quad (23)$$

$$\mathbf{q} \geq \mathbf{c} \quad (24)$$

$$\varphi' \left(\frac{q_i}{p_i} \right) + \nu \geq 0, \quad i = 1, \dots, d \quad (25)$$

$$\left(\varphi' \left(\frac{q_i}{p_i} \right) + \nu \right) (q_i - c_i) = 0, \quad i = 1, \dots, d. \quad (26)$$

It follows by inspection that $q_i^* = \max\{c_i, p_i \mu^*\}$ satisfies conditions (24), (25), (26) for $\nu^* = -\varphi'(\mu^*)$. It thus remains to find μ^* such that $F(\mu^*) = 1$ where we define the function $F(\mu) = \sum_{i=1}^d \max\{c_i, p_i \mu\}$. The function $F(\mu)$ is continuous, non-decreasing, piecewise linear, and can be represented as

$$F(\mu) = \begin{cases} \sum_{i=1}^d c_i, & \text{if } \mu < \frac{c_d}{p_d} \\ \sum_{i=1}^{k-1} c_i + \mu \sum_{i=k}^d p_i, & \text{if } \frac{c_k}{p_k} \leq \mu < \frac{c_{k-1}}{p_{k-1}} \text{ for some } k = 2, \dots, d \\ \mu, & \text{if } \frac{c_1}{p_1} \leq \mu. \end{cases}$$

As $F(c_d/p_d) < 1$ by assumption, the set $\mathcal{K} = \{1 \leq k \leq d \mid F(c_k/p_k) \leq 1\}$ is non-empty, and there exists a unique $\mu^* > c_d/p_d$ with $F(\mu^*) = 1$. It satisfies

$$\sum_{i=1}^{k^*-1} c_i + \mu^* \sum_{i=k^*}^d p_i = 1$$

with $k^* = \min \mathcal{K}$. The proof follows by observing that $\mathcal{K} = \left\{ 1 \leq k \leq d \mid \frac{c_k}{p_k} \leq \frac{1 - \sum_{i=1}^{k-1} c_i}{\sum_{i=k}^d p_i} \right\}$.

Proof of Corollary 7.3

We know from Theorem 5.1 that finding a solution of (3) is equivalent to finding a solution of (5) with two \mathbb{P} -atoms given by S and S^c . It is clear that the assumptions of the corollary fall under the umbrella of Theorem 7.2. A simple application of the theorem gives that a solution of (5) is given by $\mathbf{q}^* = (q^*, 1 - q^*)^\top$ such that $q^* = \max\{\mathbb{P}[S], c\}$ and $1 - q^* = \min\{1 - \mathbb{P}[S], 1 - c\}$.

Proof of Corollary 7.4

Given $S = \{L \geq \ell\}$ and $c = 1 - \alpha$, the solution \mathbb{Q}^* of the minimum divergence scenario aggregation in Corollary 7.3 is such that

$$\mathbb{Q}^*[L \geq \ell] = \max\{\mathbb{P}[L \geq \ell], 1 - \alpha\}.$$

Whenever $\mathbb{P}[L \geq \ell] < 1 - \alpha$, we have that ℓ is an α -quantile of \mathbb{Q}^* , i.e. $\ell \in (q_\alpha^-, q_\alpha^+]$. The expected shortfall under the alternative internal model \mathbb{Q}^* at level α becomes

$$\begin{aligned} \text{ES}_\alpha^{\mathbb{Q}^*}(L) &= \frac{1}{1 - \alpha} \mathbb{E}_{\mathbb{Q}^*} [L \mathbf{1}_{\{L \geq \ell\}}] + \frac{\ell}{1 - \alpha} (1 - \alpha - \mathbb{Q}^*[L \geq \ell]) = \frac{1}{1 - \alpha} \frac{1 - \alpha}{\mathbb{P}[L \geq \ell]} \mathbb{E} [L \mathbf{1}_{\{L \geq \ell\}}] \\ &= \mathbb{E}[L \mid L \geq \ell] = \frac{1}{\mathbb{P}[L \geq \ell]} (\mathbb{E}[L \mathbf{1}_{\{L \geq \ell\}}] - \ell \mathbb{P}[L \geq \ell] + \ell \mathbb{P}[L \geq \ell]) = \text{ES}_{\mathbb{P}[L < \ell]}(L). \end{aligned}$$

Whenever $\mathbb{P}[L \geq \ell] \geq 1 - \alpha$, $\mathbb{Q}^* = \mathbb{P}$ and $\text{ES}_\alpha^{\mathbb{Q}^*}(L) = \text{ES}_\alpha(L)$. This finishes the proof.

Proof of Lemma 8.1

- (i) For any $S \in \mathcal{C}$ we have $\mathbb{Q}[S] \geq \mathbb{A}[S]$ and $\mathbb{Q}[S^c] \geq \mathbb{A}[S^c]$. This can only hold if $\mathbb{Q} = \mathbb{A}$ on \mathcal{C} .
- (ii) Follows from (i) and the fact that two probability measures coinciding on the algebra \mathcal{C} also coincide on the σ -algebra $\sigma(\mathcal{C})$ generated by \mathcal{C} , see [124, Lemma 1.6].
- (iii) Theorem 5.1 (iii) and (ii) imply there exists a solution $\mathbb{Q}^* \in \mathcal{Q}$ of (3). Property (ii) of this lemma implies that $\mathbb{E} \left[\frac{d\mathbb{Q}^*}{d\mathbb{P}} \mathbf{1}_S \right] = \mathbb{Q}^*[S] = \mathbb{A}[S] = \mathbb{E} \left[\frac{d\mathbb{A}}{d\mathbb{P}} \mathbf{1}_S \right]$ for all $S \in \mathcal{C}$. As $\frac{d\mathbb{Q}^*}{d\mathbb{P}}$ is $\sigma(\mathcal{C})$ -measurable this implies that $\frac{d\mathbb{Q}^*}{d\mathbb{P}} = \mathbb{E} \left[\frac{d\mathbb{A}}{d\mathbb{P}} \mid \sigma(\mathcal{C}) \right]$, and \mathbb{Q}^* is the desired solution.

Proof of Theorem 8.2

Assume first that views are given on \mathcal{C}_n with target probabilities defined by $\mathbb{A}[S]$ for all $S \in \mathcal{C}_n$. By a direct application of Lemma 8.1 (iii), we have that the alternative measure $\mathbb{Q}_n^* \ll \mathbb{P}$ defined by $\frac{d\mathbb{Q}_n^*}{d\mathbb{P}} = \mathbb{E} \left[\frac{d\mathbb{A}}{d\mathbb{P}} \mid \sigma(\mathcal{C}_n) \right]$ is a solution of (3). [124, Theorem 14.2] implies that $\frac{d\mathbb{Q}_n^*}{d\mathbb{P}} \rightarrow \frac{d\mathbb{A}}{d\mathbb{P}}$ in $L^1(\mathbb{P})$. Using a similar argument as in proof of Lemma 4.4(iv), it follows that

$$\sup_{x \in \mathbb{R}} |\mathbb{Q}_n^*[X \leq x] - \mathbb{A}[X \leq x]| \rightarrow 0, \tag{27}$$

for $X \in L^0$. The last two statements of the theorem then follow from a simple modification of the proofs of Lemma 6.1 and Theorem 6.4.

Proof of Lemma 9.1

By definition of the total variation, see Lemma 4.4(iv), we obtain

$$\begin{aligned} d_{TV}(\mathbb{Q}_{\mathbf{r}}, \mathbb{P}) &= \sup_{B \in \mathcal{F}} \left| (1 - \mathbf{1}^\top \mathbf{r}) \mathbb{P}[B] + \sum_{k=1}^N r_k \delta_{\omega_k}(B) - \mathbb{P}[B] \right| = \sup_{B \in \mathcal{F}} \left| \sum_{k=1}^N r_k (\delta_{\omega_k}(B) - \mathbb{P}[B]) \right| \\ &\leq \sum_{k=1}^N r_k \sup_{B \in \mathcal{F}} |\delta_{\omega_k}(B) - \mathbb{P}[B]| = \sum_{k=1}^N r_k. \end{aligned}$$

Noting that the upper bound is attained whenever $B = \cup_{k=1}^N \{\omega_k\}$ concludes the proof.

Paper B

**An Importance Sampling Approach
for Copula Models in Insurance**

An importance sampling approach for copula models in insurance*

Philipp Arbenz[†], Mathieu Cambou[‡] Marius Hofert[§]

April 7, 2015

Abstract

An importance sampling approach for sampling copula models is introduced. We propose two algorithms that improve Monte Carlo estimators when the functional of interest depends mainly on the behaviour of the underlying random vector when at least one of the components is large. Such problems often arise from dependence models in finance and insurance. The importance sampling framework we propose is general and can be easily implemented for all classes of copula models from which sampling is feasible. We show how the proposal distribution of the two algorithms can be optimized to reduce the sampling error. In a case study inspired by a typical multivariate insurance application, we obtain variance reduction factors between 10 and 30 in comparison to standard Monte Carlo estimators.

Key words: Copula, Dependence models, Importance sampling, Insurance, Risk measure, Tail event

1 Introduction

Many insurance applications, see our motivation Section 2, lead to the problem of calculating a functional of the form $\mathbb{E}[\Psi_0(\mathbf{X})]$, where $\mathbf{X} = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ is a random vector on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\Psi_0 : \mathbb{R}^d \rightarrow \mathbb{R}$ is a measurable function. If the components of \mathbf{X} cannot be assumed to be independent, it is popular to model the distribution of \mathbf{X} with a copula, such that

$$\mathbb{P}[X_1 \leq x_1, \dots, X_d \leq x_d] = C(F_{X_1}(x_1), \dots, F_{X_d}(x_d)), \quad \mathbf{x} \in \mathbb{R}^d,$$

where $F_{X_j}(x) = \mathbb{P}[X_j \leq x]$, $j = 1, \dots, d$, are the marginal cumulative distribution functions (cdf) and $C : [0, 1]^d \rightarrow [0, 1]$ is a copula. A copula allows one to separate the dependence structure from

*The authors thank Hansjörg Albrecher, Anthony Davison, Paul Embrechts, Damir Filipovic, Christiane Lemieux and an anonymous referee for valuable feedback. As SCOR Fellow, Mathieu Cambou thanks SCOR for financial support.

[†]SCOR Global P&C, General Guisan Quai 26, 8022 Zürich, Switzerland

Email: philipp.arbenz@gmail.com

[‡]Institute of Mathematics, Station 8, EPFL, 1015 Lausanne, Switzerland

Email: mathieucambou@gmail.com

[§]Department of Statistics and Actuarial Science, University of Waterloo, Canada

Email: marius.hofert@uwaterloo.ca

the marginal distributions, which is useful for constructing multivariate stochastic models. We assume the reader to have a basic knowledge on copulas and refer to [89] or [99] for introductions.

The usual approach to estimate $\mathbb{E}[\Psi_0(\mathbf{X})]$ is by Monte Carlo simulation. In actuarial practice, very often a set of outcomes of \mathbf{X} with a low probability makes a large contribution to $\mathbb{E}[\Psi_0(\mathbf{X})]$. In this case, importance sampling can increase the number of samples lying in this set. Through a weighting approach, an unbiased estimator with a reduced variance can be obtained.

Importance sampling for copulas has been investigated by [56] and [69] for the Gauss copula only and [10] for absolutely continuous copulas. These papers are inspired by copula models in financial applications and assume the copula to be either Gaussian or having a known density. Copulas used in insurance however often deviate from these assumptions.

The main contribution of this paper is the study of importance sampling techniques that do not rely on a specific copula structure. We consider the case where the functional Ψ_0 of interest depends mainly on the behaviour of the random vector \mathbf{X} when at least one of the components is large. Such problems often arise from dependence models in the realm of finance and insurance, where distorted expectations of heavy tailed distributions are involved. We propose a new importance sampling framework for this setup which can be implemented for all classes of copula models from which sampling is feasible.

This paper is organized as follows. After motivating our work in Section 2, we introduce the importance sampling approach in Section 3. Section 4 presents a rejection sampling algorithm while Section 5 presents a direct sampling algorithm. For each of them, we expose the sampling of the proposal distribution, the calculation of the importance sampling weights and we discuss the optimal choice of the proposal distribution. Section 6 discusses the efficiency of our algorithms in rare event settings. A case study is given in Section 7 and Section 8 concludes.

2 Motivation

In a copula model, we may write

$$\mathbb{E}[\Psi_0(\mathbf{X})] = \mathbb{E}[\Psi(\mathbf{U})]$$

where $\mathbf{U} = (U_1, \dots, U_d) : \Omega \rightarrow \mathbb{R}^d$ is a random vector with distribution function C , $\Psi : [0, 1]^d \rightarrow \mathbb{R}$ is defined as

$$\Psi(u_1, \dots, u_d) = \Psi_0 \left(F_{X_1}^{-1}(u_1), \dots, F_{X_d}^{-1}(u_d) \right),$$

and $F_{X_j}^{-1}(p) = \inf\{x \in \mathbb{R} : F_{X_j}(x) \geq p\}$, for $j = 1, \dots, d$.

If C and the margins F_{X_j} are known, we can use Monte Carlo simulation to estimate $\mathbb{E}[\Psi(\mathbf{U})]$. For a random sample $\{\mathbf{U}_i : i = 1, \dots, n\}$ of \mathbf{U} , the Monte Carlo estimator of $\mathbb{E}[\Psi(\mathbf{U})]$ is given by

$$\mu_n = \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{U}_i). \quad (2.1)$$

In this paper, we consider the case where Ψ is large only when at least one of its arguments is close to 1, or equivalently, if at least one of the components of \mathbf{X} is large. This assumption is inspired by several applications in insurance, as the following examples illustrate:

- The fair premium of a stop loss cover with deductible T is $\mathbb{E} \left[\max \left\{ \sum_{j=1}^d X_j - T, 0 \right\} \right]$. The corresponding functional is $\Psi(\mathbf{u}) = \max \left\{ \sum_{j=1}^d F_{X_j}^{-1}(u_j) - T, 0 \right\}$; see the left hand side of Figure 1 for a contour plot of Ψ for two Pareto margins.

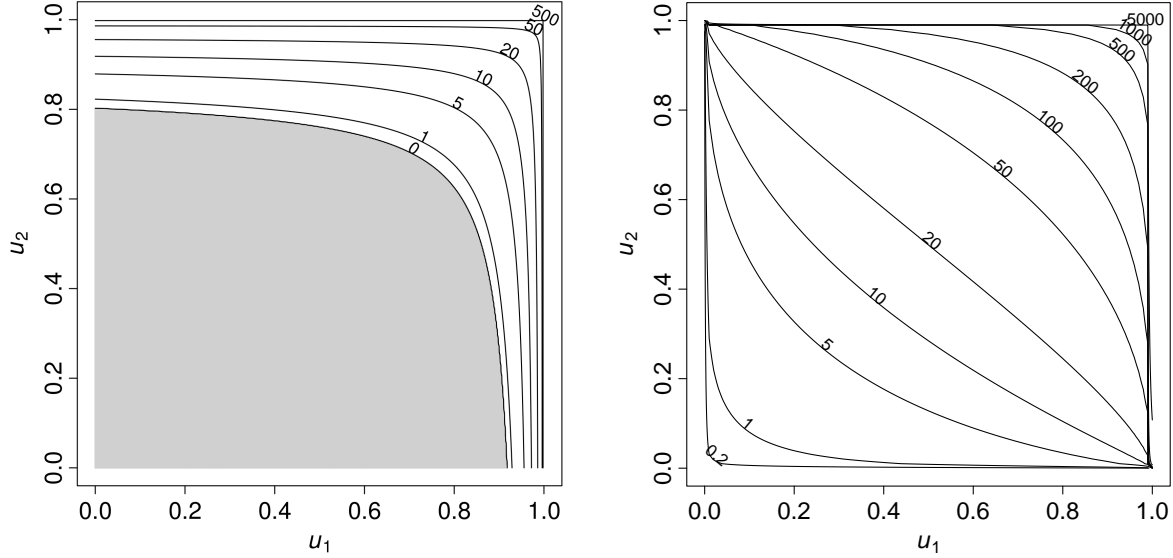


Figure 1: *Left:* Contour lines for the excess function $\Psi(u_1, u_2) = \max\{F_{X_1}^{-1}(u_1) + F_{X_2}^{-1}(u_2) - 10, 0\}$, where the margins are Pareto distributed with $F_{X_1}(x) = 1 - (1+x/4)^{-2}$ and $F_{X_2}(x) = 1 - (1+x/8)^{-2}$. The grey area indicates where Ψ is zero. *Right:* Contour lines for the product function $\Psi(u_1, u_2) = F_{X_1}^{-1}(u_1)F_{X_2}^{-1}(u_2)$, where $X_1 \sim \text{LN}(2, 1)$ and $X_2 \sim \text{LN}(1, 1.5)$.

- Risk measures for an aggregate $S = \sum_{j=1}^d X_j$, such as Value-at-Risk, $\text{VaR}_\alpha(S)$, or Expected Shortfall, $\text{ES}_\alpha(S)$, $\alpha \in (0, 1)$, cannot in general be written as an expectation of type $\mathbb{E}[\Psi_0(\mathbf{X})]$. However, they are functionals of the aggregate distribution function $F_S(x) = \mathbb{P}[S \leq x] = \mathbb{E}[\Psi^{(x)}(\mathbf{U})]$, where $\Psi^{(x)}(x \in \mathbb{R})$ is the indicator function

$$\Psi^{(x)}(\mathbf{u}) = \mathbf{1} \left\{ F_{X_1}^{-1}(u_1) + \dots + F_{X_d}^{-1}(u_d) \leq x \right\}.$$

We can therefore write

$$\text{VaR}_\alpha(S) = \inf \left\{ x \in \mathbb{R} : \mathbb{E}[\Psi^{(x)}(\mathbf{U})] \geq \alpha \right\}, \quad \text{ES}_\alpha(S) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_u(S) du,$$

which depend only on those x for which $\mathbb{E}[\Psi^{(x)}(\mathbf{U})] \geq \alpha$ holds. This is determined by the tail behaviour of S , which is strongly influenced by the properties of the copula C when at least one component is close to 1. Note that capital allocation methods such as the Euler principle for Expected Shortfall behave similarly, see [122] and [89], page 260.

- Computing the covariance (or correlation) of two positive heavy-tailed random variables X_1 and X_2 requires the calculation of $\mathbb{E}[X_1 X_2]$. The implied functional is $\Psi(u_1, u_2) =$

$F_{X_1}^{-1}(u_1)F_{X_2}^{-1}(u_2)$. A contour plot of Ψ for log-normal (LN) margins is shown in the right hand side of Figure 1. In contrast to the preceding examples, this Ψ does not only depend on the tail behaviour of (X_1, X_2) . However, $\mathbb{E}[\Psi(\mathbf{U})]$ depends mainly on the copula behaviour when at least one argument is close to 1, as Ψ becomes large in this case.

Note that in this framework we follow the convention of [89, Remark 2.1] that \mathbf{X} refers to a loss and $-\mathbf{X}$ to a profit, which is more common in an actuarial context. One could have equally well worked with the P&L random variable $-\mathbf{X}$ by changing the area of interest to where components of \mathbf{X} are small.

3 Importance sampling

The idea behind importance sampling is to sample from a proposal distribution $F_{\mathbf{V}}$ different from the target distribution C . The proposal distribution concentrates more samples in the region driving large contributions to $\mathbb{E}[\Psi(\mathbf{U})]$. With a suitable weighting approach, one obtains an unbiased estimator with lower variance.

Suppose the function Ψ under consideration is in the class illustrated above: Ψ is large if at least one of its arguments is close to 1. In this case, a drawback of the estimator μ_n in (2.1) is that, typically, for many of the samples \mathbf{U}_i , none of the components is close to 1. Therefore, most samples lie in a region of low interest. The estimation error of μ_n can thus be large, even if n is large.

Let $\mathbf{V} = (V_1, \dots, V_d) : \Omega \rightarrow [0, 1]^d$ denote a random vector with distribution function $F_{\mathbf{V}}$. We can rewrite the integral $\mathbb{E}[\Psi(\mathbf{U})]$ as

$$\mathbb{E}[\Psi(\mathbf{U})] = \int_{[0,1]^d} \Psi(\mathbf{u})dC(\mathbf{u}) = \int_{[0,1]^d} \Psi(\mathbf{u}) \frac{dC(\mathbf{u})}{dF_{\mathbf{V}}(\mathbf{u})} dF_{\mathbf{V}}(\mathbf{u}) = \mathbb{E} \left[\Psi(\mathbf{V}) \frac{dC(\mathbf{V})}{dF_{\mathbf{V}}(\mathbf{V})} \right], \quad (3.1)$$

where $dC/dF_{\mathbf{V}}$ denotes the Radon–Nikodym derivative of C with respect to $F_{\mathbf{V}}$. The Radon–Nikodym derivative exists if and only if the copula C is absolutely continuous with respect to $F_{\mathbf{V}}$. We will provide more details on this issue later in this section. If C and $F_{\mathbf{V}}$ are absolutely continuous with densities c and $f_{\mathbf{V}}$ with respect to the Lebesgue measure, the Radon–Nikodym derivative $dC/dF_{\mathbf{V}}$ is simply the ratio of the densities $c/f_{\mathbf{V}}$.

For an i.i.d. sample $\{\mathbf{V}_i : i = 1, \dots, n\}$ of \mathbf{V} , we can define the importance sampling estimator

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{V}_i)w(\mathbf{V}_i), \quad (3.2)$$

where $w(\mathbf{V}_i) = dC(\mathbf{V}_i)/dF_{\mathbf{V}}(\mathbf{V}_i)$ are the sample weights. The goal is then to find $F_{\mathbf{V}}$ such that the variance of $\hat{\mu}_n$ is smaller than the variance of μ_n .

In order to define the proposal distribution $F_{\mathbf{V}}$, we suggest a mixing approach by taking a weighted average of a multivariate cdf $C^{[\lambda]} : [0, 1]^d \rightarrow [0, 1]$ over different values of λ . Let F_{Λ} denote the distribution function of a random variable $\Lambda : \Omega \mapsto [0, 1]$. We then define the distribution $F_{\mathbf{V}}$ of \mathbf{V} as a mixture of $C^{[\lambda]}$ over the distribution F_{Λ} :

$$F_{\mathbf{V}}(\mathbf{u}) = \int_0^1 C^{[\lambda]}(\mathbf{u}) dF_{\Lambda}(\lambda), \quad \mathbf{u} \in [0, 1]^d.$$

The distribution $C^{[\lambda]}$ shall be understood as a distorted version of the copula C that will concentrate samples in specific regions of the sampling space. These regions will then be parametrized by the value of λ . More precisely, we will construct $C^{[\lambda]}$ so that it puts mass only in the region $[0, 1]^d \setminus [0, \lambda]^d$. In the sequel, we will propose two possible definitions of $C^{[\lambda]}$ that will define two importance sampling algorithms, namely a rejection sampling algorithm in Section 4 and a direct sampling algorithm in Section 5.

We will see that this mixture approach is natural in order to allow C to be absolutely continuous with respect to $F_{\mathbf{V}}$. In particular, the absolute continuity is guaranteed for any copula C if the following condition is satisfied.

Condition A. *The random variable Λ satisfies $\mathbb{P}[\Lambda = 0] > 0$.*

In order to obtain a well defined weight function $w(\mathbf{V})$ and an unbiased estimator $\hat{\mu}_n$, Condition A must be fulfilled. This condition does not require particular assumptions on C . Although it seems restrictive, we will see that it is also needed to have a consistent estimator $\hat{\mu}_n$. To that end, we assume Condition A to be satisfied in what follows.

The construction of the proposal distribution $F_{\mathbf{V}}$ as a $C^{[\lambda]}$ -mixture directly yields a sampling method, as one can draw a realization of $F_{\mathbf{V}}$ by first drawing $\Lambda \sim F_{\Lambda}$ and then $\mathbf{V} \sim C^{[\Lambda]}$. Therefore, the following algorithm can be used to calculate $\hat{\mu}_n$:

Algorithm 3.1. *Fix $n \in \mathbb{N}$. For $i = 1, \dots, n$, do:*

1. *draw $\Lambda_i \sim F_{\Lambda}$;*
2. *draw $\mathbf{V}_i \sim C^{[\Lambda_i]}$;*
3. *calculate $w(\mathbf{V}_i)$;*

Return $\hat{\mu}_n = n^{-1} \sum_{i=1}^n \Psi(\mathbf{V}_i)w(\mathbf{V}_i)$.

The following lemma establishes consistency and asymptotic normality of the estimator $\hat{\mu}_n$.

Lemma 3.2. *Suppose that $\text{var}[\Psi(\mathbf{U})] < \infty$ and that $w(\cdot) \leq B$ for some constant $B < \infty$. Then*

1. *$\hat{\mu}_n$ converges \mathbb{P} -almost surely to μ ;*
2. *$\sigma^2 = \text{var}[\Psi(\mathbf{V})w(\mathbf{V})] < \infty$ and $n^{1/2}(\hat{\mu}_n - \mu)$ converges to $\mathcal{N}(0, \sigma^2)$ in distribution.*

Proof.

1. Since $\mathbb{E}[\Psi(\mathbf{V})w(\mathbf{V})] = \mathbb{E}[\Psi(\mathbf{U})]$, consistency follows directly from the Strong Law of Large Numbers.
2. Note that

$$\mathbb{E}[\Psi(\mathbf{V})^2 w(\mathbf{V})^2] = \mathbb{E}[\Psi(\mathbf{U})^2 w(\mathbf{U})] \leq \mathbb{E}[\Psi(\mathbf{U})^2] B < \infty,$$

where the first equality is justified by a change of measure, see (3.1). We can immediately deduce asymptotic normality of $\hat{\mu}_n$ by the Central Limit Theorem, see, for example, Section 2.4 in [37], page 110. \square

We will later show that under some mild assumptions on F_{Λ} , the weight function will indeed be bounded on $[0, 1]$.

4 A rejection sampling algorithm

For this algorithm, we propose $C^{[\lambda]}$ to denote the distribution of \mathbf{U} conditioned on the event that at least one of its components exceeds λ :

$$\begin{aligned} C^{[\lambda]}(\mathbf{u}) &= \mathbb{P}[U_1 \leq u_1, \dots, U_d \leq u_d \mid \max\{U_1, \dots, U_d\} > \lambda] = \mathbb{P}[U_1 \leq u_1, \dots, U_d \leq u_d \mid \mathbf{U} \notin [0, \lambda]^d] \\ &= \frac{C(\mathbf{u}) - C(\min\{u_1, \lambda\}, \dots, \min\{u_d, \lambda\})}{1 - C(\lambda \mathbf{1})}, \end{aligned}$$

where $\lambda \mathbf{1} = \lambda(1, \dots, 1) = (\lambda, \dots, \lambda) \in [0, 1]^d$. Note that $C^{[\lambda]}$ is a copula only if $C(\lambda \mathbf{1}) = 0$, but $C^{[\lambda]}$ does not need to be copula for our algorithm to work. By putting mass of Λ on $(0, 1)$, we can put more weight on the region of the copula where at least one component is large. For instance, if F_Λ is discrete and $\mathbb{P}[\Lambda = 0] = \mathbb{P}[\Lambda = 0.9] = 0.5$, then 50% of the samples of \mathbf{V} are constrained to lie only in $[0, 1]^d \setminus [0, 0.9]^d$ while the other 50% of the samples will lie on $[0, 1]^d$. Note that the mass on $[0, 1]^d \setminus [0, 0.9]^d$ would then be higher than 50%. On the other hand, the case $\mathbb{P}[\Lambda = 0] = 1$ yields $F_{\mathbf{V}} = C$.

4.1 Sampling the proposal distribution

We shall now describe how samples from $F_{\mathbf{V}}$ can be drawn. As $F_{\mathbf{V}}$ is defined through a mixing distribution, drawing a realization from $F_{\mathbf{V}}$ is done by drawing first $\Lambda \sim F_\Lambda$ and then $\mathbf{V} \sim C^{[\Lambda]}$, see Algorithm 3.1. Unfortunately, for well-known copula classes, the conditional distribution $C^{[\lambda]}$ is not analytically tractable. We are aware of only one class of shock copulas, namely Marshall–Olkin copulas, for which it is possible to sample directly from the conditional distribution $C^{[\lambda]}$. Details and the corresponding algorithm are provided in Appendix A.

However, sampling from $C^{[\lambda]}$ for an arbitrary copula C is always possible through a rejection algorithm, which is simple to implement but may be time consuming due to the rejection step. With the following rejection algorithm, it is thus possible to draw a sample from $F_{\mathbf{V}}$ for any copula C . The only condition is that it is feasible to draw realizations from both F_Λ and C . It is not necessary to know further properties of C , such as its density. The basic idea is to first draw a realization Λ from F_Λ and then iteratively draw realizations from C until one obtains a maximum component larger than Λ .

Algorithm 4.1. *To draw one realization of $F_{\mathbf{V}}$:*

1. draw $\Lambda \sim F_\Lambda$;
2. repeatedly draw $\mathbf{V} \sim C$ until $\max\{V_1, \dots, V_d\} > \Lambda$;
3. return \mathbf{V} .

A disadvantage of Algorithm 4.1 is that typically many samples of C are discarded, because of the acceptance condition in Step 2. In practice, there are two important reasons why this approach can be justified over standard Monte-Carlo. First, the evaluation of Ψ can be numerically more expensive than sampling from the copula, if, for instance, marginal quantile functions are demanding to compute or if Ψ_0 has no closed form. Second, storing a large sample of \mathbf{U} in computer memory

can be numerically more expensive than generating it. This case may appear for example in estimating allocated capital, which requires storing the whole multivariate sample. In particular in high dimensional problems, memory constraints can be quite prohibitive. For illustration, consider the following example: for the calculation of risk capital and risk contributions in a setting with heavy tailed marginals, a sample of size 10'000'000 is often not large enough to yield sufficiently small estimation errors. However, in a 1'000-dimensional setting with double-precision floating point numbers, this sample would require about 80 gigabytes of memory, which is more than an average computer currently possesses in terms of RAM.

Algorithm 4.1 may require several realizations from \mathbf{U} in order to generate one realization of \mathbf{V} . The following lemma gives an expression for the expected number of \mathbf{U} 's for obtaining a realization of \mathbf{V} .

Lemma 4.2. *Let $N_{\mathbf{V}}$ denote the number of draws from C necessary to simulate one realization from $F_{\mathbf{V}}$. The expected number of draws is*

$$\mathbb{E}[N_{\mathbf{V}}] = \int_0^1 \frac{1}{1 - C(\lambda \mathbf{1})} dF_{\Lambda}(\lambda).$$

Proof. The probability that one draw from $\mathbf{U} \sim C$ satisfies $\max\{U_1, \dots, U_d\} > \lambda$ is $\mathbb{P}[\max\{U_1, \dots, U_d\} > \lambda] = 1 - C(\lambda \mathbf{1})$. Therefore, the number of draws necessary to simulate from $C^{[\lambda]}$ for a fixed λ is geometrically distributed with parameter $1 - C(\lambda \mathbf{1})$ and has expectation $1/[1 - C(\lambda \mathbf{1})]$. In order to simulate from \mathbf{V} , Λ is drawn from F_{Λ} . Therefore, $\mathbb{E}[N_{\mathbf{V}}]$ is given by averaging $1/[1 - C(\lambda \mathbf{1})]$ over F_{Λ} . \square

Using the Fréchet–Höfdding bounds (see Theorem 5.7 in [89]), we can give the following bounds for $\mathbb{E}[N_{\mathbf{V}}]$, which depend only on F_{Λ} and the dimension d , independent of the copula C .

Theorem 4.3. *We have*

$$\frac{1}{d} \mathbb{E} \left[\frac{1}{1 - \Lambda} \right] \leq \mathbb{E}[N_{\mathbf{V}}] \leq \mathbb{E} \left[\frac{1}{1 - \Lambda} \right].$$

Proof. Due to the upper Fréchet–Höfdding bound, we have $C(\lambda \mathbf{1}) \leq \min\{\lambda, \dots, \lambda\} = \lambda$. Hence,

$$\mathbb{E}[N_{\mathbf{V}}] = \int_0^1 \frac{1}{1 - C(\lambda \mathbf{1})} dF_{\Lambda}(\lambda) \leq \int_0^1 \frac{1}{1 - \lambda} dF_{\Lambda}(\lambda) = \mathbb{E} \left[\frac{1}{1 - \Lambda} \right].$$

Analogously, due to the lower Fréchet–Höfdding bound:

$$\mathbb{E}[N_{\mathbf{V}}] \geq \int_0^1 \frac{1}{1 - \max\{0, d\lambda - d + 1\}} dF_{\Lambda}(\lambda) = \int_0^1 \max \left\{ 1, \frac{1}{d(1 - \lambda)} \right\} dF_{\Lambda}(\lambda) \geq \frac{1}{d} \mathbb{E} \left[\frac{1}{1 - \Lambda} \right]. \quad \square$$

Due to Theorem 4.3, the number of draws from C necessary to draw one realization from \mathbf{V} has a finite expectation if and only if $\mathbb{E}[(1 - \Lambda)^{-1}] < \infty$. Intuitively, this implies that Λ should not have mass concentrated near 1 in order to be able to use Algorithm 4.1.

We shall see in the next section that specific choices for the copula C and for F_{Λ} will allow us to find analytical expressions for $\mathbb{E}[N_{\mathbf{V}}]$.

4.2 Calculation of sample weights

This section outlines how the weights $w(\mathbf{V}_i)$ used in Algorithm 3.1 can be calculated. We first deduce a useful representation.

Theorem 4.4. *The Radon–Nikodym derivative $w(\mathbf{u}) = dC(\mathbf{u})/dF_{\mathbf{V}}(\mathbf{u})$ can be written as*

$$w(\mathbf{u}) = \left(\int_0^{\max\{u_1, \dots, u_d\}} \frac{1}{1 - C(\lambda \mathbf{1})} dF_{\Lambda}(\lambda) \right)^{-1}.$$

Proof. Due to the Leibnitz integral rule, we have $dF_{\mathbf{V}}(\mathbf{u}) = \int_0^1 dC^{[\lambda]}(\mathbf{u}) dF_{\Lambda}(\lambda)$. From the definition of $C^{[\lambda]}$, we can deduce the differential

$$dC^{[\lambda]}(\mathbf{u}) = \begin{cases} 0, & \mathbf{u} \in [0, \lambda]^d, \\ \frac{dC(\mathbf{u})}{1 - C(\lambda \mathbf{1})}, & \text{otherwise.} \end{cases}$$

Using both identities, we obtain

$$dF_{\mathbf{V}}(\mathbf{u}) = dC(\mathbf{u}) \int_0^1 \frac{\mathbf{1}\{\lambda \leq \max\{u_1, \dots, u_d\}\}}{1 - C(\lambda \mathbf{1})} dF_{\Lambda}(\lambda),$$

leading to the desired result. \square

The efficiency of our approach comes from the fact that the term $dC(\mathbf{u})$ does not appear in $w(\mathbf{u})$. For instance, if C is absolutely continuous with respect to the Lebesgue measure, the density of C does not have to be evaluated to calculate $w(\mathbf{u})$. This is in an advantage in comparison to most other importance sampling algorithms, for which the existence of the density of C is required.

In order to simplify the notation, let $\tilde{w}(t) : [0, 1] \rightarrow [0, \infty)$ be defined as

$$\tilde{w}(t) = \left(\int_0^t \frac{1}{1 - C(\lambda \mathbf{1})} dF_{\Lambda}(\lambda) \right)^{-1},$$

such that $w(\mathbf{u}) = \tilde{w}(\max\{u_1, \dots, u_d\})$.

Lemma 4.5. *Under Condition A, \tilde{w} is bounded from above by $\mathbb{P}[\Lambda = 0]^{-1}$ on $[0, 1]$.*

Proof. Since $C(\lambda \mathbf{1})$, $\lambda \in [0, 1]$, the diagonal section of the copula C and the distribution function F_{Λ} are both increasing functions, the weight function $\tilde{w}(t)$ is decreasing on $[0, 1]$, it is therefore bounded above by $\tilde{w}(0) = \mathbb{P}[\Lambda = 0]^{-1} < \infty$. \square

As a consequence, Condition A is not only sufficient to obtain existence of the weights, but it also guarantees that they are bounded. In virtue of Lemma 3.2, this is needed for consistency and asymptotic normality of the importance sampling estimator.

For general C and F_{Λ} , the evaluation of the weight function \tilde{w} can be demanding. In general, numerical integration schemes could be used. To circumvent these problems, we present two cases in which the evaluation of \tilde{w} is straightforward. Section 4.2.1 illustrates the case in which F_{Λ} is discrete. In Section 4.2.2, we assume that the copula C lies in a large class of copulas satisfying a polynomial condition on the diagonal. For this class, there is a specific choice of F_{Λ} which leads to an analytical expression for \tilde{w} .

4.2.1 Discrete F_Λ

This section shows that in the case of a discrete F_Λ , calculating $\tilde{w}(t)$ is fast and can easily be implemented. To this end, suppose F_Λ is discrete with a finite number n_Λ of atoms:

$$\mathbb{P}[\Lambda = x_k] = p_k, k = 1, \dots, n_\Lambda, \sum_{k=1}^{n_\Lambda} p_k = 1, p_1 > 0, \text{ and } 0 = x_1 < \dots < x_{n_\Lambda} < 1.$$

Note that Condition A is satisfied. In this case, \tilde{w} can be written as a step function

$$\tilde{w}(t) = \left(\sum_{k=1}^{n_\Lambda} \frac{\mathbf{1}\{x_k \leq t\}}{1 - C(x_k \mathbf{1})} p_k \right)^{-1}. \quad (4.1)$$

In order to evaluate $\tilde{w}(t)$, it is sufficient to calculate (or approximate) $C(x_k \mathbf{1})$ for $k = 1, \dots, n_\Lambda$. These values must be calculated only once for the whole sample. This approach with a discrete F_Λ can be used for any copula C . For $\mathbb{E}[N_{\mathbf{V}}]$, we obtain the explicit expression

$$\mathbb{E}[N_{\mathbf{V}}] = \sum_{k=1}^{n_\Lambda} \frac{p_k}{[1 - C(x_k \mathbf{1})]}.$$

4.2.2 Continuous F_Λ

For continuous F_Λ , the weight function \tilde{w} can in general only be calculated numerically. In the following, we assume that both C and F_Λ are of a special polynomial form, which leads to an explicit \tilde{w} . Suppose that C behaves as a monomial on its diagonal:

$$C(u \mathbf{1}) = u^\alpha, \quad 0 \leq u \leq 1.$$

Due to the Fréchet–Höfding bounds, α must satisfy $1 \leq \alpha \leq d$. This class of copulas is quite large. The following list shows some popular copula families satisfying this condition.

- Marshall–Olkin copulas as proposed in Example A.2 of Appendix A. The corresponding exponent is $\alpha = \sum_{j=1}^m \min_{i:j \in I_i} (s_j / \tilde{s}_i)$.
- Sibuya copulas, as defined in [66], for which the default rate process is a homogeneous Poisson process.
- Extreme value copulas with a Pickands dependence function A . The corresponding exponent is $\alpha = dA(1/d, \dots, 1/d)$; see Section 7 in [89] for a definition of extreme value copulas. Note that this class contains the well-known Gumbel copula, for example.

Apart from the copula C , we also make some specific assumptions about $F_\Lambda : [0, 1] \rightarrow [0, 1]$. Suppose that

$$F_\Lambda(\lambda) = (1 - \gamma) + \gamma \left(1 - (1 - \lambda^\alpha)^\beta \right), \quad \beta > 1, 0 \leq \gamma \leq 1.$$

The parameter α is given by the exponent of the copula diagonal, so cannot be chosen freely. Furthermore, F_Λ has an atom of weight $1 - \gamma$ at zero. This distribution is similar to the distribution of [78]. In this case, the weight function can easily be calculated as

$$\tilde{w}(t) = \left(1 - \gamma + \gamma\beta \int_0^t \alpha\lambda^{\alpha-1}(1 - \lambda^\alpha)^{\beta-2}d\lambda\right)^{-1} = \frac{\beta - 1}{\beta - 1 + \gamma(1 - \beta(1 - t^\alpha)^{\beta-1})}. \quad (4.2)$$

As $\mathbb{E}[N_{\mathbf{V}}] = 1/\tilde{w}(1)$ (c.f. Lemma 4.2), we obtain an explicit expression for $\mathbb{E}[N_{\mathbf{V}}]$:

$$\mathbb{E}[N_{\mathbf{V}}] = 1 + \frac{\gamma}{\beta - 1}. \quad (4.3)$$

In order for Condition A to be satisfied, we assume $\gamma < 1$. In fact, using properties of the hypergeometric function, it is possible to show that for $\gamma = 1$, the weight function is unbounded and the variance of the weights $\text{var}[w(\mathbf{V})]$ is always infinite.

There are many copula classes which have an explicit diagonal. For instance, the Clayton family has a diagonal $C(t\mathbf{1}) = (dt^{-\theta} - d + 1)^{-1/\theta}$ for some $0 < \theta < \infty$. For future research, we may point out that it would be interesting to find “conjugate” F_Λ for copulas that also allow for an explicit form of $\tilde{w}(\cdot)$.

4.3 Optimal proposal distribution

This section gives an approach to calibrate the distribution F_Λ to the problem at hand. The basic idea is to choose the proposal distribution $F_{\mathbf{V}}$ in such a way that $\hat{\mu}_n$ has a smaller variance than μ_n . In our case, this reduces to optimally choosing the distribution F_Λ . In general, F_Λ must have an atom at 0 in order to satisfy Condition A. If Algorithm 4.1 is used for sampling, we also need to fulfill the constraint that $\mathbb{E}[1/(1 - \Lambda)]$ is not too large, and, in particular, finite.

Zero variance (i.e., no estimation error) would be obtained for $\hat{\mu}_n$ if

$$\Psi(\mathbf{u})w(\mathbf{u}) = \mathbb{E}[\Psi(\mathbf{U})], \quad \mathbf{u} \in [0, 1)^d, \quad (4.4)$$

see Section 4.1 in [6], page 128. This choice is obviously not possible as $\mathbb{E}[\Psi(\mathbf{U})]$ is unknown. To obtain a small variance, we should choose Λ such that $w(\mathbf{u})^{-1}$ is approximately proportional to $\Psi(\mathbf{u})$. Due to Theorem 4.4, we may write this relation as

$$K \int_0^{\max\{u_1, \dots, u_d\}} \frac{1}{1 - C(\lambda\mathbf{1})} dF_\Lambda(\lambda) \approx \Psi(\mathbf{u}), \quad (4.5)$$

for some unknown constant $K \in \mathbb{R}_+$. In order to obtain a tractable optimization scheme, we use our assumption that $\Psi(\mathbf{u})$ is large if at least one of its components is large, namely

$$\Psi(\mathbf{u}) \approx \Psi(\max\{u_1, \dots, u_d\}\mathbf{1}). \quad (4.6)$$

Plugging (4.6) into (4.5), we obtain

$$K \int_0^t \frac{1}{1 - C(\lambda\mathbf{1})} dF_\Lambda(\lambda) \approx \Psi(t\mathbf{1}), \quad t \in [0, 1]. \quad (4.7)$$

In the following, we propose methods to calibrate F_Λ such that the approximate relation (4.7) is satisfied. We illustrate this calibration with the two choices for F_Λ being discrete and continuous.

4.3.1 Discrete F_Λ

In the discrete case, as defined in Section 4.2.1, specifying the distribution F_Λ reduces to setting the atoms x_k and their weights $p_k = \mathbb{P}[\Lambda = x_k]$ for $k = 1, \dots, n_\Lambda$. By plugging F_Λ into (4.7), we obtain

$$K \sum_{k=1}^{n_\Lambda} \frac{\mathbf{1}\{x_k \leq t\} p_k}{1 - C(x_k \mathbf{1})} \approx \Psi(t \mathbf{1}), \quad t \in [0, 1]. \quad (4.8)$$

We propose to set the p_k 's by enforcing equality to hold in (4.8) only for $t = x_1, \dots, x_{n_\Lambda}$. By assuming without loss of generality, that $x_k < x_{k+1}$ for all k , Equation (4.8) leads to

$$K \sum_{l=1}^k \frac{1}{1 - C(x_l \mathbf{1})} p_l = \Psi(x_k \mathbf{1}), \quad k = 1, \dots, n_\Lambda.$$

This yields a triangular linear system of equations which can be easily solved with the following algorithm; we propose to choose the x_k 's on a finite logarithmic grid becoming denser towards 1.

Algorithm 4.6.

1. Choose $n_\Lambda \in \mathbb{N}$;
2. define $x_k = 1 - (1/2)^{k-1}$, $k = 1, \dots, n_\Lambda$;
3. define $\tilde{p}_1 = \Psi(0, \dots, 0)$ and $\tilde{p}_k = (\Psi(x_k \mathbf{1}) - \Psi(x_{k-1} \mathbf{1})) (1 - C(x_k \mathbf{1}))$, for $k = 2, \dots, n_\Lambda$;
4. define $p_k = \tilde{p}_k / (\sum_l \tilde{p}_l)$.

The use of powers of $1/2$ to set the x_k 's is arbitrary; any other factor in $(0, 1)$ can be used instead. In numerical experiments, the impact of this choice was in general small, as the calculated p_k change accordingly.

In the following situations, Algorithm 4.6 may fail:

- if $p_1 = 0$, then F_Λ does not satisfy Condition A;
- if $t \mapsto \Psi(t \mathbf{1})$ is not monotone, then Algorithm 4.6 results in some of the p_k 's being negative;
- if the function Ψ does not attain a finite value at $(0, \dots, 0)$.

Since $n_\Lambda < \infty$, the condition $\mathbb{E}[1/(1 - \Lambda)] < \infty$ is automatically satisfied. Of course, one could also use discrete distributions for Λ supported by infinitely many points. However, in experiments analogous to the case study presented in Section 7, this has led to waiting times $\mathbb{E}[N_{\mathbf{V}}]$ becoming large without providing additional accuracy when using rejection sampling.

4.3.2 Continuous F_Λ

In the continuous case, as defined in Section 4.2.2, the optimization unfortunately cannot be done as easily and explicitly as for the discrete case. By putting F_Λ , see Equation (4.2), into (4.7), we obtain

$$K \left(1 + \frac{\gamma (1 - \beta(1 - t^\alpha)^{\beta-1})}{\beta - 1} \right) \approx \Psi(t\mathbf{1}), \quad t \in [0, 1]. \quad (4.9)$$

In order to optimize F_Λ , we would need to find parameters $K \in \mathbb{R}$, $\gamma \in (0, 1)$ and $\beta > 1$ which minimize some distance between the left and right hand side of (4.9). As distance function, one can for instance use the quadratic norm. This minimization can be implemented through standard numerical minimization procedures. Recall that α is fixed through the copula's diagonal. In order to have $\mathbb{E}[N_{\mathbf{V}}]$ not excessively high, one might want to impose a further parameter constraint by bounding $\mathbb{E}[N_{\mathbf{V}}] = 1 + \gamma/(\beta - 1)$.

5 A direct sampling algorithm

As noted in the previous section, the rejection sampling algorithm may lead to large sampling time due to the rejection step. This step was necessary due to the complexity of the conditioning event in the definition of $C^{[\lambda]}$. We now consider

$$\begin{aligned} C^{[\lambda]}(\mathbf{u}) &= d^{-1} \sum_{i=1}^d \mathbb{P}[U_1 \leq u_1, \dots, U_d \leq u_d \mid U_i > \lambda] \\ &= d^{-1} \sum_{i=1}^d \frac{C(\mathbf{u}) - C(u_1, \dots, u_{i-1}, \min\{u_i, \lambda\}, u_{i+1}, \dots, u_d)}{1 - \lambda}, \quad \mathbf{u} \in [0, 1]^d. \end{aligned} \quad (5.1)$$

This distribution only involves the conditional copula where the conditioning event is only on one element of the random vector \mathbf{U} . This will have the practical advantage that a direct sampling algorithm, i.e. with no rejection step, can be provided.

5.1 Sampling the proposal distribution

Let us denote C_{u_k} the conditional copula given that the k -th component equals u_k , that is

$$C_{u_k}(u_1, \dots, u_{k-1}, u_{k+1}, \dots, u_d) = \mathbb{P}[U_1 \leq u_1, \dots, U_{k-1} \leq u_{k-1}, U_{k+1} \leq u_{k+1}, \dots, U_d \leq u_d \mid U_k = u_k].$$

Sampling from $F_{\mathbf{V}}$ can then be performed using the following algorithm.

Algorithm 5.1. *To draw one realization of $F_{\mathbf{V}}$:*

1. draw $\Lambda \sim F_\Lambda$;
2. draw $I \in \{1, \dots, d\}$, with $\mathbb{P}[I = i] = d^{-1}$;
3. draw $V_I \sim U(\Lambda, 1)$;

4. draw $(V_1, \dots, V_{I-1}, V_{I+1}, \dots, V_d) \sim C_{V_I}$;

5. return $\mathbf{V} = (V_1, \dots, V_d)$.

The main advantage of this algorithm is that it does not reject any sample and as a consequence, in contrast to Algorithm 4.1, its run time does not depend on the distribution F_Λ . In addition, one can show that using a rejection algorithm for producing samples from (5.1) would yield an expected waiting time of $\mathbb{E}[(1 - \Lambda)^{-1}]$, which would be higher than the expected waiting time of the rejection sampling presented in Section 4, see Theorem 4.3. This justifies the fact that we propose two specific distributions of $C^{[\lambda]}$ for each of these two algorithms.

In Step 4 of Algorithm 5.1, a sampling algorithm for the conditional copula C_{u_k} , where k can be any of the d components, is required. Depending on the form of the copula C_{u_k} , efficient sampling algorithms may be available, see for instance Examples 5.3 and 5.4 below, or one can use the conditional distribution method. Note that the conditional distribution method is applied, for example, for sampling vine copulas; see the `VineCopula R` package.

Along the lines of [39], we then propose the following general algorithm to sample from C_{u_k} .

Algorithm 5.2. Given $u_k \in \mathbb{R}$, to draw one realization of C_{u_k} , do:

1. draw $\mathbf{U}' = (U'_1, \dots, U'_{k-1}, U'_{k+1}, \dots, U'_d) \sim U(0, 1)^{d-1}$;

2. set

$$\begin{aligned} U_1 &= C^{-1}(U'_1 | u_k) \\ &\vdots \\ U_{k-1} &= C^{-1}(U'_{k-1} | U_1, \dots, U_{k-2}, u_k) \\ U_{k+1} &= C^{-1}(U'_{k+1} | U_1, \dots, U_{k-2}, U_{k-1}, u_k) \\ &\vdots \\ U_d &= C^{-1}(U'_d | U_1, \dots, U_{k-1}, u_k, U_{k+1}, \dots, U_{d-1}) \end{aligned}$$

3. return $(U_1, \dots, U_{k-1}, U_{k+1}, \dots, U_d)$.

Following Theorem 2.27 and Remark 2.29 in [117], we have that for $k > j$

$$C(u_j | u_1, \dots, u_{j-1}, u_k) = \frac{D_{1, \dots, j-1, k} C_{1, \dots, j-1, j, k}(u_1, \dots, u_{j-1}, u_j, u_k)}{D_{1, \dots, j-1, k} C_{1, \dots, j-1, k}(u_1, \dots, u_{j-1}, u_k)}, \quad (5.2)$$

which simplifies to

$$C(u_j | u_1, \dots, u_{j-1}) = \frac{D_{1, \dots, j-1} C_{1, \dots, j-1, j}(u_1, \dots, u_{j-1}, u_j)}{D_{1, \dots, j-1} C_{1, \dots, j-1}(u_1, \dots, u_{j-1})},$$

whenever $k < j$. Here, $D_{1, \dots, j, k}$ denotes the partial derivatives with respect to the components $1, \dots, j, k$ and $C_{1, \dots, j, k}$ denotes the copula corresponding to the distribution of these components. In general, tractable inverses of the conditional distributions (5.2) are not always available, and numerical root-finding would need to be applied. However, there are cases where one can derive explicitly such inverses, see, e.g., Example 5.5. In consequence, although this algorithm does not involve a rejection step, it may require more implementation effort.

Example 5.3 (Direct sampling of Farlie–Gumbel–Morgenstern copula). *The Farlie–Gumbel–Morgenstern (FGM) copula is defined by*

$$C^\theta(\mathbf{u}) = \prod_{i=1}^d u_i \left(1 + \theta \prod_{j=1}^d (1 - u_j) \right), \quad \mathbf{u} \in \mathbb{R}^d,$$

with $\theta \in [-1, 1]$, see, e.g., [54]. This copula is a special form of the more general Eyrraud–Farlie–Gumbel–Morgenstern copula, see page 19 in [72]. It is easily seen that

$$\begin{aligned} \frac{\partial}{\partial u_k} C^\theta(\mathbf{u}) &= \prod_{i=1, i \neq k}^d u_i \left(1 + \theta(1 - 2u_k) \prod_{i=1, i \neq k}^d (1 - u_i) \right) \\ &= C^{\theta(1-2u_k)}(u_1, \dots, u_{k-1}, u_{k+1}, \dots, u_d), \end{aligned}$$

where $C^{\theta(1-2u_k)}$ is a FGM copula with parameter $\theta(1 - 2u_k) \in [-1, 1]$. As a consequence, sampling from $C_{u_k}^\theta$ is reduced to sampling from $C^{\theta(1-2u_k)}$. To this end, the conditional distribution method can be used. Producing a sample $\mathbf{U} \sim C^\theta$ can indeed be reduced to drawing $\mathbf{U}' \sim U(0, 1)^d$ and setting $U_1 = U'_1, \dots, U_{d-1} = U'_{d-1}$, and

$$U_d = \frac{2U'_d}{1 + W + \sqrt{(1 + W)^2 - 4WU'_d}},$$

where $W = \theta \prod_{j=1}^{d-1} (1 - 2U'_j)$, see Section 8.7.12 in [114] for more details.

Example 5.4 (Direct sampling of Frank copula). According to Section 6 in [92], if $C(\mathbf{u}) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d))$ is a d -dimensional Archimedean copula with generator ψ , then the $(d - 1)$ -dimensional copula \tilde{C} of the multivariate distribution C_{u_k} is also Archimedean, with generator

$$\psi_{u_k}(t) = \frac{\psi'(t + \psi^{-1}(u_k))}{\psi'(\psi^{-1}(u_k))}, \quad t \in [0, \infty].$$

This can be used to show that if C is a Frank copula with parameter $\alpha \in \mathbb{R}$ and generator $\psi_\alpha(t) = -\alpha^{-1} \log(1 - (1 - e^{-\alpha})e^{-t})$, then C_{u_k} can be modeled by a multivariate distribution with copula of type Ali–Mikhail–Haq with parameter $\theta(\alpha, u_k) = 1 - e^{-\alpha u_k}$, generator

$$\psi_{\theta(\alpha, u_k)}(t) = \frac{1 - \theta(\alpha, u_k)}{e^t - \theta(\alpha, u_k)} \quad (5.3)$$

and marginal distributions that have quantile functions

$$F_{\alpha, u_k}^{-1}(u) = -\frac{1}{\alpha} \log \left(\frac{e^{-\alpha} - 1}{1 + e^{-\alpha u_k}(u^{-1} - 1)} + 1 \right), \quad u \in [0, 1]. \quad (5.4)$$

In consequence, sampling from C_{u_k} is reduced to sampling from a Ali–Mikhail–Haq copula with generator (5.3), for example using the fast Marshall–Olkin algorithm, see Sections 2.4 and 2.5 in [63], and then applying the quantile function (5.4) to the copula sample. In a similar fashion, if C is Archimedean such that C_{u_k} is easy to sample with the Marshall–Olkin algorithm (many examples and techniques are known), and, additionally, the marginal distributions are easy to invert, then one obtains a fast sampling technique for Step 4 in Algorithm 5.1.

Example 5.5 (Conditional distribution method for Clayton copula). *The Clayton copula is defined by*

$$C^\theta(\mathbf{u}) = \left(1 + \sum_{i=1}^d (u_i^{-\theta} - 1)\right)^{-1/\theta}, \quad \mathbf{u} \in \mathbb{R}^d,$$

with $\theta > 0$. Using (5.2), one can show that

$$C^{\theta(-1)}(u'_j | u_1, \dots, u_{j-1}, u_k) = \left(1 + \left(1 - (j-1) + \sum_{k=1}^{j-1} u_k^{-\theta}\right) \left((u'_j)^{-\frac{1}{j-1+1/\theta}} - 1\right)\right)^{-1/\theta},$$

which allows one to easily implement Algorithm 5.2.

5.2 Calculation of sample weights

As for the rejection sampling approach, we derive a representation for the weights $w(\mathbf{V}_i)$ used in Algorithm 5.1.

Theorem 5.6. *The Radon–Nikodym derivative $w(\mathbf{u}) = dC(\mathbf{u})/dF_{\mathbf{V}}(\mathbf{u})$ can be written as*

$$w(\mathbf{u}) = \left(d^{-1} \sum_{i=1}^d \int_0^{u_i} \frac{1}{1-\lambda} dF_\Lambda(\lambda)\right)^{-1}.$$

Proof. Noting that

$$dC^{[\lambda]}(\mathbf{u}) = \frac{dC(\mathbf{u})}{d(1-\lambda)} \sum_{i=1}^d \mathbf{1}\{u_i > \lambda\},$$

we proceed similarly as in the proof of Theorem 4.4. □

As in the rejection sampling algorithm, we note that $dC(\mathbf{u})$ does not appear in $w(\mathbf{u})$, so that the existence of the density of C is not a requirement for the derivation of the weights. In order to insure consistency and asymptotic normality of the importance sampling estimator, we shall also check the boundedness of the weight function.

Lemma 5.7. *Under Condition A, the weight function w is bounded from above by $\mathbb{P}[\Lambda = 0]^{-1}$ on $[0, 1]$.*

Proof. We note that $w(\mathbf{u})$ is decreasing in all components. Hence, it is bounded above by $w(0, \dots, 0) = \mathbb{P}[\Lambda = 0]^{-1} < \infty$. □

For general F_Λ , the evaluation of the weight function w could be demanding. In general, numerical integration schemes could be used. To circumvent these problems we propose to use the same setup for F_Λ as in Section 4, i.e., a discrete case and a continuous case.

5.2.1 Discrete F_Λ

If F_Λ is discrete such that $\mathbb{P}[\Lambda = x_k] = p_k$, $p_1 > 0$, $k = 1, \dots, n_\Lambda$, $0 = x_1 < \dots < x_{n_\Lambda} < 1$, then w can be written as

$$w(\mathbf{u}) = d \left(\sum_{i=1}^d \sum_{k=1}^{n_\Lambda} \frac{\mathbf{1}\{x_k \leq u_i\}}{1 - x_k} p_k \right)^{-1}. \quad (5.5)$$

5.2.2 Continuous F_Λ

Taking the cdf of Λ as

$$F_\Lambda(\lambda) = (1 - \gamma) + \gamma \left(1 - (1 - \lambda)^\beta \right), \quad \beta > 1, 0 \leq \gamma < 1,$$

gives, for any copula C , the following closed form for the weights

$$w(\mathbf{u}) = \frac{\beta - 1}{\beta - 1 + \gamma - \gamma \beta d^{-1} \sum_{i=1}^d (1 - u_i)^{\beta-1}}. \quad (5.6)$$

Note that we do not need any restriction on the copula diagonal, in contrast to Section 4.2.2

5.3 Optimal proposal distribution

To obtain a small variance, we should choose Λ such that $w(\mathbf{u})^{-1}$ is approximately proportional to $\Psi(\mathbf{u})$. Due to Theorem 5.6, we may write this relation as

$$K d^{-1} \sum_{i=1}^d \int_0^{u_i} \frac{1}{1 - \lambda} dF_\Lambda(\lambda) \approx \Psi(\mathbf{u}), \quad (5.7)$$

for some unknown constant $K \in \mathbb{R}_+$. As per the rejection sampling approach, we shall restrict the calibration to the diagonal in order to obtain a tractable optimization scheme and we use our assumption that $\Psi(\mathbf{u}) \approx \Psi(\max\{u_1, \dots, u_d\} \mathbf{1})$. Therefore (5.7) reduces to

$$K \int_0^t \frac{1}{1 - \lambda} dF_\Lambda(\lambda) \approx \Psi(t \mathbf{1}), \quad t \in [0, 1]. \quad (5.8)$$

In the following, we propose methods to calibrate F_Λ such that the approximate relation (5.8) is satisfied. We illustrate this calibration with the two choices for F_Λ as outlined in Sections 5.2.1 and 5.2.2.

5.3.1 Discrete F_Λ

In the discrete case, we obtain

$$K \sum_{k=1}^{n_\Lambda} \frac{\mathbf{1}\{x_k \leq t\} p_k}{1 - x_k} \approx \Psi(t \mathbf{1}), \quad t \in [0, 1]. \quad (5.9)$$

We propose to determine the p_k 's by enforcing equality to hold in (5.9) only for $t = x_1, \dots, x_{n_\Lambda}$ which leads to the triangular system

$$K \sum_{l=1}^k \frac{1}{1-x_l} p_l = \Psi(x_k \mathbf{1}), \quad k = 1, \dots, n_\Lambda.$$

Choosing the x_k 's as in the rejection sampling approach, we can solve for the p_k 's using the following algorithm:

Algorithm 5.8.

1. Choose $n_\Lambda \in \mathbb{N}$;
2. Define $x_k = 1 - (1/2)^{k-1}$, $k = 1, \dots, n_\Lambda$;
3. Define $\tilde{p}_1 = \Psi(0, \dots, 0)$ and $\tilde{p}_k = (\Psi(x_k \mathbf{1}) - \Psi(x_{k-1} \mathbf{1})) (1 - x_k)$, for $k = 2, \dots, n_\Lambda$;
4. Define $p_k = \tilde{p}_k / (\sum_l \tilde{p}_l)$.

5.3.2 Continuous F_Λ

In the continuous case, the optimization unfortunately cannot be done as easily and explicitly as for the discrete case. In this case, the optimization on $K \in \mathbb{R}$, $\gamma \in (0, 1)$ and $\beta > 1$ is performed such that

$$K \left[1 + \frac{\gamma (1 - \beta(1-t)^{\beta-1})}{\beta - 1} \right] \approx \Psi(t \mathbf{1}), \quad t \in [0, 1]. \quad (5.10)$$

6 Rare event analysis

As the importance sampling technique is intended to be used in cases where the functional Ψ is large on sets which relate to rare events, we may want to study the efficiency of the algorithm in a rare event setting. We shall consider $\Psi^{(s)}(\mathbf{u})$ as a functional that will take non-zero values only on a small probability set. Let $p^{(s)} = \mathbb{E} [\Psi^{(s)}(\mathbf{U})]$ be the probability of interest. The rare event setting assumes that $\lim_{s \rightarrow 1} p^{(s)} = 0$. For each s , we would select a new mixing distribution $F_\Lambda^{(s)}$, therefore changing the calibration of the proposal distribution $F_{\mathbf{V}}^{(s)}$ and its sampling cost that we shall denote $T(s)$. In the direct sampling algorithm, see Section 5, this sampling cost is finite and constant in s , whereas it is of order $\mathbb{E}[(1 - \Lambda)^{-1}]$, see Theorem 4.3, in the rejection sampling algorithm from Section 4.

Let $\hat{\mu}_n^{(s)} = n^{-1} \sum_{i=1}^n \Psi^{(s)}(\mathbf{V}_i) w^{(s)}(\mathbf{V}_i)$ be the importance sampling estimate. In a rare-event setting, we would ideally aim for a bounded relative error as $s \rightarrow 1$, see Chapter VI in [6], that is

$$\limsup_{s \rightarrow 1} \frac{\text{var} [\hat{\mu}_n^{(s)}]}{(p^{(s)})^2} T(s) < \infty. \quad (6.1)$$

Replacing $\text{var} [\widehat{\mu}_n^{(s)}]$ by its upper bound $n^{-1} \mathbb{E} [\Psi^{(s)}(\mathbf{V})^2 w^{(s)}(\mathbf{V})^2]$, we shall aim for an algorithm that satisfies

$$\limsup_{s \rightarrow 1} \frac{n^{-1} \mathbb{E} [\Psi^{(s)}(\mathbf{V})^2 w^{(s)}(\mathbf{V})^2]}{(p^{(s)})^2} T(s) < \infty. \quad (6.2)$$

Note first that the optimality condition (4.4) guarantees that $\mathbb{E} [\Psi^{(s)}(\mathbf{V})^2 w^{(s)}(\mathbf{V})^2] / (p^{(s)})^2 \propto 1$. We now assume a mild condition for the calibration of $F_\Lambda^{(s)}$ that will be needed to obtain the efficiency criteria (6.2).

Condition B. *For all $s > 0$, the discrete distribution of Λ is constructed such that there exists $k \in \{1, \dots, n_\Lambda\}$ with $x_k = s$ and $p_k > 0$.*

We first study the case of rejection sampling from Section 4, limiting ourselves to the setup of a discrete distribution for Λ . Although typical rare event sets in the literature consider the sum of margins, we will consider the maximum instead, which allows us to stay within a rare event framework since $\{\max_i u_i > s\} \subseteq \{\sum_i u_i > s\}$, $\mathbf{u} \in [0, 1]^d$.

Theorem 6.1. *Assume that $\Psi^{(s)}(\mathbf{u}) = \mathbf{1}\{\max_i u_i > s\}$ and that the proposal distribution $F_{\mathbf{V}}$ and the corresponding weight function $w(\mathbf{u})$ are chosen as in Section 4. In addition, assume F_Λ is a discrete distribution with a finite number n_Λ of atoms $\mathbb{P}[\Lambda = x_k] = p_k$, $k = 1, \dots, n_\Lambda$, $0 = x_1 < \dots < x_{n_\Lambda} < 1$, calibrated as in Algorithm 4.6 and that the Condition B holds. Denote $k_{\mathbf{u}}^* = \max\{1 \leq k \leq n_\Lambda : x_k \leq \max_i u_i\}$, $\mathbf{u} \in [0, 1]^d$. Then*

$$\limsup_{s \rightarrow 1} \frac{\mathbb{E} [\Psi^{(s)}(\mathbf{V})^2 w^{(s)}(\mathbf{V})^2]}{(p^{(s)})^2} < \infty.$$

Proof. Under Condition B, we have that $x_{k_{\mathbf{u}}^*} \geq s$ on the event $\{\max_i u_i > s\}$. Therefore,

$$\begin{aligned} & \int_{[0,1]^d} \Psi^{(s)}(\mathbf{v})^2 w^{(s)}(\mathbf{v})^2 dF_{\mathbf{V}}(\mathbf{v}) = \int_{[0,1]^d} \Psi^{(s)}(\mathbf{u}) w^{(s)}(\mathbf{u}) dC(\mathbf{u}) \\ &= \int_{[0,1]^d} \Psi^{(s)}(\mathbf{u}) \left(\sum_{k=1}^{n_\Lambda} \tilde{p}_k \right) \left(\sum_{k=1}^{n_\Lambda} \frac{\mathbf{1}\{x_k \leq \max_i u_i\} \tilde{p}_k}{1 - C(x_k \mathbf{1})} \right)^{-1} dC(\mathbf{u}) \\ &= \int_{[0,1]^d} \Psi^{(s)}(\mathbf{u}) \left(\sum_{k=1}^{n_\Lambda} \Psi^{(s)}(x_k \mathbf{1}) (C(x_k \mathbf{1}) - C(x_{k-1} \mathbf{1})) \right) \Psi^{(s)}(x_{k_{\mathbf{u}}^*} \mathbf{1})^{-1} dC(\mathbf{u}) \\ &= \int_{[0,1]^d} \Psi^{(s)}(\mathbf{u}) (C(x_{n_\Lambda} \mathbf{1}) - C(s \mathbf{1})) dC(\mathbf{u}) \\ &\leq (1 - C(s \mathbf{1})) \int_{[0,1]^d} \Psi^{(s)}(\mathbf{u}) dC(\mathbf{u}) = (p^{(s)})^2, \end{aligned} \quad (6.3)$$

which proves the theorem. \square

Note that Theorem 6.1 guarantees a bounded relative error as in (6.1) whenever $\limsup_{s \rightarrow 1} T(s) < \infty$. This would not hold for the rejection algorithm. Indeed, since $\mathbb{E}[(1 - \Lambda)^{-1}] = \sum_{k=1}^{n_\Lambda} p_k / (1 - x_k)$, we obtain in virtue of Theorem 4.3 that $\limsup_{s \rightarrow 1} T(s) = \infty$ under Condition B.

In the case of direct sampling, we can prove the corresponding version of Theorem 6.1 by taking $\Psi^{(s)}(\mathbf{u}) = \mathbf{1}\{u_i > s\}$ for any $i = 1, \dots, d$ and $k_{\mathbf{u}}^* = \max\{1 \leq k \leq n_{\Lambda} : x_k \leq u_i\}$, $\mathbf{u} \in [0, 1]^d$. Since this algorithm has a computational cost $T(s)$ constant in s , it shall then be preferred to the rejection sampling algorithm in rare event settings, although it may require more implementation efforts.

The calibration of the proposal distribution $F_{\mathbf{V}}$ is profiled on the assumption that $\Psi(\mathbf{u}) \approx \Psi(\max_i u_i \mathbf{1})$. In Theorem 6.1 we have been able to show that when $\Psi(\mathbf{u}) = \mathbf{1}\{\max_i u_i > s\}$ for some $s \in (0, 1)$, i.e. when the assumption holds with equality, then $\mathbb{E}[\Psi(\mathbf{V})^2 w(\mathbf{V})^2] \leq \mathbb{E}[\Psi(\mathbf{U})]^2$. By Jensen's inequality we obtain that $\mathbb{E}[\Psi(\mathbf{V})^2 w(\mathbf{V})^2] \leq \mathbb{E}[\Psi(\mathbf{U})^2]$, and thus that $\text{var}(\hat{\mu}_n) \leq \text{var}(\mu_n)$, so a smaller estimator's variance. Although the assumption that $\Psi(\mathbf{u}) \approx \Psi(\max_i u_i \mathbf{1})$ is typical of application in insurance mathematics, it does not often hold with equality and thus cannot be easily incorporated into an analytical framework that would allow to prove a certain variance reduction factor. However, we illustrate in the numerical Case Study of Section 7 that we obtain a substantial variance reduction for several typical insurance problems.

7 Case study

In this section, we illustrate the performance of our importance sampling algorithms for functionals Ψ relevant for insurance applications. We shall use the two importance sampling algorithms defined in Section 4 and Section 5 on the same example. We use three random vectors, of dimension $d = 2$, $d = 5$, and $d = 25$, respectively. Our case study will assume that marginal distributions of $\mathbf{X} = (X_1, \dots, X_d)$ are lognormal, parametrized as $X_j \sim \text{LN}(10 - 0.1j, 1 + 0.2j)$, $j = 1, \dots, d$, which yields equal expectation for each margin, i.e. $\mathbb{E}[X_j] = 36\,315.5$ and standard deviation $\mathbb{E}[X_j] \sqrt{e^{1+0.2j} - 1}$. We will consider two examples of copulas, namely Clayton and Gumbel. Kendall's tau, see, e.g., Section 5.1.1 in [99], between each pair of margins is $1/3$, which yields a Clayton parameter of 1 and a Gumbel parameter of 1.5. Note that our importance sampling method does not rely on particular assumptions on the copula. In consequence, the general behavior of the algorithm does not significantly change with the strength of the dependence. This case study has been implemented using the R package `copula`.

We investigate the estimation of the following five functionals of \mathbf{X} . All are formulated in terms of the aggregate losses $S = \sum_{j=1}^d X_j$, which is inspired by risk aggregation problems arising frequently in actuarial practice:

- $\mathbb{E}[\max\{S - T, 0\}]$, which is the fair premium of a stop-loss cover with deductible T . For T we use $T = 10^5 d$, which is approximately 3 times the expectation of S ;
- $\text{VaR}_{0.995}(S)$ and $\text{ES}_{0.99}(S)$, which are the risk measures determining solvency capital under Solvency II and the Swiss Solvency Test, respectively (see [52] and [24]);
- $\mathbb{E}[X_1 | S > F_S^{-1}(0.99)]$ and $\mathbb{E}[X_d | S > F_S^{-1}(0.99)]$, which represent the capital allocated to the first and last margin under the Euler principle, see [122].

For ease of calibration and simulation, we use the discrete framework for F_{Λ} . As we want to use the same sample to estimate all objective functions, we only conduct one calibration of F_{Λ} for each problem dimension. Recall from Section 2 that VaR_{α} and ES_{α} cannot be written as an

expectation of type $\mathbb{E}[\Psi_0(\mathbf{X})]$. We thus calibrate F_Λ using the stop-loss objective function $\tilde{\Psi}(\mathbf{u}) = \max\{\sum_{j=1}^d F_j^{-1}(u_j) - T, 0\}$. This is non-zero only for $\sum_{j=1}^d F_j^{-1}(u_j)$ above the deductible T , so that calibration with this function will favour a high concentration of distorted samples in the region of interest for our applications. Note that the calibration of F_Λ depends on the choice of copula and of the importance sampling algorithm, see Sections 4.3.1 and 5.3.1. The number of discretization points is set to $n_\Lambda = 10$. As shown in Table 1, the highest point $x_{10} = 1 - (1/2)^9 \approx 0.998$, which is well beyond the highest VaR level under consideration. In order to satisfy Condition A, we manually set the weight of x_0 to be $p_0 = 0.1$ and decrease the other weights proportionally. The weights p_k for $k = 1, \dots, n_\Lambda$ resulting from the calibration using the Gumbel copula and the rejection sampling approach are shown in Table 1 for dimensions $d = 2, 5$ and 25.

	k	1	2	3	4	5	6	7	8	9	10
	x_k	0.000	0.500	0.750	0.875	0.937	0.968	0.984	0.992	0.996	0.998
$d = 2$	p_k	0.100	0.000	0.000	0.000	0.115	0.325	0.206	0.128	0.787	0.048
$d = 5$	p_k	0.100	0.000	0.000	0.000	0.129	0.302	0.202	0.131	0.084	0.053
$d = 25$	p_k	0.100	0.000	0.000	0.000	0.022	0.252	0.216	0.174	0.135	0.102

Table 1: Calibrated probability weights p_k using the Gumbel copula.

The weight functions $\tilde{w}(\cdot) = w(\cdot \mathbf{1})$ and a scatter plot of 5000 samples of \mathbf{V} are plotted in Figure 2, when the reference copula is Gumbel and the rejection sampling approach is used. Given the setup of this case study, it is easy to check that Lemma 3.2 is satisfied. Due to the construction of $F_{\mathbf{V}}$, more samples lie close to the upper or right border than what would be observed for a copula sample.

As the objective functions use estimates of the distribution function of S , we normalize the sample weights to sum to 1. This further reduces the estimation error as advocated in Section 4.2.2 in [55] or Section 2.5.2 in [85].

In order to assess the improvements provided by importance sampling, we present a simulation study for $d \in \{2, 5, 25\}$. We use a sample size of $n = 10\,000$ to calculate the importance sampling estimators $\hat{\mu}_n$ for each of the two algorithms and the standard Monte Carlo estimators μ_n for all objective functions. A total of 500 repetitions is used to obtain an empirical distribution of these estimators, and thus to estimate their variance. Although the sampling size has an impact on the value of the sample variance, it should not play a significant role in the study of algorithmic efficiency. The results are presented in Tables 2, 3 for the Gumbel and Clayton cases using the rejection algorithm, and Tables 4, 5 for the Gumbel and Clayton cases using the direct algorithm. Although the value of the estimates may be different depending on which algorithm is used for sampling, we only present the reference value obtained from the plain Monte Carlo simulation since our empirical study has shown only negligible differences. The main results in these tables are in the form of variance reduction factors, which represent the sample variance of the plain Monte Carlo estimator divided by the sample variance of our importance sampling estimator.

Tables 2, 3, 4 and 5 show that the importance sampling algorithms greatly decrease the estimation error for all objective functions. It is not surprising that the largest reduction is for the

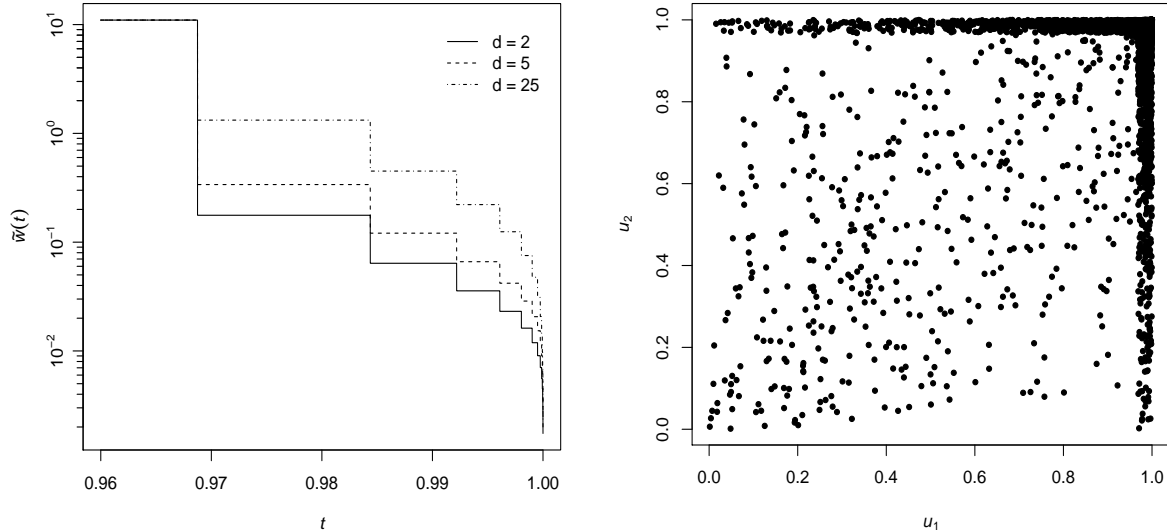


Figure 2: *Left:* Discrete weight function $\tilde{w}(t) = w(t\mathbf{1})$ for $t \in [0.96, 1]$ for the Gumbel copula, for dimensions $d \in \{2, 5, 25\}$ and using the rejection sampling approach. From 0 to 0.968 the function \tilde{w} is constant. *Right:* A scatter plot of 5000 samples of \mathbf{V} sampled using the rejection sampling approach.

stop-loss cover, since F_Λ is calibrated to this functional. A larger reduction for the other functionals could be achieved with a specific calibration for each of them. The smallest reduction factors are for $\text{VaR}_{0.995}(S)$, because this functional does not depend on the tail behaviour of S beyond the 99.5% quantile, where the largest gain in accuracy is obtained by our importance sampling approach. The variance reduction can also be observed from the boxplots in Figures 3 and 4 that allow us to compare the entire distribution of the plain Monte Carlo to the importance sampling estimators for $\text{VaR}_{0.995}(S)$ and $\text{ES}_{0.99}(S)$, using the two algorithms for $d = 5$. We note that the bias indeed appears negligible and that variance of these empirical distributions has been greatly reduced by the importance sampling approaches.

In order to fairly assess the efficiency of the method and to compare the two sampling algorithms, one should divide the variance reduction factors in Tables 2 and 3 by the expected waiting time of the rejection sampling algorithm, $\mathbb{E}[N_{\mathbf{V}}]$, given in Table 6.

In most cases, the expected waiting time is larger than the variance reduction ratio, hence rendering the rejection algorithm inefficient for this case study. We recall that this waiting time issue is only a concern when using the rejection sampling algorithm from Section 4.

Although the conditional sampling algorithm might be a bit more computationally intensive than the direct sampling of the copula C , e.g., when inverting the conditional distributions in Algorithm 5.2 for certain copulas, this complexity is insignificant and does not become more pronounced if one puts more mass of Λ towards 1. For this reason, we conclude that the efficiency of the importance sampling method is not reduced with the conditional sampling approach.

Objective function	$d = 2$		$d = 5$		$d = 25$	
	Ref. val.	Red. fact.	Ref. val.	Red. fact.	Ref. val.	Red. fact.
$\mathbb{E}[(S - T)^+]$	10 498	80.8	29 648	39.1	310 499	17.3
$\text{VaR}_{0.995}(S)$	645 162	12.4	1 795 071	11.5	15 183 823	9.9
$\text{ES}_{0.99}(S)$	774 616	18.6	2 241 589	17.5	24 541 482	13.9
$\mathbb{E}[X_1 S > F_S^{-1}(0.99)]$	351 077	21.4	332 560	19.3	324 231	11.3
$\mathbb{E}[X_d S > F_S^{-1}(0.99)]$	423 539	22.3	570 105	18.1	1 676 897	17.7

Table 2: Reference values (Ref. val.) of the objective functions and variance reduction factors (Red. fact.) with F_λ discrete and C Gumbel copula, using the rejection sampling algorithm.

Objective function	$d = 2$		$d = 5$		$d = 25$	
	Ref. val.	Red. fact.	Ref. val.	Red. fact.	Ref. val.	Red. fact.
$\mathbb{E}[(S - T)^+]$	7 765	63.08	13 657	23.59	119 531	9.05
$\text{VaR}_{0.995}(S)$	526 254	14.14	1 101 395	10.60	7 235 669	6.05
$\text{ES}_{0.99}(S)$	610 928	19.74	1 272 925	14.84	9 963 262	8.47
$\mathbb{E}[X_1 S > F_S^{-1}(0.99)]$	259 814	31.18	139 127	19.18	68 702	5.62
$\mathbb{E}[X_d S > F_S^{-1}(0.99)]$	351 113	26.04	384 475	16.92	1 009 675	15.81

Table 3: Reference values (Ref. val.) of the objective functions and variance reduction factors (Red. fact.) with F_λ discrete and C Clayton copula, using the rejection sampling algorithm.

8 Conclusion

We proposed an importance sampling approach for copula models with two algorithms, specifically designed for problems arising frequently in insurance and financial applications.

The starting point for the construction of an alternative sampling distribution was to consider the copula conditional on the event that some of its components exceed a certain threshold. In the rejection sampling approach, we require that the maximum of all components exceeds the threshold. In the direct sampling approach, we require that a specific component exceeds the threshold. The proposal distribution has then been constructed by mixing over different thresholds.

In order to minimize the estimation error of the importance sampling estimator, we proposed several procedures to set up and optimize the mixing distribution. Unlike other importance sampling approaches, our method does not have requirements on the original copula and it can be applied to any copula from which sampling is feasible.

The variance reduction of our approach has only been shown analytically for a simplified case. Through a case study inspired by a typical insurance application, we have shown that the rejection and the direct sampling algorithms are able to largely reduce simulation errors in more general estimation problems relevant to actuarial practitioners.

In the rejection sampling approach, sampling the proposal distribution can easily be implemented through a rejection algorithm, which only requires that samples from the original copula

Objective function	$d = 2$		$d = 5$		$d = 25$	
	Ref. val.	Red. fact.	Ref. val.	Red. fact.	Ref. val.	Red. fact.
$\mathbb{E}[(S - T)^+]$	10 498	116.03	29 648	80.27	310 499	21.71
$\text{VaR}_{0.995}(S)$	645 162	14.25	1 795 071	15.83	15 183 823	8.97
$\text{ES}_{0.99}(S)$	774 616	20.98	2 241 589	19.78	24 541 482	12.14
$\mathbb{E}[X_1 S > F_S^{-1}(0.99)]$	351 077	23.84	332 560	19.01	324 231	11.85
$\mathbb{E}[X_d S > F_S^{-1}(0.99)]$	423 539	23.87	570 105	20.67	1 676 897	19.52

Table 4: Reference values (Ref. val.) of the objective functions and variance reduction factors (Red. fact.) with F_λ discrete and C Gumbel copula, using the direct sampling algorithm.

Objective function	$d = 2$		$d = 5$		$d = 25$	
	Ref. val.	Red. fact.	Ref. val.	Red. fact.	Ref. val.	Red. fact.
$\mathbb{E}[(S - T)^+]$	7 765	72.17	13 657	22.34	119 531	5.82
$\text{VaR}_{0.995}(S)$	526 254	14.74	1 101 395	11.05	7 235 669	6.33
$\text{ES}_{0.99}(S)$	610 928	20.18	1 272 925	12.60	9 963 262	5.23
$\mathbb{E}[X_1 S > F_S^{-1}(0.99)]$	259 814	31.41	139 127	14.93	68 702	10.55
$\mathbb{E}[X_d S > F_S^{-1}(0.99)]$	351 113	25.57	384 475	14.84	1 009 675	10.98

Table 5: Reference values (Ref. val.) of the objective functions and variance reduction factors (Red. fact.) with F_λ discrete and C Clayton copula, using the direct sampling algorithm.

can be drawn. It is acknowledged that the computational cost of the algorithm is increased due to the rejection sampling procedure, which however is not always a disadvantage. In addition, the direct sampling algorithm based on the inversion of conditional distributions has been proposed. Although it requires a more advanced implementation, this algorithm has the striking advantage that it has a reduced computational complexity, of order of the cost of sampling C and that it does not depend on the calibration of the proposal distribution $F_{\mathbf{V}}$. We have also shown that the later algorithm is efficient in a rare-event setting in the sense of [6].

For further research, we emphasize the problem of sampling from conditional distributions such as the $C^{[\lambda]}$'s proposed in Section 4 and 5. One could aim at finding copulas for which the sampling of $C^{[\lambda]}$ is sufficiently simple, an example is given in Appendix A. More generally, families of copulas such that $C^{[\lambda]}$ stays within the same class for all $\lambda \in [0, 1]^d$ could be of interest. Note that copulas that are invariant under conditioning on subregions of $[0, 1]^d$ have been investigated in [26], [71] or [36]. However, the conditional regions are always d -rectangles such as $\prod_{i=1}^d [\alpha_i, \beta_i] \subseteq [0, 1]^d$. The regions we condition on in the rejection sampling approach, $\{\max_i u_i > \lambda\}$, are unions of stripes.

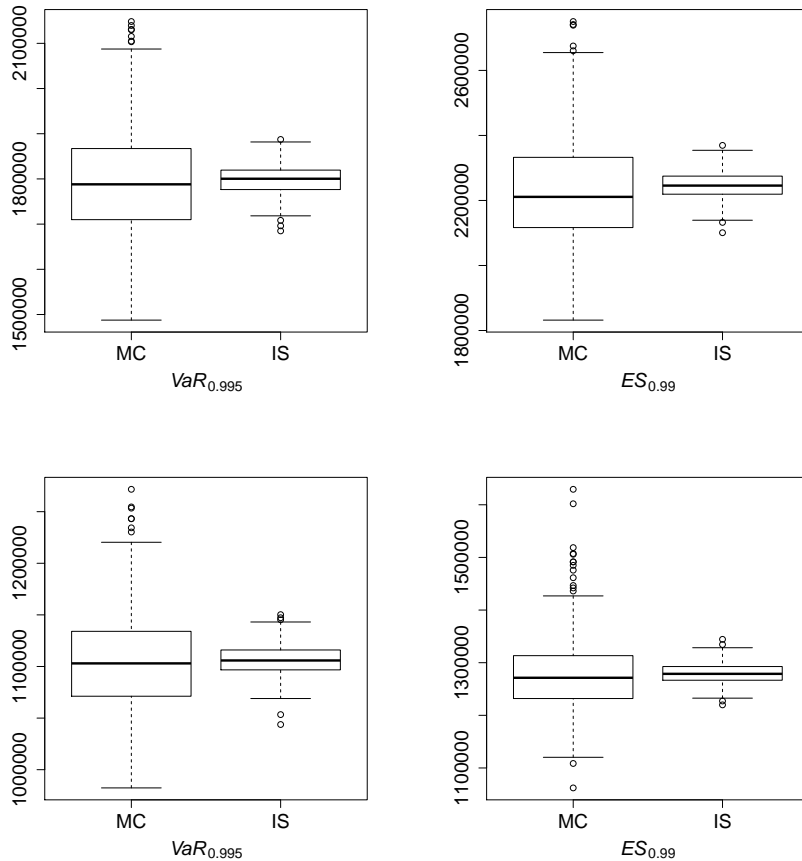


Figure 3: Boxplots for the importance sampling estimators of $\text{VaR}_{0.995}(S)$ and $\text{ES}_{0.99}(S)$ from the $N = 500$ independent copies of the estimator, using the rejection sampling algorithm in the Gumbel (*top*) and Clayton (*bottom*) cases for $d = 5$.

A Direct sampling of $C^{[\lambda]}$ for shock copulas

This appendix shows that for a certain class of shock copulas, it is possible to directly sample from the conditional distribution $C^{[\lambda]}$, as defined in Section 4. The Marshall–Olkin copula is a special case of this class.

We now introduce a multivariate construction for shock copulas. Let $Z_j : \Omega \rightarrow \mathbb{R}$, $j = 1, \dots, m$, for some $m \in \mathbb{N}$, denote continuous independent random variables. We call the Z_j “shocks” and denote their cdf’s by F_{Z_j} . Suppose each component X_j of $\mathbf{X} = (X_1, \dots, X_d)$ is exposed to a subset of shocks with indices $I_j \subset \{1, \dots, m\}$ through the maximum:

$$(X_1, \dots, X_d) = \left(\max_{k \in I_1} Z_k, \dots, \max_{k \in I_d} Z_k \right). \quad (\text{A.1})$$

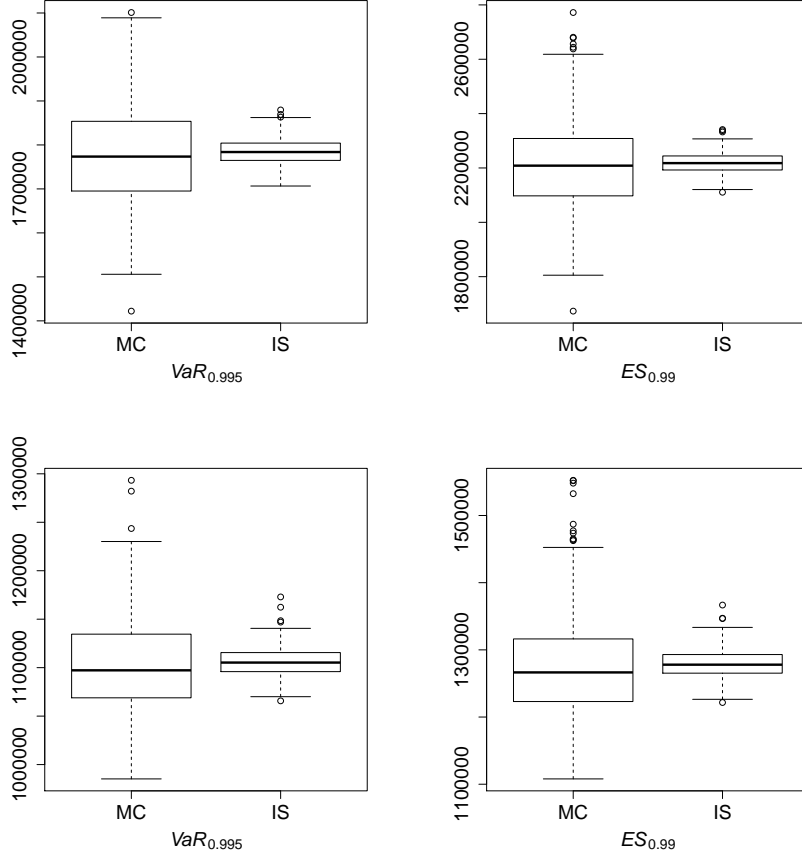


Figure 4: Boxplots for the importance sampling estimators of $\text{VaR}_{0.995}(S)$ and $\text{ES}_{0.99}(S)$ from the $N = 500$ independent copies of the estimator, using the direct sampling algorithm in the Gumbel (*top*) and Clayton (*bottom*) cases for $d = 5$.

As the Z_j 's are independent, the marginal cdf's F_{X_j} can be calculated as

$$F_{X_j}(x) = \prod_{k \in I_j} F_{Z_k}(x), \quad x \in \mathbb{R}. \quad (\text{A.2})$$

By rearranging the arguments, and due to the fact that the Z_j 's are independent, we can write the joint distribution of \mathbf{X} as

$$\mathbb{P}[X_1 \leq x_1, \dots, X_d \leq x_d] = \prod_{j=1}^m \mathbb{P} \left[Z_j \leq \min_{k: j \in I_k} x_k \right].$$

Hence, the copula induced by \mathbf{X} is given by

$$C(\mathbf{u}) = \prod_{j=1}^m F_{Z_j} \left(\min_{k: j \in I_k} F_{X_k}^{-1}(u_k) \right). \quad (\text{A.3})$$

	$d = 2$	$d = 5$	$d = 25$
Clayton ($\theta = 1$)	44.16	19.48	6.89
Gumbel ($\theta = 1.5$)	54.69	31.11	15.83

Table 6: Expected waiting time, $\mathbb{E}[N_{\mathbf{V}}]$, of the rejection sampling algorithm.

As the copula can be expressed in terms of the independent shocks, we can write the conditional distribution $C^{[\lambda]}$ in a tractable form. To this end, let the constants ϕ_j and the random variables B_j for $j = 1, \dots, m$ be defined by

$$\phi_j = \min_{k: j \in I_k} F_{X_k}^{-1}(\lambda), \quad B_j = \mathbf{1}\{Z_j > \phi_j\}.$$

Then, we can express conditioning on $\mathbf{U} \notin [0, \lambda]^d$ through the following equivalent statements

$$\begin{aligned} \mathbf{U} \notin [0, \lambda]^d &\Leftrightarrow \max\{U_1, \dots, U_d\} > \lambda \\ &\Leftrightarrow X_i > F_{X_i}^{-1}(\lambda) \quad \text{for at least one } i \in \{1, \dots, d\} \\ &\Leftrightarrow \max_{k \in I_i} Z_k > F_{X_i}^{-1}(\lambda) \quad \text{for at least one } i \in \{1, \dots, d\} \\ &\Leftrightarrow Z_j > F_{X_k}^{-1}(\lambda) \quad \text{for at least one } j \in \{1, \dots, m\} \text{ and } k \text{ s.t. } j \in I_k \\ &\Leftrightarrow Z_j > \phi_j \quad \text{for at least one } j \in \{1, \dots, m\} \\ &\Leftrightarrow \max\{B_1, \dots, B_m\} = 1. \end{aligned}$$

Note that, unconditionally, the B_j 's are independent Bernoulli variables with parameter

$$p_j = \mathbb{P}[B_j = 1] = 1 - F_{Z_j}(\phi_j), \quad j = 1, \dots, m.$$

The following algorithm can be used to draw a realization from $C^{[\lambda]}$. First, a realization from the conditional distribution of (B_1, \dots, B_m) given that $\max\{B_1, \dots, B_m\} = 1$ is drawn through iterative conditioning. Then the shocks are simulated conditionally on the B_j 's, which is easy as the shocks are independent under this conditioning. Finally, by calculating the corresponding realization of \mathbf{X} with (A.1), we obtain the sample from $C^{[\lambda]}$. This approach is fast because the conditional distribution of (B_1, \dots, B_m) given $\max\{B_1, \dots, B_m\} = 1$ is analytically tractable, as the following algorithm also shows.

Algorithm A.1. *In order to draw a realization from $C^{[\lambda]}$, do:*

1. Draw a realization from (B_1, \dots, B_m) given that $\max\{B_1, \dots, B_m\} = 1$ through iterative conditioning as follows.

For $k = 1, \dots, m$:

(a) Set

$$\tilde{p}_k = \begin{cases} \frac{p_k}{1 - \prod_{l=k}^m (1 - p_l)}, & \text{if } k = 1 \text{ or } \max_{1 \leq j < k} B_j = 0, \\ p_k, & \text{if } \max_{1 \leq j < k} B_j = 1. \end{cases}$$

(b) Draw $B_k \sim \text{Bernoulli}(\tilde{p}_k)$.

2. Draw a realization from (Z_1, \dots, Z_m) given (B_1, \dots, B_m) as follows:

For $k = 1, \dots, m$:

(a) Draw $\tilde{U}_k \sim \text{U}(0, 1)$

(b) Set

$$Z_k = \begin{cases} F_{Z_k}^{-1}\left(p_k \tilde{U}_k\right), & \text{if } B_k = 0, \\ F_{Z_k}^{-1}\left(p_k + (1 - p_k)\tilde{U}_k\right), & \text{if } B_k = 1. \end{cases}$$

3. Set $X_j = \max_{k \in I_j} Z_k$ and $U_j = F_{X_j}(X_j)$, where F_{X_j} is defined in (A.2).

4. Return $\mathbf{U} = (U_1, \dots, U_d)$.

Although it is not an issue for the purpose of sampling, note that for most choices of shock distributions F_{Z_j} , the copula C in (A.3) does not have an analytic form. One possible choice for F_{Z_j} yielding an analytic expression for C is illustrated in the following example.

Example A.2 (Marshall–Olkin copula). *Suppose the Z_j 's are Fréchet distributed with $F_{Z_j}(x) = \exp(-s_j/x)$, $x > 0$, $j = 1, \dots, m$, with scale parameters $s_j > 0$. Then the X_j are also Fréchet distributed, with $F_{X_j}(x) = \exp(-\tilde{s}_j/x)$, where $\tilde{s}_j = \sum_{k \in I_j} s_k$, $j = 1, \dots, d$. The copula (A.3) then reduces to*

$$C(\mathbf{u}) = \prod_{j=1}^m \exp \left\{ -s_j \left(\min_{i: j \in I_i} \left(\frac{-\log u_i}{\tilde{s}_i} \right)^{-1} \right)^{-1} \right\} = \prod_{j=1}^m \min_{i: j \in I_i} u_i^{s_j/\tilde{s}_i}.$$

This copula is of the Marshall–Olkin type, see [87]. As an example, consider $d = 2$, $m = 3$, $I_1 = \{1, 3\}$, and $I_2 = \{2, 3\}$. In this case, $X_1 = \max\{Z_1, Z_3\}$, $X_2 = \max\{Z_2, Z_3\}$, and the copula can be written as

$$C(u_1, u_2) = u_1^{s_1/(s_1+s_3)} \cdot u_2^{s_2/(s_2+s_3)} \cdot \min \left\{ u_1^{s_3/(s_1+s_3)}, u_2^{s_3/(s_2+s_3)} \right\} = \min \left\{ u_1 u_2^{s_2/(s_2+s_3)}, u_2 u_1^{s_1/(s_1+s_3)} \right\}.$$

Paper C

Quasi-Random Numbers for Copula Models

Quasi-random numbers for copula models*

Mathieu Cambou[†] Marius Hofert[‡] Christiane Lemieux[§]

March 01, 2016

Abstract

The present work addresses the question how sampling algorithms for commonly applied copula models can be adapted to account for quasi-random numbers. Besides sampling methods such as the conditional distribution method (based on a one-to-one transformation), it is also shown that typically faster sampling methods (based on stochastic representations) can be used to improve upon classical Monte Carlo methods when pseudo-random number generators are replaced by quasi-random number generators. This opens the door to quasi-random numbers for models well beyond independent margins or the multivariate normal distribution. Detailed examples (in the context of finance and insurance), illustrations and simulations are given and software has been developed and provided in the R packages `copula` and `qrng`.

Key words: Quasi-random numbers, copulas, conditional distribution method, Marshall–Olkin algorithm, tail events, risk measures

1 Introduction

In many applications, in particular in finance and insurance, the quantities of interest can be written as $\mathbb{E}[\Psi_0(\mathbf{X})]$, where $\mathbf{X} = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ is a random vector with distribution function H on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\Psi_0 : \mathbb{R}^d \rightarrow \mathbb{R}$ is a measurable function. The components of \mathbf{X} are typically dependent. To account for this dependence, the distribution of \mathbf{X} can be modeled by

$$H(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d)), \quad \mathbf{x} \in \mathbb{R}^d, \quad (1)$$

where $F_j(x) = \mathbb{P}(X_j \leq x)$, $j \in \{1, \dots, d\}$, are the marginal distribution functions of H and $C : [0, 1]^d \rightarrow [0, 1]$ is a *copula*, i.e., a distribution function with standard uniform univariate margins; for an introduction to copulas, see [89, Chapter 5], [99] or [73]. A copula model such as (1) allows one to separate the dependence structure from the marginal distributions. This is especially useful

*We thank the Associate Editor and the two anonymous reviewers for their helpful comments, which helped us improve this paper. The first author wishes to thank SCOR for their financial support. The second and third authors acknowledge the support of NSERC through grants #5010 and #238959, respectively.

[†]EPFL, Institute of Mathematics, Station 8, EPFL, 1015 Lausanne, Switzerland
email: mathieucambou@gmail.com

[‡]Department of Statistics and Actuarial Science, University of Waterloo, Canada
email: marius.hofert@uwaterloo.ca

[§]Department of Statistics and Actuarial Science, University of Waterloo, Canada
email: clemieux@uwaterloo.ca

in the context of model building and sampling in the case where $\mathbb{E}[\Psi_0(\mathbf{X})]$ mainly depends on the dependence between the components of \mathbf{X} , so on C ; for examples of this type, see Section 5.

In terms of copula model (1), we may then write

$$\mathbb{E}[\Psi_0(\mathbf{X})] = \mathbb{E}[\Psi(\mathbf{U})]$$

where $\mathbf{U} = (U_1, \dots, U_d) : \Omega \rightarrow \mathbb{R}^d$ is a random vector with distribution function C , $\Psi : [0, 1]^d \rightarrow \mathbb{R}$ is defined as

$$\Psi(u_1, \dots, u_d) = \Psi_0(F_1^-(u_1), \dots, F_d^-(u_d)),$$

and $F_j^-(p) = \inf\{x \in \mathbb{R} : F_j(x) \geq p\}$, $j \in \{1, \dots, d\}$, are the marginal quantile functions. If C and the margins F_j , $j \in \{1, \dots, d\}$, are known, we can use Monte Carlo simulation to estimate $\mathbb{E}[\Psi(\mathbf{U})]$. For a (pseudo-)random sample $\{\mathbf{U}_i : i = 1, \dots, n\}$ from C , the (classical) Monte Carlo estimator of $\mathbb{E}[\Psi(\mathbf{U})]$ is given by

$$\frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{U}_i) \approx \mathbb{E}[\Psi(\mathbf{U})].$$

The main challenge of a Monte Carlo simulation is thus the sampling of the copula. This challenge also holds for quasi-Monte Carlo (QMC) methods, and is actually amplified by the fact that these methods are more sensitive to certain properties of the function Ψ . Thus the choice of the construction method of a stochastic representation for C can have complex effects on the performance of QMC methods, a feature that does not show up when using Monte Carlo methods. The present work includes a careful analysis of these effects, as they must be thoroughly understood in order to successfully replace pseudorandom numbers by quasi-random numbers into different copula sampling algorithms.

Let us briefly summarize the idea behind QMC methods and how they can be used for copula models; more precise definitions on some of the concepts used here will be given later. The idea is to start with a so-called *low-discrepancy point set* $P_n = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subseteq [0, 1]^k$, with $k \geq d$, which is designed so that its empirical distribution over $[0, 1]^k$ is closer (in a sense to be defined later) to the uniform distribution $U[0, 1]^k$ than a set of independent and identically (i.i.d.) random points is. Assuming that for $\mathbf{U}' \sim U[0, 1]^k$ we have a transformation ϕ_C such that $\phi_C(\mathbf{U}') \sim C$, we can then construct the approximation

$$\frac{1}{n} \sum_{i=1}^n \Psi(\phi_C(\mathbf{v}_i)) \approx \mathbb{E}[\Psi(\mathbf{U})]. \quad (2)$$

Figure 1 shows the points $\phi_C(\mathbf{v}_i)$ obtained using either pseudo-random or quasi-random numbers, for a transformation ϕ_C designed for the *Clayton copula*.

QMC methods are typically used to approximate integrals of functions over the unit cube via

$$Q_n = \frac{1}{n} \sum_{i=1}^n f(\mathbf{v}_i) \approx \int_{[0,1]^k} f(\mathbf{v}) d\mathbf{v} = I(f). \quad (3)$$

A widely used upper bound for the integration error $|I(f) - Q_n|$ is the Koksma–Hlawka inequality (see, e.g., [101]), which is of the form $V(f)D^*(P_n)$, where $V(f)$ measures the variation of f in the

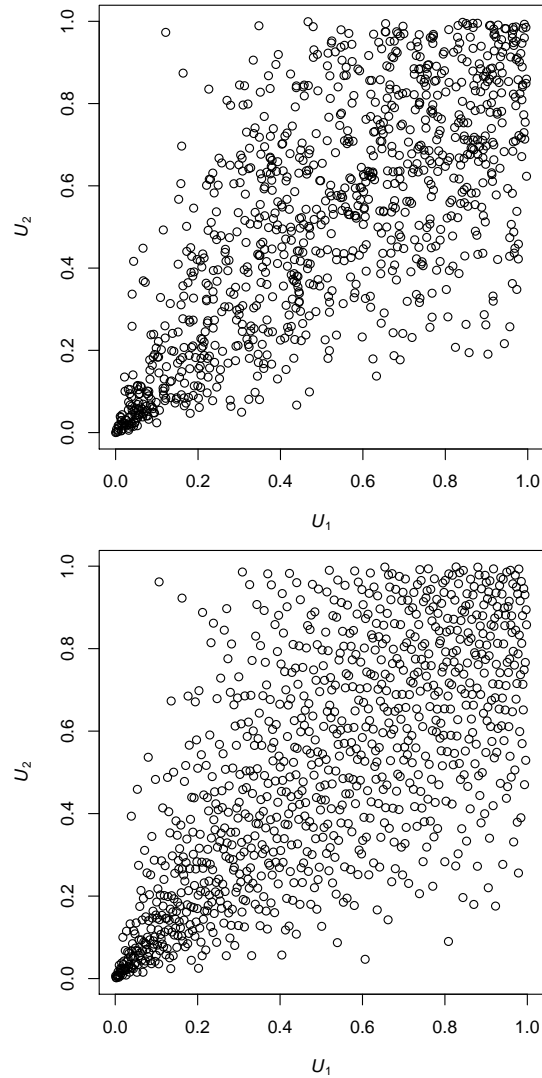


Figure 1: 1000 realizations of a bivariate Clayton copula with $\theta = 2$ (Kendall's tau equals 0.5), generated by a pseudo-random number generator (top) and by a quasi-random number generator (bottom).

sense of Hardy and Krause, while $D^*(P_n)$ measures the discrepancy of P_n , i.e., how far P_n is from $U[0, 1]^k$.

To analyze the properties of the QMC approximation (2) for $\mathbb{E}[\Psi(\mathbf{U})]$, there are two possible approaches. The first one is to define $f = \Psi \circ \phi_C$ and work within the traditional framework given by (3), the Koksma–Hlawka inequality with this composed function and the low-discrepancy point set P_n . The second one is to think of (2) as approximating

$$\mathbb{E}[\Psi(\mathbf{U})] = \int_{[0,1]^d} \Psi(\mathbf{u})dC(\mathbf{u})$$

and work with generalizations of the Koksma–Hlawka inequality that apply to measures other than the Lebesgue measure; see [62] or [2]. In the latter case, we work with the function Ψ and view the transformation ϕ_C as one that is applied to the low-discrepancy point set P_n rather than to Ψ . That is, here we work with the transformed point set $\tilde{P}_n = \{\phi_C(\mathbf{v}_1), \dots, \phi_C(\mathbf{v}_n)\}$ and analyze its quality via measures of discrepancy that quantify its distance to C rather than comparing P_n to $U[0, 1]^k$.

QMC methods have been used in a variety of applications, but so far most of the problems considered have been such that the stochastic models used can be formulated using independent random variables (e.g., a vector of dependent normal variates can be written as a linear transformation of independent normal variates). In such cases, the transformation of the uniform vector \mathbf{v} into observations from the desired stochastic model can be easily obtained by transforming each component v_j of \mathbf{v} using the inverse transform method, which is deemed to work well with QMC in part because of its monotonicity, and also because it corresponds to an overall one-to-one transformation from $[0, 1]^d$ to \mathbb{R}^d .

In the more general copula setting considered in this paper, at first sight the so-called conditional distribution method (CDM) (which is the inverse of the copula-based version of the Rosenblatt transform) appears to be a good choice to use with quasi-random numbers, as it is the direct multivariate extension of the inverse transform mentioned in the previous paragraph. Namely, it is a one-to-one transformation that maps $[0, 1]^d$ to $[0, 1]^d$ and it is monotone in each variable.

A transformation with $k = d$ is certainly desirable (and preferable to a many-to-one transformation with $k > d$) when used in conjunction with QMC methods, since these methods do better on problems of lower dimension. Also, intuitively the monotonicity should be helpful to preserve the smoothness of Ψ (for the first approach) and the low-discrepancy of P_n (for the second approach).

An additional advantage of the CDM is that it is applicable to any copula C (and the only known algorithm such general). However, the involved (inverses of the) conditional copulas are often challenging to evaluate which has led to other sampling algorithms being more frequently used. An example is the Marshall–Olkin algorithm for sampling Archimedean copulas, which we also address in this work.

The paper is organized as follows. Section 2 provides a short introduction to quasi-random numbers. Section 3 shows how quasi-random samples from various copulas (and thus multivariate models with these dependence structures) can be obtained using different sampling algorithms. Detailed examples are given. In Section 4 we discuss the theoretical background supporting each of the two approaches mentioned earlier to analyze the use of low-discrepancy sequences for copula sampling. Numerical results are provided in Section 5. Finally, Section 6 includes concluding remarks and a discussion of future work. Note that most results and figures presented in this paper (as well as additional experiments conducted) can be found in the R packages `copula` (see the vignette `qrng`) and `qrng` (see `demo(basket_options)` and `demo(test_functions)`).

2 Quasi-random numbers

Here we assume that a random sample $\{\mathbf{U}_i : i = 1, \dots, n\}$ from a copula C can be generated by transforming a random sample $\{\mathbf{U}'_i : i = 1, \dots, n\}$ from $U[0, 1]^k$ with $k \geq d$; several algorithms for copula models fall under this setup. Due to the independence of the vectors \mathbf{U}'_i , realizations of the sample $\{\mathbf{U}'_i : i = 1, \dots, n\}$ (obtained by so-called pseudo-random number generators (PRNGs)) will inevitably show regions of $[0, 1]^k$ which are lacking points and other areas of $[0, 1]^k$ which contain more samples than expected by the uniform distribution. To reduce this problem of an inhomogeneous concentration of samples, quasi-random number generators (QRNGs) do not aim at mimicking i.i.d. samples but instead at producing a homogeneous coverage of $[0, 1]^k$.

The homogeneity of a sequence of points over the unit hypercube can be measured by its discrepancy, which relates to the error incurred by representing the (Lebesgue-)measure of subsets of the unit hypercube by the fraction of points in these subsets. Quasi-random sequences aim at achieving smaller discrepancy than pseudo-random number sequences and are thus also called *low-discrepancy sequences*.

2.1 Discrepancy

The notion of discrepancy applies to sequences of points $X = \{\mathbf{v}_1, \mathbf{v}_2, \dots\}$ in the unit hypercube $[0, 1]^k$. Denote by $P_n = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subseteq [0, 1]^k$ the first n points of the sequence. Let \mathcal{J}^* be the set of intervals of $[0, 1]^k$ of the form $[\mathbf{0}, \mathbf{z}] = \prod_{j=1}^k [0, z_j)$, where $0 < z_j \leq 1, j = 1, \dots, k$. Then the *discrepancy function* of P_n on an interval $[\mathbf{0}, \mathbf{z}]$ is the difference

$$E([\mathbf{0}, \mathbf{z}]; P_n) = \frac{A([\mathbf{0}, \mathbf{z}]; P_n)}{n} - \lambda([\mathbf{0}, \mathbf{z}]),$$

where $A([\mathbf{0}, \mathbf{z}]; P_n) = \#\{i; 1 \leq i \leq n, \mathbf{v}_i \in [\mathbf{0}, \mathbf{z}]\}$ is the number of points from P_n that fall in $[\mathbf{0}, \mathbf{z})$ and $\lambda([\mathbf{0}, \mathbf{z}]) = \prod_{j=1}^k z_j$ is the Lebesgue measure of $[\mathbf{0}, \mathbf{z})$.

The *star discrepancy* D^* of P_n is defined by

$$D^*(P_n) = \sup_{[\mathbf{0}, \mathbf{z}] \in \mathcal{J}^*} |E([\mathbf{0}, \mathbf{z}]; P_n)|.$$

An infinite sequence X satisfying $D^*(P_n) \in O(n^{-1} \log^k n)$ is said to be a *low-discrepancy sequence*.

For a function $\Psi : [0, 1]^k \rightarrow \mathbb{R}$, we have the well-known Koksma–Hlawka error bound given by

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{v}_i) - \mathbb{E}[\Psi(\mathbf{U}')] \right| \leq V(\Psi) D^*(P_n), \quad (4)$$

where $\mathbf{U}' \sim U[0, 1]^k$ and $V(\Psi)$ denotes the variation of the function Ψ in the sense of Hardy and Krause. See [110] for a detailed account of the notion of variation and its applicability in practice. We also refer the interested reader to [60] and [119] for results handling unbounded functions (and thus of unbounded variation).

2.2 Low-discrepancy sequences

There are two main approaches for constructing low-discrepancy sequences: integration lattices and digital sequences. Only the latter are used in this paper, so our discussion will focus on those.

Digital sequences contain the well-known constructions of [118], [48], and [100], and are also closely related to the sequence proposed in [58]. The basic building block for this construction is the *van der Corput sequence in base $b \geq 2$* , defined as

$$S_b(i) = \sum_{r=0}^{\infty} a_r(i) b^{-r-1}, \quad i \in \mathbb{N}, \quad (5)$$

where $a_r(i)$ is the r th digit of the b -adic expansion of $i - 1 = \sum_{r=0}^{\infty} a_r(i) b^r$. To construct a sequence of points in $[0, 1)^k$ from this one-dimensional sequence, one approach is the one proposed by [58], which consists of choosing k pairwise relatively prime integers b_1, \dots, b_k and defining the i th point of the sequence as

$$\mathbf{v}_i = (S_{b_1}(i), \dots, S_{b_k}(i)), \quad i \in \mathbb{N}.$$

Another possibility is to fix the base b , and choose k linear transformations that are then applied to the digits $a_r(i)$ from the expansion of $i - 1$ before being used in (5) to define a real number between 0 and 1. More precisely, let M_1, \dots, M_k be (unbounded) “ $\infty \times \infty$ ” matrices with entries in \mathbb{Z}_b and let

$$S_b^{M_j}(i) = \sum_{r=0}^{\infty} \sum_{l=0}^{\infty} m_{r+1, l+1} a_l(i) b^{-r-1}, \quad (6)$$

where $m_{r,l}$ is the element in the r th row and l th column of M_j . Here we assume for simplicity that b is prime and all operations in (6) are performed in the finite field \mathbb{F}_b . One can then define a sequence of points in $[0, 1)^k$ by taking

$$\mathbf{v}_i = (S_b^{M_1}(i), \dots, S_b^{M_k}(i)) \quad (7)$$

as its i th point. Sobol’ was the first to propose such a construction, working in base 2 and defining the matrices M_1, \dots, M_k so that he was able to prove that the obtained sequence has $D^*(P_n) \in O(n^{-1} \log^k n)$; see [118].

We also point out that Halton sequences can be generalized using the same idea as in (7). That is, one can choose matrices M_1, \dots, M_k with the elements of M_j in \mathbb{Z}_{b_j} , and “scramble” the digits of the expansion of $i - 1$ before reverting them via (5) to produce a number between 0 and 1. A very simple way to achieve this is via diagonal matrices M_j , each containing a well-chosen element (or factor) of \mathbb{Z}_{b_j} . In our numerical experiments, we use such an approach, with the factors provided in [49]; see the R package `qrng` for an implementation.

2.3 Randomized quasi-Monte Carlo

In contrast to the error rate $O(n^{-1/2})$ for Monte Carlo methods based on PRNGs, approximations based on QRNGs have the advantage of having an error in $O(n^{-1} \log^k n)$ when the function of interest Ψ is of bounded variation. However, in practice it is also important to be able to estimate the corresponding error. While bounds such as the Koksma–Hlawka inequality are useful to understand the behaviour of approximations based on quasi-random sequences, they do not provide a practical way to estimate the error. To circumvent this problem, an approach that is often used is to *randomize* a low-discrepancy point set in such a way that its high uniformity (or low discrepancy)

is preserved, but at the same time unbiased estimators can be constructed (and sampled) from it. Another advantage of this approach is that variance expressions can be derived and compared with Monte Carlo sampling for wider classes of functions, i.e., not necessarily of bounded variation (see [107], [82] and the references therein). This approach gives rise to *randomized quasi-Monte Carlo (RQMC)* methods.

To apply this approach, we need a randomization function $r : [0, 1]^s \times [0, 1]^k \rightarrow [0, 1]^k$ with $s \geq k$ such that for any fixed $\mathbf{v} \in [0, 1]^k$, we have that if $\mathbf{U}' \sim \text{U}[0, 1]^s$, then $r(\mathbf{U}', \mathbf{v}) \sim \text{U}[0, 1]^k$. Hence the individual RQMC samples have the same properties as those from a random sample; the difference lies in the fact that the RQMC samples are dependent.

An early randomization scheme originally proposed by [30] is to take

$$r(\mathbf{U}', \mathbf{v}) = (\mathbf{v} + \mathbf{U}') \bmod 1. \quad (8)$$

A randomized point set is then obtained by generating a uniform vector \mathbf{U}' and letting $\tilde{P}_n(\mathbf{U}') = \{\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_n\}$, where $\tilde{\mathbf{v}}_i = r(\mathbf{U}', \mathbf{v}_i)$, $i \in \{1, \dots, n\}$.

Hence the same shift \mathbf{U}' is applied to all points in P_n . If we let $\mathbf{U}'_1, \dots, \mathbf{U}'_B$ be independent $\text{U}[0, 1]^s$ vectors, we can construct B i.i.d. unbiased estimators

$$\hat{\mu}_n^l = \frac{1}{n} \sum_{\tilde{\mathbf{v}}_i \in \tilde{P}_n(\mathbf{U}'_l)} \Psi(\phi_C(\tilde{\mathbf{v}}_i)), \quad l \in \{1, \dots, B\}$$

for $\mathbb{E}[\Psi(\mathbf{U})]$, whose variances can be estimated via the sample variance of $\hat{\mu}_n^1, \dots, \hat{\mu}_n^B$.

In addition to the simple random shift described in (8), several other randomization schemes have been proposed and studied. A popular randomization method for digital nets is to “scramble” them, an idea originally proposed by [106] and subsequently studied by [107], [108], [109], [88] and [67], among others.

A simpler randomization for digital nets is to use the digital counterpart of (8), where instead of adding two real numbers modulo 1, we add (in \mathbb{Z}_b) the digits of their base b expansion. That is, for $u = \sum_{r=0}^{\infty} u_r b^{-r-1}$ and $v = \sum_{r=0}^{\infty} v_r b^{-r-1}$, we let

$$u \oplus_b v = \sum_{r=0}^{\infty} ((u_r + v_r) \bmod b) b^{-r-1}$$

and define $r(\mathbf{u}, \mathbf{v}) = \mathbf{u} \oplus_b \mathbf{v}$, where the \oplus_b operation is applied component-wise to the k coordinates of \mathbf{u} and \mathbf{v} . The same idea can be applied to randomize Halton sequences (as shown, e.g., in [82]), but where a different base b is used in each of the k coordinates. Digital shifts for the Sobol’ and generalized Halton sequences are available in our R package `qrng`.

3 Quasi-random copula samples

Sampling procedures for a d -dimensional copula C can be viewed as transformations $\phi_C : [0, 1]^k \rightarrow [0, 1]^d$ for some $k \geq d$, such that, for $\mathbf{U}' \sim \text{U}[0, 1]^k$, $\mathbf{U} := \phi_C(\mathbf{U}') \sim C$; that is, ϕ_C transforms independent $\text{U}[0, 1]$ random variables to d dependent random variables with distribution function C .

The case $k = d$ is mostly known and applied as *conditional distribution method (CDM)* and involves the inversion method for sampling univariate conditional copulas (although, for example,

for Archimedean copulas another transformations ϕ_C with $k = d$ is known; see [125]). This approach thus naturally uses d independent $U[0, 1]$ random variables as input. The case $k \geq d$ (often: $k > d$) is typically known as *stochastic representation* and is usually based on sampling k univariate random variables from elementary probability distributions, as we will see in Section 3.2.

In what follows we consider the above two approaches and show how they can be adapted to quasi-random number generation.

3.1 Conditional distribution method and other one-to-one transformations ($k = d$)

3.1.1 Theoretical background

The only known general sampling approach which works for any copula is the CDM. For $j \in \{2, \dots, d\}$, let

$$C(u_j | u_1, \dots, u_{j-1}) = \mathbb{P}(U_j \leq u_j | U_1 = u_1, \dots, U_{j-1} = u_{j-1})$$

denote the *conditional copula of U_j at u_j given $U_1 = u_1, \dots,$*

$U_{j-1} = u_{j-1}$. If $C^-(u_j | u_1, \dots, u_{j-1})$ denotes the corresponding quantile function, the CDM is given as follows; see [39] or [63, p. 45].

Theorem 3.1 (Conditional distribution method). *Let C be a d -dimensional copula, $\mathbf{U}' \sim U[0, 1]^d$, and ϕ_C^{CDM} be given by*

$$\begin{aligned} U_1 &= U'_1, \\ U_2 &= C^-(U'_2 | U_1), \\ &\vdots \\ U_d &= C^-(U'_d | U_1, \dots, U_{d-1}). \end{aligned}$$

Then $\mathbf{U} = (U_1, \dots, U_d) = \phi_C^{CDM}(\mathbf{U}') \sim C$.

To find the conditional copulas $C(u_j | u_1, \dots, u_{j-1})$, for $j \in \{2, \dots, d\}$, for a specific copula C , the following result (which holds under mild assumptions) is often applied. A rigorous proof can be found in [117, p. 20], an implementation is provided by the function `rtrafo()` in the R package `copula`. The corollary that follows is an immediate consequence of Sklar's theorem, for example.

Theorem 3.2 (Computing conditional copulas). *Let C be a d -dimensional copula, which, for $d \geq 3$, admits continuous partial derivatives with respect to the first $d - 1$ arguments. For $j \in \{2, \dots, d\}$ and $u_l \in [0, 1]$, $l \in \{1, \dots, j\}$,*

$$\begin{aligned} C(u_j | u_1, \dots, u_{j-1}) &= \frac{D_{j-1 \dots 1} C(u_1, \dots, u_j)}{D_{j-1 \dots 1} C(u_1, \dots, u_{j-1})} \\ &= \frac{D_{j-1 \dots 1} C(u_1, \dots, u_j)}{c(u_1, \dots, u_{j-1})}, \end{aligned} \tag{9}$$

where $D_{j-1 \dots 1}$ denotes the derivative with respect to the first $j - 1$ arguments, $C(u_1, \dots, u_j)$ denotes the marginal copula corresponding to the first j components and $c(u_1, \dots, u_{j-1})$ denotes the density of $C(u_1, \dots, u_{j-1})$. If C admits a density, then (9) equals

$$C(u_j | u_1, \dots, u_{j-1}) = \frac{\int_0^{u_j} c(u_1, \dots, u_{j-1}, z_j) dz_j}{c(u_1, \dots, u_{j-1})}. \tag{10}$$

Corollary 3.3 (Conditional copulas for general multivariate distributions). *Let H be a d -dimensional absolutely continuous distribution function with margins F_1, \dots, F_d and copula C . For $j \in \{2, \dots, d\}$ and $u_l \in [0, 1]$, $l \in \{1, \dots, j\}$,*

$$C(u_j | u_1, \dots, u_{j-1}) = H(F_j^-(u_j) | F_1^-(u_1), \dots, F_{j-1}^-(u_{j-1})). \quad (11)$$

3.1.2 Examples

We now present several copula families and show how the corresponding conditional copulas and their inverses can be computed. To the best of our knowledge, several of these results have not appeared in the literature before.

Elliptical copulas

An *elliptical copula* describes the dependence structure of an elliptical distribution; for the latter, see [19], [40], [39], or [89, Sections 3.3, 5]. The most prominent two families in the class of elliptical copulas are the Gauss and the t copulas.

Gauss copulas. Gauss copulas are given by

$$C_P^{\text{Ga}}(\mathbf{u}) = \Phi_P(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)),$$

where Φ_P denotes the d -variate normal distribution function with location vector $\mathbf{0}$ and scale matrix P (a correlation matrix) and Φ^{-1} is the standard normal quantile function. Consider the dimension to be j and let $\mathbf{X} \sim \Phi_P$ with $\mathbf{X} = (\mathbf{X}_{1:(j-1)}, X_j)$. Furthermore, assume

$$P = \begin{pmatrix} P_{1:(j-1),1:(j-1)} & P_{1:(j-1),j} \\ P_{j,1:(j-1)} & P_{j,j} \end{pmatrix}$$

to be positive definite. It follows from [47, p. 45 and 78] that

$$X_j | \mathbf{X}_{1:(j-1)} = \mathbf{x}_{1:(j-1)} \sim \mathcal{N}(\mu_{j|1:(j-1)}(\mathbf{x}_{1:(j-1)}), P_{j|1:(j-1)}),$$

where

$$\begin{aligned} \mu_{j|1:(j-1)}(\mathbf{x}_{1:(j-1)}) &= P_{j,1:(j-1)} (P_{1:(j-1),1:(j-1)})^{-1} \mathbf{x}_{1:(j-1)}, \\ P_{j|1:(j-1)} &= P_{j,j} - P_{j,1:(j-1)} (P_{1:(j-1),1:(j-1)})^{-1} P_{1:(j-1),j}; \end{aligned} \quad (12)$$

so $H(x_j | x_1, \dots, x_{j-1})$ is again normal. With $\Phi^{-1}(\mathbf{u}_{1:(j-1)}) = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_{j-1}))$, it follows from (11) that

$$\begin{aligned} C(u_j | u_1, \dots, u_{j-1}) &= H(\Phi^{-1}(u_j) | \Phi^{-1}(\mathbf{u}_{1:(j-1)})) \\ &= \Phi_{\mu_{j|1:(j-1)}(\Phi^{-1}(\mathbf{u}_{1:(j-1)})), P_{j|1:(j-1)}}(\Phi^{-1}(u_j)) \\ &= \Phi\left(\frac{\Phi^{-1}(u_j) - \mu_{j|1:(j-1)}(\Phi^{-1}(\mathbf{u}_{1:(j-1)}))}{\sqrt{P_{j|1:(j-1)}}}\right) \end{aligned}$$

and thus that

$$\begin{aligned}
& C^-(u_j | u_1, \dots, u_{j-1}) \\
&= \Phi\left(\Phi_{\mu_{j|1:(j-1)}^{-1}(\Phi^{-1}(\mathbf{u}_{1:(j-1)})), P_{j|1:(j-1)}}^{-1}(u_j)\right) \\
&= \Phi\left(\mu_{j|1:(j-1)}(\Phi^{-1}(\mathbf{u}_{1:(j-1)})) + \sqrt{P_{j|1:(j-1)}}\Phi^{-1}(u_j)\right).
\end{aligned}$$

An implementation of this inverse is provided via `rtrafo(, inverse=TRUE)` in the R package `copula`.

***t* copulas.** *t* copulas are given by

$$C_{\nu, P}^t(\mathbf{u}) = t_{\nu, P}(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_d)),$$

where $t_{\nu, P}$ denotes the d -variate t_{ν} distribution function with location vector $\mathbf{0}$ and scale matrix P (a correlation matrix) and t_{ν}^{-1} is the standard t_{ν} quantile function. The following proposition guarantees stability of the *t* copula upon conditioning; see the appendix for its proof and `rtrafo()` for an implementation. We are not aware that this result has appeared before. Given the importance of *t* copulas in practice, this is rather remarkable.

Proposition 3.4 (Conditional *t* copulas and inverses). *With the notation as in the Gauss case, the conditional t copula at u_j , given u_1, \dots, u_{j-1} , and its inverse are given by*

$$\begin{aligned}
C(u_j | u_1, \dots, u_{j-1}) &= t_{\nu+j-1}\left(s_1\left(\sqrt{P_{j,j}^{-1}}t_{\nu}^{-1}(u_j) + s_2\right)\right), \\
C^-(u_j | u_1, \dots, u_{j-1}) &= t_{\nu}\left(\frac{t_{\nu+j-1}^{-1}(u_j)/s_1 - s_2}{\sqrt{P_{j,j}^{-1}}}\right),
\end{aligned}$$

for s_1, s_2 as given in the proof.

Figure 2 displays 1000 samples from a *t* copula with three degrees of freedom and correlation parameter $\rho = P_{1,2} = 1/\sqrt{2}$ (Kendall's tau equals 0.5), once drawn with a PRNG (top) and once with a QRNG (bottom). We can visually confirm in this case that the low discrepancy of the latter is preserved. How this seemingly good feature translates into better estimators of the form (2) will be studied further through the theoretical results of Section 4 and the numerical experiments of Section 5.

Archimedean copulas

An (*Archimedean*) generator is a continuous, decreasing function $\psi : [0, \infty] \rightarrow [0, 1]$ which satisfies $\psi(0) = 1$, $\psi(\infty) = \lim_{t \rightarrow \infty} \psi(t) = 0$, and which is strictly decreasing on $[0, \inf\{t : \psi(t) = 0\}]$. A d -dimensional copula C is called *Archimedean* if it permits the representation

$$C(\mathbf{u}) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d)),$$

where $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$, and for some generator ψ with inverse $\psi^{-1} : [0, 1] \rightarrow [0, \infty]$, where $\psi^{-1}(0) = \inf\{t : \psi(t) = 0\}$. For applications and the importance of Archimedean copulas in the realm of finance and insurance, see, e.g., [65].

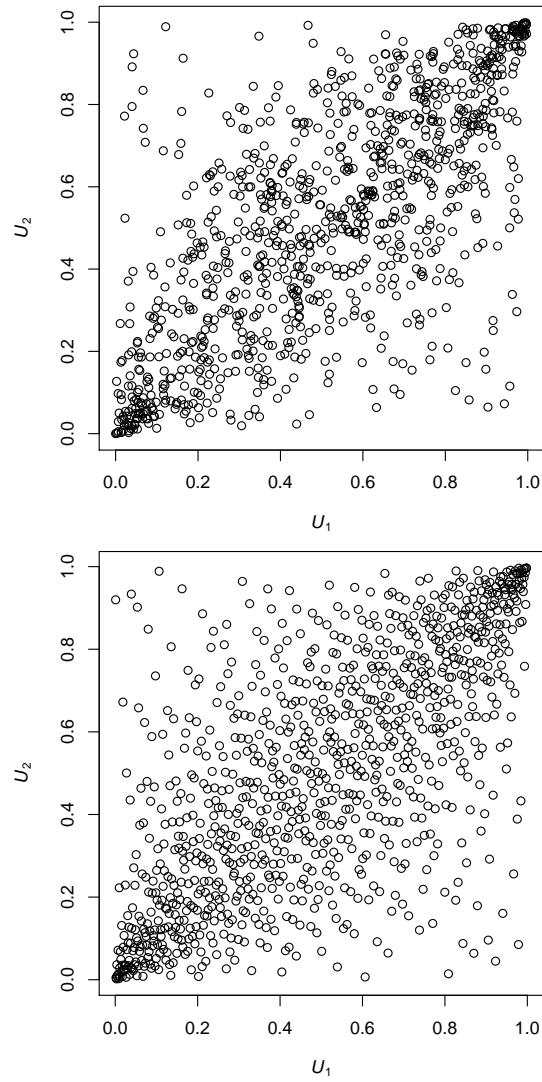


Figure 2: 1000 realizations of a t copula with three degrees of freedom and correlation parameter $\rho = 1/\sqrt{2}$ (Kendall's tau equals 0.5), generated by a PRNG (top) and by a QRNG (bottom).

[90] show that a generator defines an Archimedean copula if and only if ψ is *d-monotone*, meaning that ψ is continuous on $[0, \infty]$, admits derivatives $\psi^{(l)}$ up to the order $l = d - 2$ satisfying $(-1)^l \psi^{(l)}(t) \geq 0$ for all $l \in \{0, \dots, d - 2\}$, $t \in (0, \infty)$, and $(-1)^{d-2} \psi^{(d-2)}(t)$ is decreasing and convex on $(0, \infty)$.

Assuming ψ to be sufficiently often differentiable, conditional Archimedean copulas follow from Theorem 3.2 and are given by

$$C(u_j | u_1, \dots, u_{j-1}) = \frac{\psi^{(j-1)}(\sum_{l=1}^j \psi^{-1}(u_l))}{\psi^{(j-1)}(\sum_{l=1}^{j-1} \psi^{-1}(u_l))}, \quad (13)$$

where $u_l \in [0, 1]$, $l \in \{1, \dots, j\}$, and thus

$$\begin{aligned} C^-(u_j | u_1, \dots, u_{j-1}) \\ = \psi \left(\psi^{(j-1)-1} \left(u_j \psi^{(j-1)} \left(\sum_{l=1}^{j-1} \psi^{-1}(u_l) \right) \right) - \sum_{l=1}^{j-1} \psi^{-1}(u_l) \right). \end{aligned} \quad (14)$$

The generator derivatives $\psi^{(j-1)}$ and their inverses $\psi^{(j-1)-1}$ can be challenging to compute. The former are known explicitly for several Archimedean families and certain generator transformations; see [64] for more details. To compute the inverses, one can use numerical root-finding on $[0, 1]$; see `rtrafo(..., inverse=TRUE)` in the R package `copula`. This can be applied, e.g., in the case of Gumbel copulas.

The following example shows the case of a Clayton copula family, for which (14) can be given explicitly and thus where the CDM is tractable; this explicit formula is also utilized by `rtrafo(..., inverse=TRUE)`.

Example 3.5. [Clayton copulas] If $\psi(t) = (1 + t)^{-1/\theta}$, $t \geq 0$, $\theta > 0$, denotes a generator of the Archimedean Clayton copula, then $\psi^{(j)}(t) = (-1)^j (1 + t)^{-(j+1/\theta)} \prod_{l=0}^{j-1} (l + 1/\theta)$. Therefore, (13) equals

$$C(u_j | u_1, \dots, u_{j-1}) = \left(\frac{1 - j + \sum_{l=1}^j u_l^{-\theta}}{2 - j + \sum_{l=1}^{j-1} u_l^{-\theta}} \right)^{j-1/\theta}$$

and (14) equals

$$\begin{aligned} C^-(u_j | u_1, \dots, u_{j-1}) = \\ \left(1 + \left(1 - (j - 1) + \sum_{l=1}^{j-1} u_l^{-\theta} \right) \left(u_j^{-\frac{1}{j-1+1/\theta}} - 1 \right) \right)^{-\frac{1}{\theta}}. \end{aligned}$$

Figure 1 displays 1000 samples from a Clayton copula with $\theta = 2$ (Kendall's tau equals 0.5), once drawn with a PRNG (top) and once with a QRNG (bottom).

Marshall–Olkin copulas

A class of bivariate copulas for which $C^-(u_2 | u_1)$ is explicit is the class of Marshall–Olkin copulas $C(u_1, u_2) = \min\{u_1^{1-\alpha_1}u_2, u_1u_2^{1-\alpha_2}\}$, $\alpha_1, \alpha_2 \in (0, 1)$, where one can show that

$$C^-(u_2 | u_1) = \begin{cases} \frac{u_1^{\alpha_1}}{1-\alpha_1}u_2, & \text{if } u_2 \in [0, (1-\alpha_1)u_1^{\alpha_1(1/\alpha_2-1)}], \\ u_1^{\alpha_1/\alpha_2}, & \text{if } u_2 \in ((1-\alpha_1)u_1^{\alpha_1(1/\alpha_2-1)}, u_1^{\alpha_1(1/\alpha_2-1)}), \\ u_2^{\frac{1}{1-\alpha_2}}, & \text{if } u_2 \in [u_1^{\alpha_1(1/\alpha_2-1)}, 1]. \end{cases}$$

Figure 3 shows 1000 samples, once drawn from a PRNG (top) and once from a QRNG (bottom). Here again we can visually confirm the low discrepancy.

Another class of copulas not discussed in this section which is naturally sampled with the CDM and thus can easily be adapted to construct corresponding quasi-random numbers is the class of pair copula constructions; see, e.g., [80]. For this purpose, we modified the function `RVineSim()` in the R package `VineCopula` (version ≥ 1.3). It now allows to pass a matrix of quasi-random numbers to be transformed to the corresponding samples from a pair copula construction; see the vignette `qrng` in the R package `copula` for examples. Note that if sampling of the R-vine involves numerical root-finding (required for certain copula families), the corresponding numerical inaccuracy may have an effect on the low discrepancy of the generated samples.

3.2 Stochastic representations ($k \geq d$, typically $k > d$)

3.2.1 Theoretical background

As mentioned above, pair-copula constructions are one of the rare copula classes for which the CDM is applied in practice. For most other copula classes and families, faster sampling algorithms derived from stochastic representations of $\mathbf{U} \sim C$ are known, especially for $d \gg 2$. They are mostly class- and family-specific, as can be seen in the examples below.

3.2.2 Examples

Elliptical copulas

Gauss and t copulas are typically sampled via their stochastic representations.

Gauss copulas. A random vector $\mathbf{X} \sim \Phi_P$ admits the stochastic representation $\mathbf{X} = \mathbf{AZ}$ where A denotes the lower triangular matrix from the Cholesky decomposition $P = AA^\top$ and \mathbf{Z} is a vector of d independent standard normal random variables. A random vector $\mathbf{U} \sim C_P^{\text{Ga}}$ thus admits the stochastic representation $\Phi(\mathbf{X}) = \Phi(\mathbf{AZ})$ for $\mathbf{Z} = (\Phi^{-1}(U'_1), \dots, \Phi^{-1}(U'_d))$ and $\mathbf{U}' \sim U[0, 1]^d$; here Φ is assumed to act on \mathbf{AZ} component-wise. Note that for Gauss copulas, this sampling approach is equivalent to the CDM.

t copulas. A random vector $\mathbf{X} \sim t_{\nu, P}$ admits the stochastic representation $\mathbf{X} = \sqrt{W}\mathbf{AZ}$ where A and \mathbf{Z} are as above and $W = 1/\Gamma$ for Γ following a Gamma distribution with shape and rate equal to $\nu/2$. A random vector $\mathbf{U} \sim C_{\nu, P}^t$ thus admits the stochastic representation $t_\nu(\mathbf{X}) = t_\nu(\sqrt{W}\mathbf{AZ})$;

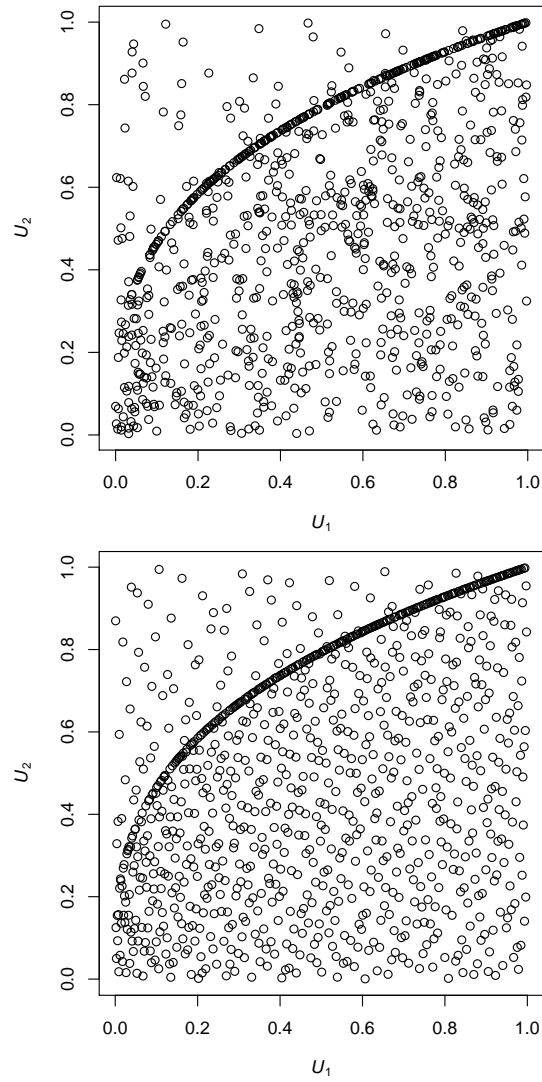


Figure 3: 1000 realizations of a Marshall–Olkin copula with $\alpha_1 = 0.25$ and $\alpha_2 = 0.75$ (Kendall’s tau equals roughly 0.23), generated by a PRNG (top) and by a QRNG (bottom).

as before, t_ν is assumed to act on $\sqrt{W}AZ$ component-wise. Note that for t copulas with finite ν , this sampling approach is different from the CDM.

Archimedean copulas

The conditional independence approach behind the Marshall–Olkin algorithm for sampling Archimedean copulas is one example for transformations ϕ_C for $k > d$; see [86]. For this algorithm, $k = d + 1$ and one uses the fact that for an Archimedean copula C with completely monotone generator ψ ,

$$\mathbf{U} = (\psi(E_1/V), \dots, \psi(E_d/V)) \sim C, \quad (15)$$

where $V \sim F = \mathcal{L}\mathcal{S}^{-1}[\psi]$, independent of $E_1, \dots, E_d \sim \text{Exp}(1)$; here, $F = \mathcal{L}\mathcal{S}^{-1}[\psi]$ denotes the distribution function corresponding to ψ by Bernstein’s Theorem ($\mathcal{L}\mathcal{S}^{-1}[\cdot]$ denotes the inverse Laplace–Stieltjes transform). To give an explicit expression for the transformation $\phi_C = \phi_C^{\text{MO}}$ in this case, we assume that v_1 is used to generate V via the inversion method, and v_{j+1} is used to generate E_j , for $j \in \{1, \dots, d\}$. Then we have that $\phi_C^{\text{MO}} = (\phi_{C,1}^{\text{MO}}, \dots, \phi_{C,d}^{\text{MO}})$, where

$$\phi_{C,j}^{\text{MO}} = \phi_{C,j}^{\text{MO}}(v_1, v_{j+1}) = \psi\left(\frac{-\log v_{j+1}}{F^{-1}(v_1)}\right), \quad j \in \{1, \dots, d\}. \quad (16)$$

We can use a low-discrepancy sequence in $k = d + 1$ dimensions to produce a sample based on this method. Having $k = d + 1$ instead of $k = d$ is a slight disadvantage, since it is well known that the performance of QMC methods tends to deteriorate with increasing dimensions.

To explore the effect of the transformation ϕ_C on P_n , we generated 1000 realizations of a three-dimensional Halton sequence; see the top of Figure 4 where we colored points falling in two non-overlapping regions in $[0, 1]^2$. The first two of the three dimensions are then mapped via ϕ_C^{CDM} (see the bottom of Figure 4) to a Clayton copula with parameter $\theta = 2$ (such that Kendall’s tau equals 0.5). As we can see, the non-overlapping colored regions remain non-overlapping after the one-to-one transformations have been applied. To study the effect of the Marshall–Olkin algorithm, we look at when the first dimension of the Halton sequence is mapped to a Gamma $\Gamma(1/\theta, 1)$ distribution by inversion of v_1 (the distribution of V in (15) for a Clayton copula) and the last two to unit exponential distributions (by inversion of $1 - v_j$ for $j = 2, 3$, so that the obtained u_j is increasing in each of v_1 and v_{j+1} for $j = 1, 2$). The top of Figure 5 shows the second and third coordinates of the Halton sequence, and colors the points belonging to two different three-dimensional intervals (this is why not all two-dimensional points are coloured in the two-dimensional projected regions). We see on the bottom of Figure 5 that here again, the colored regions remain non-overlapping. However, it should also be clear that two points in a given interval defined over the second and third dimension could end up in very different locations after this transform, if the corresponding first coordinates are far apart. Hence, the fact that the Marshall–Olkin transform uses $k = d + 1$ uniforms (and thus is not one-to-one) makes it more challenging to understand its effect when used with quasi-random numbers. On the other hand, because it is designed so that the first uniform v_1 is very important, it may work quite well with QMC since these methods are known to perform better when a small number of variables are important (i.e., see [35, 82]). This combination (QRNG with the Marshall–Olkin approach) is studied further in Section 4, with numerical results provided in Section 5.

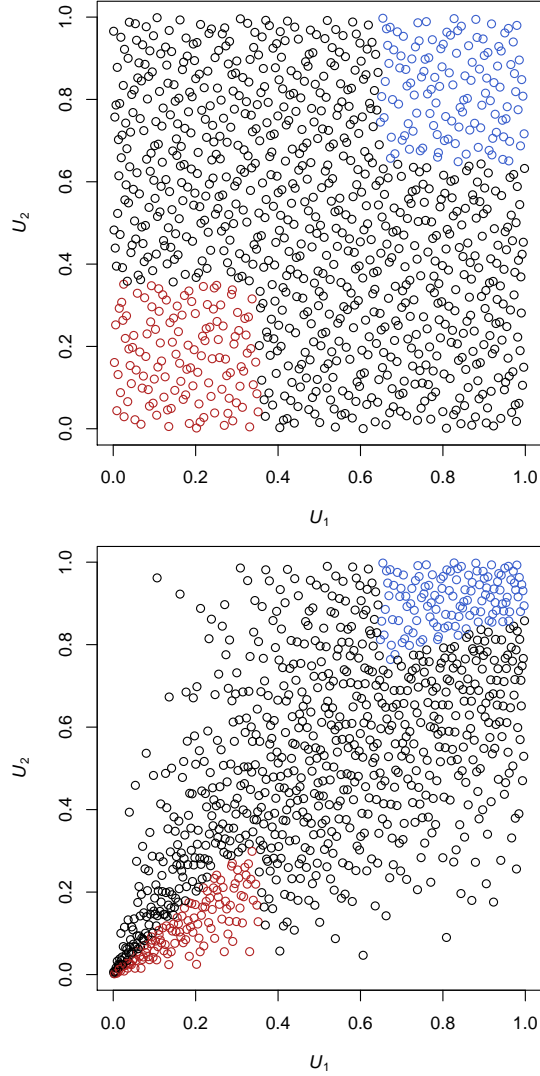


Figure 4: 1000 realizations of the first two components of a three-dimensional Halton sequence with colored points in the regions $[0, \sqrt{1/8}]^2$ and $[1 - \sqrt{1/8}, 1]^2$ (top): corresponding ϕ_C^{CDM} -transformed points (bottom) to a Clayton copula with $\theta = 2$ (Kendall's tau equals 0.5).

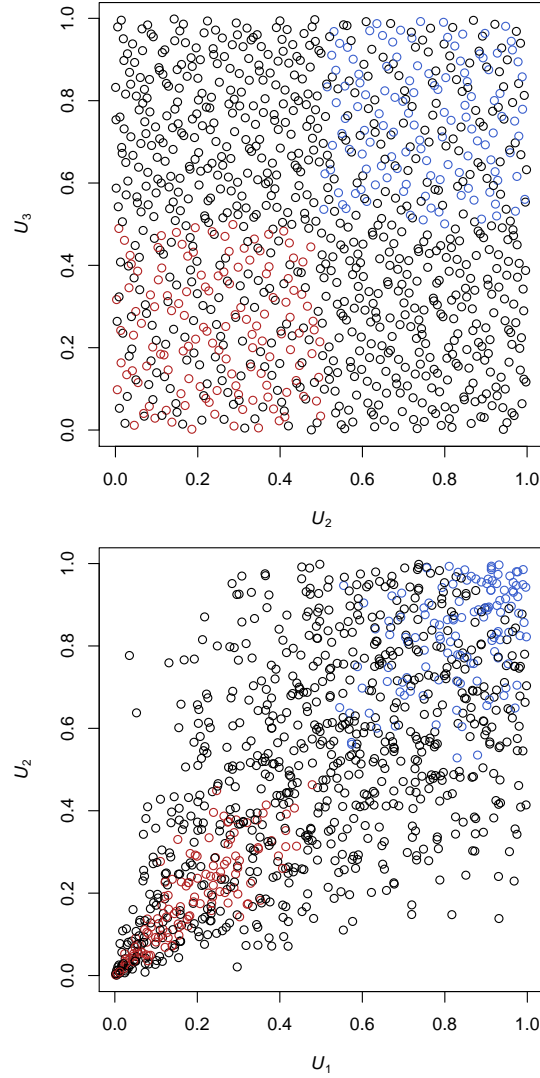


Figure 5: 1000 realizations of the second and third components of a three-dimensional Halton sequence with colored points corresponding to the regions $[0, 0.5]^3$ and $[0.5, 1]^3$ (top): corresponding ϕ_C^{MO} -transformed points (bottom) to a Clayton copula with $\theta = 2$ (Kendall's tau equals 0.5).

Marshall–Olkin copulas

Bivariate ($d = 2$) Marshall–Olkin copulas C also allow for a stochastic representation in our framework ϕ_C for $k > d$. For example, it is easy to check that for $(U'_1, U'_2, U'_3) \sim U[0, 1]^3$,

$$(\max\{U_1'^{\frac{1}{1-\alpha_1}}, U_3'^{\frac{1}{\alpha_1}}\}, \max\{U_2'^{\frac{1}{1-\alpha_2}}, U_3'^{\frac{1}{\alpha_2}}\}) \sim C.$$

This construction can be generalized to $d > 2$ (but we omit further details about Marshall–Olkin copulas in the remaining part of this paper).

3.3 Words of caution

The plots showing copula samples obtained from QRNGs that we have seen so far have been promising, in that the additional uniformity (or low discrepancy) compared to pseudo-sampling was visible. Here we want to add a word of caution to the effect that it is crucial to work with high quality quasi-random numbers, as defects that exist with respect to their uniformity on the unit cube will translate into poor copula samples. Figure 6 illustrates this by showing two-dimensional copula samples obtained from quasi-random numbers of poor quality, corresponding to the projection on coordinates (20,21) of the first 1000 points of the Halton sequence (top) and a similar sample obtained from a generalized Halton sequence (bottom), which was designed to address defects of this type in the Halton sequence. More precisely, here the problem is that this particular projection is based on the twin prime numbers 71 and 73 for the base. Defects of this type are discussed further, e.g., in [95].

4 Analyzing the performance of copula sampling with quasi-random numbers

In this section, we discuss the two approaches outlined in the introduction to analyze the validity of sampling algorithms for copulas that are based on low-discrepancy sequences.

4.1 Composing the sampling method with the function of interest Ψ

Our goal here is to assess the quality of a quasi-random sampling method for copula models by viewing the transformation ϕ_C as being composed with the function Ψ of interest, so that we can work in the usual Koksma–Hlawka setting based on uniform discrepancy measures.

Given that a copula transform $\phi_C = (\phi_{C,1}, \dots, \phi_{C,d})$ is regular enough, denote its Jacobian by

$$J_{\phi_C} = \frac{\partial(\phi_{C,1}, \dots, \phi_{C,d})}{\partial(u_1, \dots, u_d)},$$

and write

$$\begin{aligned} \mathbb{E}[\Psi(\mathbf{U})] &= \int_{[0,1]^d} \Psi(\mathbf{u})c(\mathbf{u})d\mathbf{u} \\ &= \int_{[0,1]^d} \Psi(\phi_C(\mathbf{v}))c(\phi_C(\mathbf{v}))|J_{\phi_C}(\phi_C(\mathbf{v}))|d\mathbf{v}. \end{aligned}$$

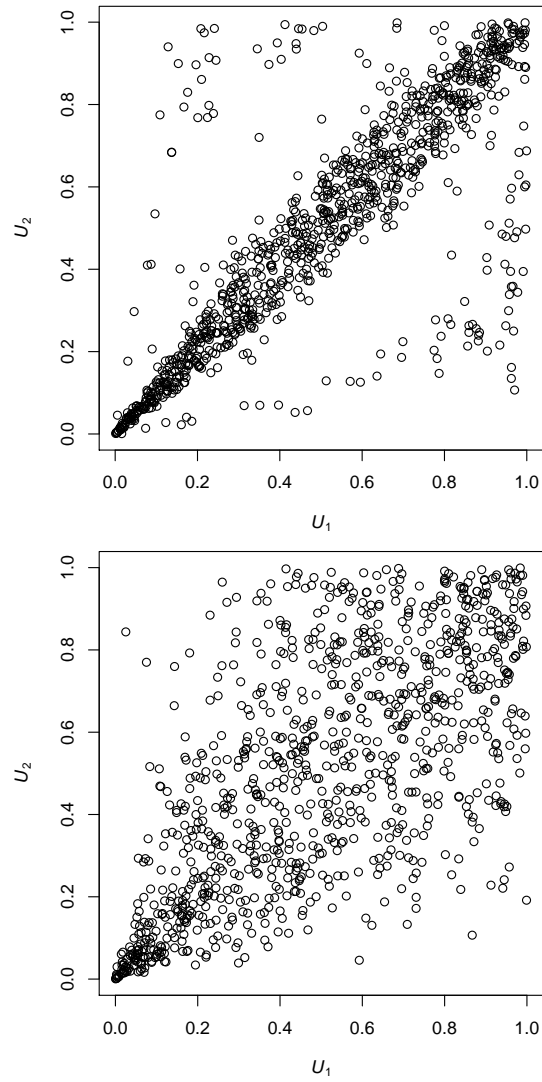


Figure 6: Samples obtained from Clayton copula with $\theta = 2$ with the CDM method based on coordinates 20 and 21 of the Halton sequence (top) and the generalized Halton sequence (bottom).

In the case of $\phi_C = \phi_C^{\text{CDM}}$, one can easily show that $|J_{\phi_C}(\phi_C(\mathbf{v}))| = c(\phi_C(\mathbf{v}))^{-1}$, and thus

$$\mathbb{E}[\Psi(\mathbf{U})] = \int_{[0,1]^d} \Psi(\phi_C(\mathbf{v}))d\mathbf{v}. \quad (17)$$

While the properties of the CDM approach allow one to directly show (17) in its integral form as done above, this equality holds more generally for any transformation $\phi_C : [0, 1]^k \rightarrow [0, 1]^d$ such that $\phi_C(\mathbf{U}) \sim C$ whenever $\mathbf{U} \sim U[0, 1]^k$; see also [18, 112].

In the case where (17) holds, one can apply the Koksma–Hlawka error bound (4) to transformed samples.

Proposition 4.1 (Koksma–Hlawka bound for a change of variables). *Let $\mathbf{U} \sim C$, ϕ_C such that (17) holds, and $\mathbf{u}_i = \phi_C(\mathbf{v}_i)$ for $P_n = \{\mathbf{v}_i, i = 1, \dots, n\}$ in $[0, 1]^d$. Then*

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq D^*(P_n)V(\Psi \circ \phi_C).$$

Note that $V(\Psi) < \infty$ does not imply $V(\Psi \circ \phi_C) < \infty$ in general. To get further insight into the conditions required to have a finite bound on the integration error, we work with a slight variation of the above bound that is given in [101, pp. 19–20] (see also [62, (4')] and [61, (4)]), where the term $V(\Psi \circ \phi_C)$ is replaced by an expression given in terms of the partial derivatives of $\Psi \circ \phi_C$ assuming the latter exist and are continuous. It is given by

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq D^*(P_n)\|\Psi \circ \phi_C\|_{d,1},$$

where

$$\begin{aligned} \|\Psi \circ \phi_C\|_{d,1} = & \\ & \sum_{l=1}^s \sum_{\alpha} \int_{[0,1]^l} \left| \frac{\partial^l \Psi \circ \phi_C(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial v_{\alpha_1} \dots \partial v_{\alpha_l}} \right| dv_{\alpha_1} \dots dv_{\alpha_l} \end{aligned}$$

and the second sum is taken over all nonempty subsets $\alpha = \{\alpha_1, \dots, \alpha_l\} \subseteq \{1, \dots, d\}$. Furthermore, the notation $\mathbf{1}$ in $\Psi \circ \phi_C(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})$ means that each variable v_j with $j \notin \{\alpha_1, \dots, \alpha_l\}$ is set to 1.

The following proposition provides sufficient conditions on the functional Ψ and on the copula C to ensure that $\|\Psi \circ \phi_C\|_{d,1} < \infty$ when $\phi_C = \phi_C^{\text{CDM}}$.

Proposition 4.2 (Conditions to have bounded variation with variable change in the CDM). *Assume that Ψ has continuous mixed partial derivatives up to total order d and there exist $m, M, K > 0$ such that for all $\mathbf{u} \in (0, 1)^d$, $c(\mathbf{u}) \geq m > 0$ and*

$$\left| \frac{\partial^k C(u_i | u_1, \dots, u_{i-1})}{\partial u_{\alpha_1} \dots \partial u_{\alpha_k}} \right| \leq M, \quad \alpha_1, \dots, \alpha_k \in \{1, \dots, i\}, \quad (18)$$

for each $1 \leq k \leq i \leq d$. Furthermore, assume that for all $1 \leq k \leq l \leq d$ and $\{\alpha_1, \dots, \alpha_l\} \subseteq \{1, \dots, d\}$, we have

$$\left| \frac{\partial^k \Psi(u_1, \dots, u_d)}{\partial u_{\beta_1} \dots \partial u_{\beta_k}} \right| \leq K, \quad \beta_j \in \{\alpha_1, \dots, \alpha_l\}, \quad 1 \leq j \leq k. \quad (19)$$

Then there exists a constant $C^{(d)}$ (independent of n but dependent on Ψ) such that for the choice $\phi_C = \phi_C^{CDM}$, we have

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq D^*(\mathbf{v}_1, \dots, \mathbf{v}_n) K C^{(d)} (M^d/m)^{2d-1},$$

where $\mathbf{u}_i = \phi_C^{CDM}(\mathbf{v}_i)$, $i = 1, \dots, n$.

Proof. See [62, (11) and the remark thereafter]. □

Remark 4.3. 1. As we will see in the discussion preceding the next proposition, in general, to ensure that $\|\Psi \circ \phi_C\|_{d,1} < \infty$ holds, a possible approach is to bound the mixed partial derivatives involving Ψ and then to verify that the mixed partial derivatives involving ϕ_C are integrable. As explained in [62], Condition (18) ensures that the latter condition is verified in the case of the CDM (or Rosenblatt) transform, and avoids having to deal with the function ϕ_C and its partial derivatives. Unfortunately (and while it may seem easier to work with the conditional copulas $C(u_j | u_1, \dots, u_{j-1})$ than with ϕ_C), in many cases the copulas involved do not have bounded mixed partial derivatives everywhere, with singularities appearing near the boundaries when one or more arguments are 0 or 1. A non-trivial case where we were able to verify (18) is for the Eyrraud-Farlie-Gumbel-Morgenstern copula (see [72]), assuming the parameters are chosen so that the density $c(\mathbf{u})$ and thus the denominator of $C(u_j | u_1, \dots, u_{j-1})$ is bounded away from 0 for all $\mathbf{u} \in [0, 1]^d$.

2. We note that the conditions given in (19) are not the same as those required to prove that

$$\|\Psi\|_{d,1} = \sum_{l=1}^d \sum_{\alpha} \int_{[0,1]^l} \left| \frac{\partial^l \Psi(u_{\alpha_1}, \dots, u_{\alpha_l}, \mathbf{1})}{\partial u_{\alpha_1} \dots \partial u_{\alpha_l}} \right| du_{\alpha_1} \dots du_{\alpha_l}$$

is bounded; in the latter case, we only need to consider mixed partial derivatives of order at most one in each variable (since the α_j 's are distinct). However, in (19), the β_j 's are not necessarily distinct. In particular, this means that we need to consider the partial derivative of Ψ of order d with respect to each variable and make sure it is bounded.

Remark 4.4. 1. As we will see in the discussion preceding the next proposition, in general, to ensure that $\|\Psi \circ \phi_C\|_{d,1} < \infty$ holds, a possible approach is to bound the mixed partial derivatives involving Ψ and then to verify that the mixed partial derivatives involving ϕ_C are integrable. As explained in [62], Condition (18) ensures that the latter condition is verified in the case of the CDM (or Rosenblatt) transform, and avoids having to deal with the function ϕ_C and its partial derivatives. Unfortunately (and while it may seem easier to work with the conditional copulas $C(u_j | u_1, \dots, u_{j-1})$ than with ϕ_C), in many cases the copulas involved do not have bounded mixed partial derivatives everywhere, with singularities appearing near the boundaries when one or more arguments are 0 or 1. A non-trivial case where we were able to verify (18) is for the Eyrraud-Farlie-Gumbel-Morgenstern copula (see [72]), assuming the parameters are chosen so that the density $c(\mathbf{u})$ and thus the denominator of $C(u_j | u_1, \dots, u_{j-1})$ is bounded away from 0 for all $\mathbf{u} \in [0, 1]^d$.

2. We note that the conditions given in (19) are not the same as those required to prove that

$$\|\Psi\|_{d,1} = \sum_{l=1}^d \sum_{\alpha} \int_{[0,1]^l} \left| \frac{\partial^l \Psi(u_{\alpha_1}, \dots, u_{\alpha_l}, \mathbf{1})}{\partial u_{\alpha_1} \dots \partial u_{\alpha_l}} \right| du_{\alpha_1} \dots du_{\alpha_l}$$

is bounded; in the latter case, we only need to consider mixed partial derivatives of order at most one in each variable (since the α_j 's are distinct). However, in (19), the β_j 's are not necessarily distinct. In particular, this means that we need to consider the partial derivative of Ψ of order d with respect to each variable and make sure it is bounded.

Let us now move away from the CDM method and consider a general transformation ϕ_C . In order to study $\|\Psi \circ \phi_C\|_{d,1}$, we first need to decompose mixed partial derivatives of the form

$$\frac{\partial^l (\Psi \circ \phi_C)(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial v_{\alpha_1} \dots \partial v_{\alpha_l}}$$

in terms of Ψ and ϕ_C separately. To do so, we follow [62], as well as [27, Theorem 2.1], and obtain an expression for the mixed partial derivative of a composition of functions via the representation

$$\begin{aligned} & \frac{\partial^l \Psi \circ \phi_C(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial v_{\alpha_1} \dots \partial v_{\alpha_l}} \\ &= \sum_{1 \leq |\beta| \leq l} \frac{\partial^{|\beta|} \Psi}{\partial^{\beta_1} u_1 \dots \partial^{\beta_d} u_d} \sum_{s=1}^l \sum_{\gamma, \mathbf{k}} c_{\gamma} \prod_{j=1}^s \frac{\partial^{|\gamma_j|} \phi_{C, k_j}(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial^{\gamma_{j,1}} v_{\alpha_1} \dots \partial^{\gamma_{j,l}} v_{\alpha_l}} \end{aligned} \quad (20)$$

where $\beta \in \mathbb{N}_0^d$ and $|\beta| = \sum_{j=1}^d \beta_j$. Here we do not specify over which values of γ_j and k_j the inner sum in the above expression is taken: details can be found in the proof of Proposition 4.5. But let us point out that in the product over j , each index $\alpha_1, \dots, \alpha_l$ appears exactly once. On the other hand – and as noted in item 2 of Remark 4.4 above – in the mixed partial derivative of Ψ , a given variable can appear with order larger than 1.

From (20), we see that a sufficient condition to show that $\|\Psi \circ \phi_C\|_{d,1} < \infty$ is to establish that all products of the form

$$\frac{\partial^{|\beta|} \Psi}{\partial^{\beta_1} u_1 \dots \partial^{\beta_d} u_d} \prod_{j=1}^s \frac{\partial^{|\gamma_j|} \phi_{C, k_j}(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial^{\gamma_{j,1}} v_{\alpha_1} \dots \partial^{\gamma_{j,l}} v_{\alpha_l}}, \quad s \in \{1, \dots, l\}, \quad (21)$$

are in L_1 .

We note that for the MO algorithm (assuming as we did in (16) that v_1 is used to generate V and v_{j+1} is used to generate E_j), $\phi_{C,j}$ is a function of v_1 and v_{j+1} only, for $j \in \{1, \dots, d\}$. Hence the only partial derivatives of $\phi_{C,j}$ that are nonzero are those with respect to variables in $\{v_1, v_{j+1}\}$. This observation is helpful to prove the following result, which shows that the error bound obtained when using the MO algorithm has the desired behavior induced by the low-discrepancy point set used to generate the copula samples; note that it does not show that $\Psi \circ \phi_C$ has bounded variation. Its proof can be found in the appendix.

Proposition 4.5 (Error behaviour for MO for continuous V). *Let ϕ_C^{MO} be the transformation associated with the Marshall–Olkin algorithm, as given in (16), and that $V \sim F$ is continuously distributed. Let $P_n = \{\mathbf{v}_i, i = 1, \dots, n\}$ be the point set in $[0, 1]^{d+1}$ used to produce copula samples via the transformation ϕ_C^{MO} and let $\mathbf{u}_i = \phi_C^{MO}(\mathbf{v}_i)$. If*

1. the point set P_n excludes the origin and there exists some $p \geq 1$ such that $\min_{1 \leq i \leq n} v_{i,1} \geq 1/pn$;
2. the function Ψ satisfies $|\Psi(\mathbf{u})| < \infty$ for all $\mathbf{u} \in [0, 1]^{d+1}$ and

$$\frac{\partial^{|\boldsymbol{\beta}|} \Psi}{\partial^{\beta_1} u_1 \dots \partial^{\beta_d} u_d} < \infty \text{ for all } \boldsymbol{\beta} = (\beta_1, \dots, \beta_d), \quad (22)$$

with $\beta_l \in \{0, \dots, d\}$ and $|\boldsymbol{\beta}| \leq d$;

3. and the generator $\psi(\cdot)$ of the Archimedean copula C is such that

- (a) $\psi'(t) + t\psi''(t)$ has at most one zero t^* in $(0, \infty)$ and it satisfies $-t^*\psi'(t^*) < \infty$; and
- (b) $F^{-1}(1 - 1/pn)$ is in $O(n^a)$ for some constant $a > 0$;

then there exists a constant $C^{(d)}$ (independent of n but dependent of Ψ and ϕ_C^{MO}) such that

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq C^{(d)}(\log n) D^*(P_n).$$

Remark 4.6. We note that if $\mathbb{E}[V] < \infty$, as is the case for Clayton's copula family, Condition 3 3b can be easily checked via Markov's inequality. In the case of the Gumbel copula, V has an α -stable distribution and it can be shown that $P(V > x) \leq cx^{-\alpha}$ for $x \geq x_0$ and for some constant c , where c and x_0 depend both on the parameters of the distribution; see [103, Theorem 1.12]. Therefore $F^{-1}(1 - 1/pn)$ can be bounded by a constant time n^a in this case (namely by $\max\{x_0, (cpn)^{1/\alpha}\}$). As for Condition 3 3a, one can show that $t^* = \theta$ and $t^* = 1$ are the only zeros for the Clayton and Gumbel copulas, respectively.

When F is discrete, we can split the problem into subproblems based on the value taken by V . Then, in each case, the bounded variation condition is much easier to verify, because the transformation ϕ_C given V is essentially one-dimensional as it is mapping each v_j to an exponential E_{j-1} for $j \in \{2, \dots, d+1\}$. Its proof can be found in the appendix.

Proposition 4.7 (Error behaviour for the Marshall–Olkin algorithm for discrete V). *Let ϕ_C^{MO} be the transformation associated with the Marshall–Olkin algorithm, as given in (16) and assume C is an Archimedean copula whose distribution function F of V is discrete. Let $P_n = \{\mathbf{v}_i, i = 1, \dots, n\}$ be the point set in $[0, 1]^{d+1}$ used to produce copula samples via the transformation ϕ_C^{MO} and let $\mathbf{u}_i = \phi_C^{MO}(\mathbf{v}_i)$. If (22) holds and*

1. there exists some $p \geq 1$ such that the point set P_n satisfies $\max_{1 \leq i \leq n} v_{i,1} \leq 1 - 1/pn$;
2. there exist constants $c > 0$ and $q \in (0, 1)$ such that $1 - F(l) \leq cq^l$ for $l \geq 1$;

then there exists a constant $C^{(d)}$ (independent of n but dependent of Ψ and ϕ_C^{MO}) such that

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq C^{(d)}(\log n) D^*(P_n).$$

Remark 4.8. We note that the Frank, Joe and Ali-Mikhail-Haq copulas are such that F is discrete. The condition on the tail of F stated in the proposition can be shown to hold for the Frank and Ali-Mikhail-Haq copulas, but not for the Joe copula (the distribution of V in this case has a Sibuya distribution, for which no moments exists, i.e., it has a very fat tail).

Let us now move on to RQMC methods. We already mentioned that an advantage they have over their deterministic counterparts is that much weaker conditions are required to provide variance expressions for their corresponding estimators. The following result shows that this also holds after composing Ψ with ϕ_C .

Proposition 4.9 (Variance expression with a change of variables). *If $\tilde{P}_n = \{\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_n\}$ is a randomly digitally shifted net with corresponding RQMC estimator $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \Psi(\phi_C(\tilde{\mathbf{v}}_i))$ and if $\text{Var}(\Psi(\mathbf{U})) < \infty$ with $\mathbf{U} \sim C$, then we have that*

$$\text{Var}(\hat{\mu}_n) = \sum_{\mathbf{0} \neq \mathbf{h} \in \mathcal{L}_d^*} |\widehat{\Psi \circ \phi_C}(\mathbf{h})|^2, \quad (23)$$

where \mathcal{L}_d^* is the dual net of the deterministic net that has been shifted to get \tilde{P}_n , and $\hat{f}(\mathbf{h})$ is the Walsh coefficient of f at \mathbf{h} , while

$$\begin{aligned} (\widehat{\Psi \circ \phi_C})(\mathbf{h}) &= \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{\Psi}(\mathbf{k}) P(\mathbf{h}, \mathbf{k}), \\ P(\mathbf{h}, \mathbf{k}) &= \int_{[0,1]^d} e^{2\pi i(\mathbf{k} \cdot \phi_C(\mathbf{w}) - \mathbf{h} \cdot \mathbf{w})} d\mathbf{w} \\ &= \mathbb{E} \left[e^{2\pi i(\mathbf{k} \cdot \phi_C(\mathbf{W}) - \mathbf{h} \cdot \mathbf{W})} \right], \quad \mathbf{W} \sim \text{U}[0, 1]^d. \end{aligned}$$

Proof. It is clear from Theorem 6.1 in the appendix and using Representation (17) that (23) holds and the condition $\text{Var}(\Psi(\mathbf{U})) < \infty$ with $\mathbf{U} \sim C$ ensures it is finite. So what remains to be shown is the expression for the Walsh coefficient of the composed function $\Psi \circ \phi_C$. It is obtained as follows:

$$\begin{aligned} (\widehat{\Psi \circ \phi_C})(\mathbf{h}) &= \int_{[0,1]^d} \Psi(\phi_C(\mathbf{w})) e^{-2\pi i \langle \mathbf{h}, \mathbf{w} \rangle_b} d\mathbf{w} \\ &= \int_{[0,1]^d} \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{\Psi}(\mathbf{k}) e^{2\pi i \langle \mathbf{k}, \phi_C(\mathbf{w}) \rangle_b} e^{-2\pi i \langle \mathbf{h}, \mathbf{w} \rangle_b} d\mathbf{w} \\ &= \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{\Psi}(\mathbf{k}) \int_{[0,1]^d} e^{2\pi i (\langle \mathbf{k}, \phi_C(\mathbf{w}) \rangle_b - \langle \mathbf{h}, \mathbf{w} \rangle_b)} d\mathbf{w} \\ &= \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{\Psi}(\mathbf{k}) P(\mathbf{h}, \mathbf{k}), \end{aligned}$$

where the third equality holds thanks to Fubini's theorem. □

By adding assumptions on the smoothness of Ψ and thus on the behavior of its Walsh coefficients, one could obtain improved convergence rates for the variance given in (23) compared to the $O(1/n)$ we get with MC, something we plan to study in future work.

4.2 Transforming the low-discrepancy samples

As mentioned in the introduction, we can think of ϕ_C as transforming the point set P_n instead of being composed with Ψ . The integration error can then be analyzed via a generalized version of the Koksma–Hlawka inequality such as the one studied in [2], which we now explain.

Similarly to the Lebesgue case we define the *copula-discrepancy function* with respect to a copula-induced measure P_C on an interval B (i.e., $P_C(B) = \mathbb{P}(\mathbf{U} \in B)$ for $\mathbf{U} \sim C$) as

$$E_C(B; P_n) = \frac{A(B; P_n)}{n} - P_C(B).$$

Let \mathcal{J} be the set of intervals of $[0, 1]^d$ of the form $[\mathbf{a}, \mathbf{b}] = \prod_{j=1}^d [a_j, b_j]$, where $0 \leq a_j \leq b_j \leq 1$. The *copula-discrepancy* D_C of P_n is then defined as

$$D_C(P_n) = \sup_{B \in \mathcal{J}} |E_C(B; P_n)|, \quad (24)$$

and similarly for $D_C^*(P_n)$, the *star-copula-discrepancy function* when the sup in (24) is taken over \mathcal{J}^* instead.

The generalization of the Koksma–Hlawka inequality studied in [2, Theorem 1] then provides

$$\left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \leq V(\Psi) D_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n),$$

where we assume $\mathbf{u}_i = \phi_C(\mathbf{v}_i)$, $i \in \{1, \dots, n\}$. To get some insight on this upper bound, we need to know how $D_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n)$ behaves as a function of n . Unfortunately, in general we cannot prove that $D^*(\mathbf{v}_1, \dots, \mathbf{v}_n) \in O(n^{-1} \log^d n)$ implies that $D_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n) \in O(n^{-1} \log^d n)$. Here are a few things we can say, though.

First, an obvious case for which discrepancy is preserved is when ϕ_C maps rectangles to rectangles, because then $\phi_C(B) \in \mathcal{J}$ for all $B \in \mathcal{J}$, and thus

$$\begin{aligned} D_C(\tilde{P}_n) &\leq D(P_n), \\ D_C^*(\tilde{P}_n) &\leq D^*(P_n), \end{aligned}$$

where $\tilde{P}_n = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$. However, this only happens when C is the independence copula, and in this case the equality holds. This is not a very interesting case since our focus here is on dependence modelling.

For the more realistic setting where ϕ_C does not map rectangles to rectangles, the following result from [62] holds and gives a much slower convergence rate for $D_C^*(\tilde{P}_n)$.

Proposition 4.10. *Let C be such that the Rosenblatt transform ϕ_C^{-1} is Lipschitz continuous on $[0, 1]^d$ w.r.t. the sup-norm $\|\cdot\|_\infty$, and $\{\mathbf{u}_i = \phi_C(\mathbf{v}_i)\}$ for some sequence of points $\{\mathbf{v}_i\}$ in $[0, 1]^d$. Then*

$$D_C(\{\mathbf{u}_1, \dots, \mathbf{u}_n\}) \leq c(d) D(\{\mathbf{v}_1, \dots, \mathbf{v}_n\})^{1/d},$$

for some function $c(d)$, constant in n .

Note that the above results fully depend on the properties of ϕ_C . The aim would then be to choose ϕ_C such that a low-discrepancy sequence $\{\phi_C(\mathbf{v}_i)\}$ w.r.t. the copula measure P_C results whenever applied to a low-(Lebesgue-)discrepancy sequence $\{\mathbf{v}_i\}$. A more fundamental approach would be to directly produce a low-discrepancy sequence $\{\mathbf{u}_i\}$ where the discrepancy is measured w.r.t. the copula measure C . This is something we intend to study in future work.

Now, computing D_C or D_C^* is usually not feasible in practice. If we replace the sup-norm by the L_2 -norm, we obtain L_2 -discrepancies which are usually more practical to compute. Let L_2 -discrepancies $T_C(\mathbf{u}_1, \dots, \mathbf{u}_n)$ and $T_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n)$ be defined by

$$T_C(\mathbf{u}_1, \dots, \mathbf{u}_n) = \left(\int_{\{(\mathbf{y}, \mathbf{z}) \in [0,1]^{2d}; y_i < z_i\}} \left(\frac{A([\mathbf{y}, \mathbf{z}]; P_n)}{n} - P_C([\mathbf{y}, \mathbf{z}]) \right)^2 d\mathbf{y}d\mathbf{z} \right)^{1/2},$$

and

$$T_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n) = \left(\int_{[0,1]^d} \left(\frac{A([\mathbf{0}, \mathbf{z}]; P_n)}{n} - C(\mathbf{z}) \right)^2 d\mathbf{z} \right)^{1/2},$$

respectively. Proceeding similarly to [95], T_C^* can be computed as

$$\begin{aligned} T_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n) &= \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n \prod_{i=1}^d (1 - \max(u_{k,i}, u_{l,i})) + \int_{[0,1]^d} C(\mathbf{z})^2 d\mathbf{z} \\ &\quad - \frac{2}{n} \sum_{k=1}^n \int_{u_{k,1}}^1 \dots \int_{u_{k,d}}^1 C(\mathbf{z}) d\mathbf{z}. \end{aligned}$$

If we consider a convex combination $C(u_1, \dots, u_d) = \lambda \prod_{i=1}^d u_i + (1 - \lambda) \min(u_1, \dots, u_d)$, $\lambda \in (0, 1)$, of the independence copula and the upper Fréchet–Hoeffding bound, then one can compute T_C^* explicitly via

$$\begin{aligned} T_C^*(\mathbf{u}_1, \dots, \mathbf{u}_n) &= \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n \prod_{i=1}^d (1 - \max(u_{k,i}, u_{l,i})) + \frac{\lambda^2}{3^d} \\ &\quad + \frac{2(1 - \lambda)^2}{(d + 1)(d + 2)} + \frac{2\lambda(1 - \lambda)d!}{\prod_{i=1}^d (2i + 1)} - \frac{\lambda}{n2^{d-1}} \sum_{k=1}^n \prod_{i=1}^d (1 - u_{k,i}^2) \\ &\quad - \frac{2(1 - \lambda)}{n} \sum_{k=1}^n \left(\sum_{i_1=1}^d \sum_{i_2 \neq i_1}^d \sum_{i_d \neq i_1, \dots, i_{d-1}}^d \frac{1 - u_{k,i_d}^{d+1}}{(d + 1)!} \right. \\ &\quad \left. - \sum_{l=1}^{d-1} \sum_{i_1=1}^d \dots \sum_{i_l \neq i_1, \dots, i_{l-1}}^d \frac{u_{k,i_l}^{l+1} (1 - u_{k,i_{l+1}})}{(l + 1)!} \right). \end{aligned}$$

5 Numerical results

Through typical examples from the realm of finance and insurance and a few test functions, we now illustrate in this section the efficiency of QRNG in comparison to standard (P)RNG for copula

sampling. More precisely, we compare Monte Carlo sampling approaches with two types of QRNGs based on randomized low-discrepancy sequences: The Sobol' sequence and the generalized Halton sequence, both randomized with a digital shift. Variance/error estimates are obtained by using $B = 25$ i.i.d. copies of the randomized sequence and comparisons are made with MC sampling based on the same total number of replications. Each plot includes lines showing $n^{-0.5}$, n^{-1} and/or $n^{1.5}$ convergence rates. In addition, on top of each plot and for each QRNG method, we provide the regression estimate of $\alpha < 0$ such that the variance/error is in $O(n^\alpha)$. For PRNG, we only show the results with the CDM sampling algorithm, since the choice of method does not affect the error or variance very much. On the other hand, for QRNG we show the results both with CDM and MO (when applicable), as this seems to sometimes make a difference. Understanding better why it is so and under what circumstances a sampling algorithm perform better when used in conjunction with QRNG will be a subject of further research.

While the examples given in the next section illustrates the use of our proposed method in typical contexts where they might be used, the test functions results in the section that follows are meant to focus on assessing the performance of QRNG compared to PRNG on the sole basis of generating copula samples \mathbf{U} – without including the effect of the marginal distributions – and also to see if the sampling algorithm (CDM or MO) has an effect on the performance of QRNG.

Finally, we note that QRNG based on Sobol' point sets is typically slightly faster than PRNG, while the generalized Halton sequence runs slower than PRNG.

5.1 Examples from the realm of finance and insurance

Consider a random vector $\mathbf{X} = (X_1, \dots, X_d)$ modeling d risks in a portfolio of stocks or insurance losses. We assume that the j th marginal distribution is either log-normal with $X_j \sim \text{LN}(\log(100) + \mu - \sigma^2/2, \sigma^2)$, $j \in \{1, \dots, d\}$, where $\mu = 0.0001$ and $\sigma = 0.2$, or Pareto distributed with the same mean and variance as in the log-normal case. The copula C of \mathbf{X} throughout this numerical study is either a Clayton or an exchangeable t copula with three degrees of freedom. To allow a comparable degree of dependence, we will use the same Kendall's tau for both models. This easily translates to the parameter θ of a Clayton copula via the relationship $\theta = 2\tau(1 - \tau)^{-1}$ and to the correlation parameter ρ of an exchangeable t copula via $\rho = \sin(\pi\tau/2)$. We denote $S = \sum_{j=1}^d X_j$ and consider the estimation of the following functionals $\Psi(\mathbf{X})$:

- the Best-Of Call option payoff $(\max X_i - K)^+$;
- the Basket Call option payoff $(d^{-1}S - K)^+$;
- the Value-at-Risk at level 0.99 on the aggregated risks

$$\text{VaR}_{0.99}(S) = F_S^{-1}(0.99) = \inf \{x \in \mathbb{R} : F_S(x) \geq 0.99\},$$

- the expected shortfall at level 0.99 on the aggregated risks

$$\text{ES}_{0.99}(S) = \frac{1}{1 - 0.99} \int_{0.99}^1 F_S^{-1}(u) du;$$

- the contribution of the first and middle margin to $\text{ES}_{0.99}$ of the sum under the Euler principle, see [122],

$$\mathbb{E}[X_1 | S > F_S^{-1}(\alpha)] \text{ and } \mathbb{E}[X_{d/2} | S > F_S^{-1}(\alpha)].$$

These two functionals are referred to as Allocation First and Allocation Middle, respectively.

Figures 7, 8, 9, and 10 (as well as Figures 13 and 14 in the online supplement) display selected variance estimates for Clayton and t copulas with Kendall's tau parameter equal to 0.2 and 0.5, using either lognormal or Pareto margins, in dimensions $d = 5, 10, 20$ (displayed in different rows) and sample sizes $n \in \{10\,000, 15\,000, \dots, 200\,000\}$. In the Clayton case, the experiment uses both the MO and CDM sampling methods. For the t copulas, while there is a sampling approach based on a stochastic representation (as seen in Section 3.2.2), there is no version of the MO algorithm available, so we only use the CDM method. In addition, both the Sobol' and generalized Halton QRNGs are used. In all cases, we see that the variances associated with the Sobol' and generalized Halton quasi-random sequences are smaller and converge faster than the Monte Carlo variance. It is not clearly determined whether one sampling method is performing considerably better than the other. But we note that in some cases, such as the estimate of the Basket Call with $\tau = 0.2$ in $d = 20$ dimensions (Figure 7, bottom) the MO sampling seems to perform better than CDM.

5.2 Test functions

We now consider integration results on two different test functions. The results are shown in Figures 11, 12, 15, and 16 (the latter two are in the online supplement), which are based on a Clayton (or t) copula with $\tau = 0.2$ and $\tau = 0.5$, respectively. The first test function is given by

$$\Psi_1(\mathbf{u}) = 3(u_1^2 + \dots + u_d^2)/d,$$

where the vector \mathbf{u} is obtained after transforming the uniform points \mathbf{v} using either the CDM transform or the MO algorithm. Recall that the former requires d -dimensional points (using either a PRNG or a QRNG), whereas the latter requires $(d+1)$ -dimensional points. Note that Ψ_1 integrates exactly to 1 with respect to the copula-induced measure, since $U_j \sim U[0, 1]$, $j \in \{1, \dots, d\}$. While we know the exact value of the integral in this case, we still compare estimators based on B i.i.d. copies of either MC or RQMC, but we plot the average absolute error rather than the estimated variance.

The second test function is given by

$$\Psi_2(\mathbf{u}) = g_1((\phi^{\text{CDM}})^{-1}(\mathbf{u})),$$

where

$$g_1(\mathbf{v}) = \prod_{j=1}^d \frac{|4v_j - 1| + \alpha_j}{1 + \alpha_j}, \quad \alpha_j = j,$$

which is often used as a test function in the QMC literature; see, e.g., [49] and the references therein. So here we first apply the inverse of the CDM transform to the copula-transformed points obtained either using the CDM approach or MO, and then apply the d -dimensional function g_1 . While this has the effect of simply applying the standard test function g_1 to the original sample points \mathbf{v}_i in the case of the CDM, in the case of the MO algorithm, we are not falling back on the original points \mathbf{v}_i . The hope is that if MO does not preserve so well the low discrepancy of the original points, this function would be able to detect this problem.

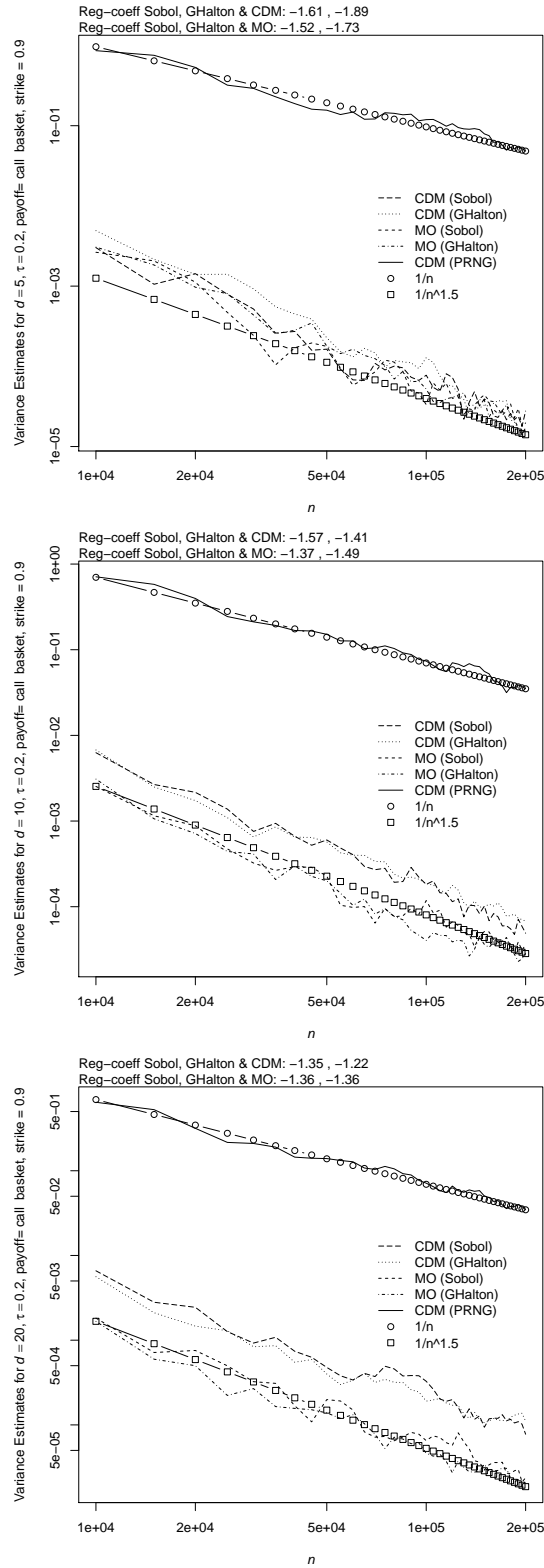


Figure 7: Variance estimates for the functional Basket Call with lognormal margins based on $B = 25$ repetitions for a Clayton copula with parameter such that Kendall's tau equals 0.2 for $d = 5$ (top), $d = 10$ (middle) and $d = 20$ (bottom).

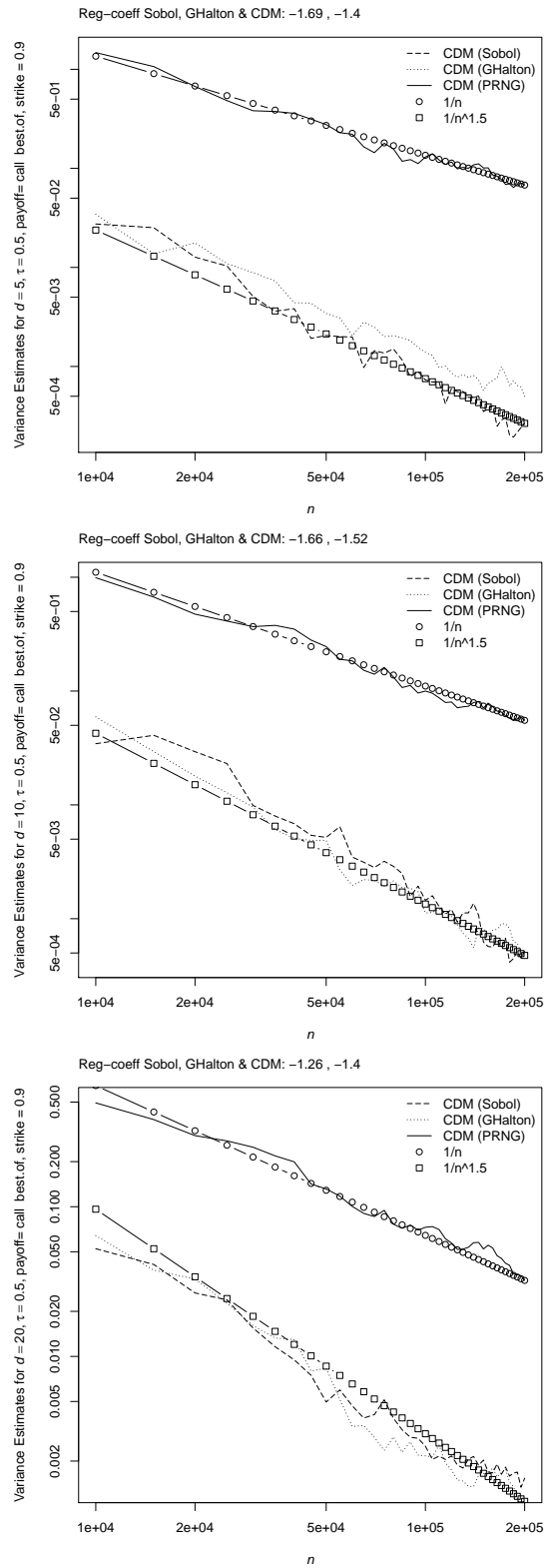


Figure 8: Variance estimates for the functional Best-Of Call with Pareto margins based on $B = 25$ repetitions for an exchangeable t copula with three degrees of freedom such that Kendall's tau equals 0.5 for $d = 5$ (top), $d = 10$ (middle) and $d = 20$ (bottom).

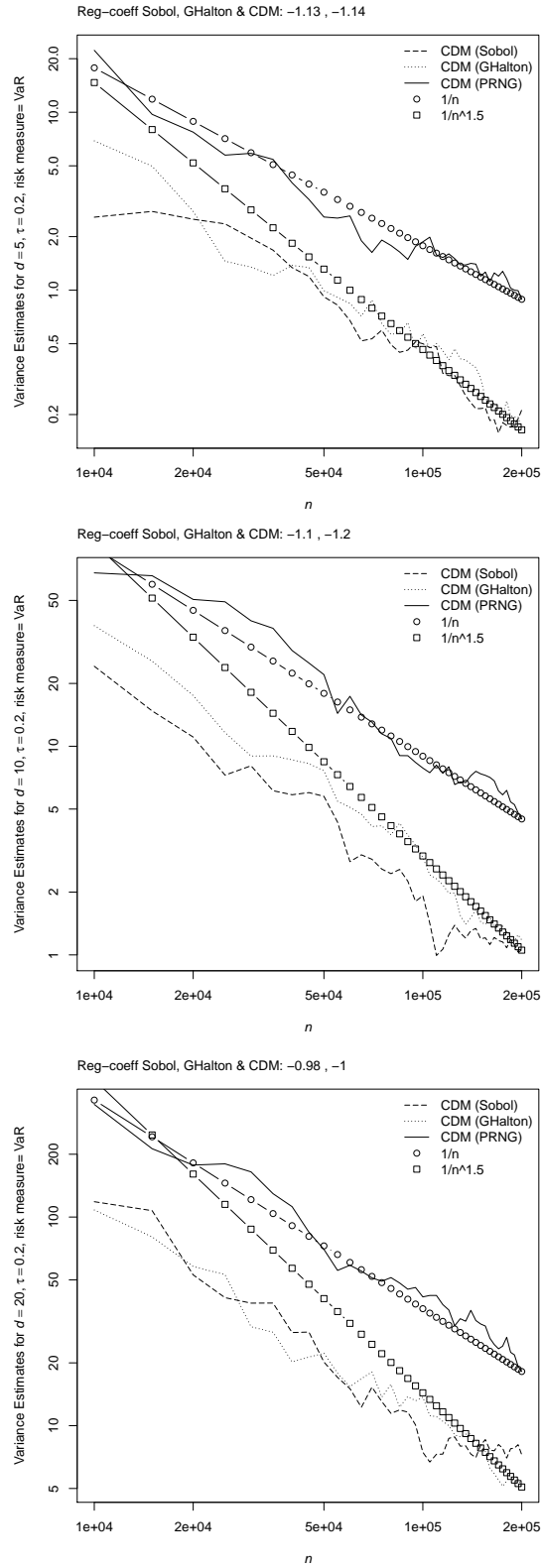


Figure 9: Variance estimates for the functional $\text{VaR}_{0.99}$ with lognormal margins for an exchangeable t copula with three degrees of freedom such that Kendall's tau equals 0.2 based on $B = 25$ repetitions for $d = 5$ (top), $d = 10$ (middle) and $d = 20$ (bottom).

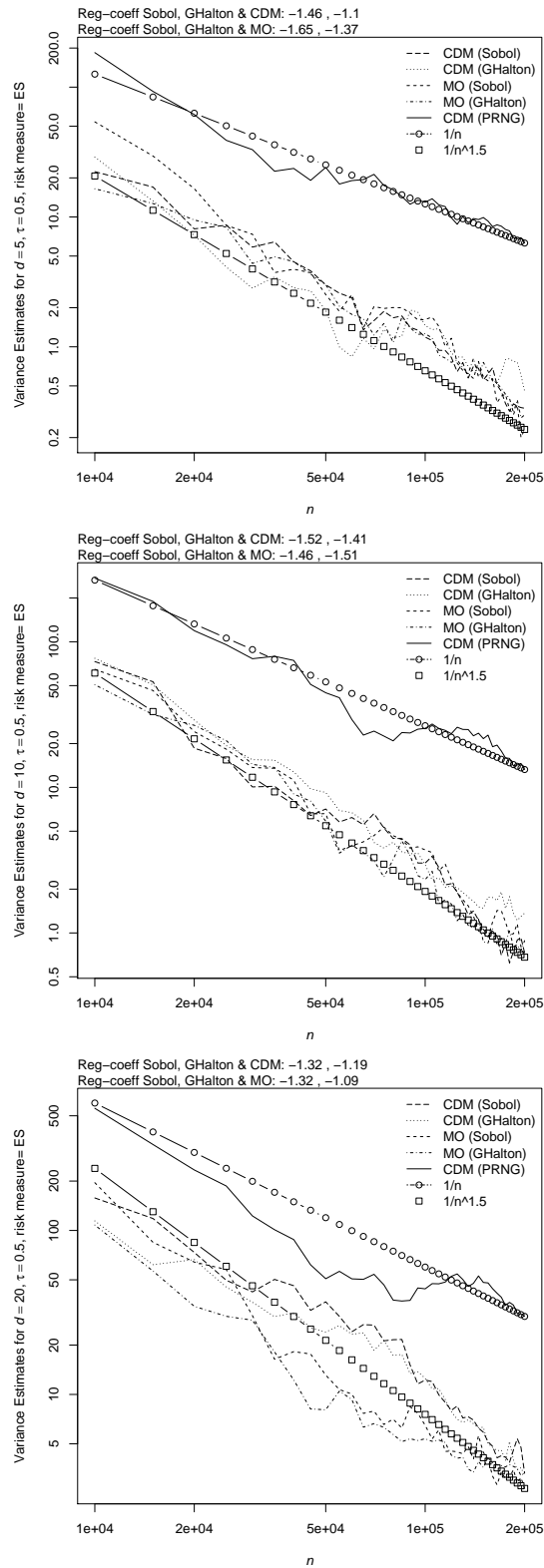


Figure 10: Variance estimates for the functional $\text{ES}_{0.99}$ with Pareto margins for a Clayton copula with parameter such that Kendall's tau equals 0.5 based on $B = 25$ repetitions for $d = 5$ (top), $d = 10$ (middle) and $d = 20$ (bottom).

While the second test function is mostly interesting for Archimedean copulas, the first one can be used more generally. This is why in the results reported in Figures 11 and 12, we also consider an exchangeable t copula with three degrees of freedom and Kendall’s τ equal to 0.2.

For both test functions, we see that the Sobol’ and generalized Halton sequences always clearly outperform Monte Carlo, with an error often converging in $O(n^{-1})$ rather than the $O(n^{-0.5})$ corresponding to Monte Carlo.

For the first function Ψ_1 , both the CDM and MO methods perform about the same. We believe this might be due to the simplicity of Ψ_1 —a sum of univariate powers of the u_j ’s—and the fact that both methods perform equally well in the univariate case when combined with RQMC.

Looking at the results for Ψ_2 , we see that with RQMC the CDM method performs better than MO, as there is no copula transformation performed in the case of CDM. However, RQMC with MO is still better than Monte Carlo, which suggests that the MO algorithm manages to preserve the low discrepancy of the original point set.

6 Conclusion and discussion

The main goal of this paper was to show how copula samples can be generated using quasi-random numbers. This is of interest when replacing PRNGs by QRNGs in applications involving dependent samples, possibly in higher dimensions. We have seen that different sampling approaches can be used, with a focus on the CDM approach and, additionally for Archimedean copulas, on the Marshall–Olkin algorithm. We have studied the error behaviour for both methods and have seen that in order to prove that the composed function $\Psi \circ \phi_C$ is smooth enough to satisfy the Koksma–Hlawka bound for the error, sufficient conditions on the function Ψ are that it must have smooth higher order mixed partial derivatives. For the Marshall–Olkin algorithm, we have shown that for several Archimedean copula families, the corresponding transformation ϕ_C is smooth enough to guarantee the good behaviour of the error bound. The superiority of QRNG over PRNG for copula sampling was illustrated on several examples, including a simulation addressing an application in the realm of finance and insurance. Most of the results in this paper are reproducible using the R packages `copula` and `qrng`.

Some ideas for future work would be to follow-up on Proposition 4.9 and to analyze the speed of decay of the Walsh coefficients of the composed function $\Psi \circ \phi_C$, based on assumptions on the speed of decay of the Walsh coefficients of Ψ and the properties of the sampling method ϕ_C .

Concerning the copula-induced discrepancy studied in Section 4.2, a possible avenue for future research would be to construct point sets that directly minimize this discrepancy, without first transforming a (uniform-based) low-discrepancy sample. In addition, proving error bounds based on the L_2 -discrepancy would be useful, as this discrepancy measure is easier to compute. Finally, numerically computing the copula-induced discrepancy for samples transformed either using the CDM or the MO algorithm would give us some insight as to how conservative the bound given in Proposition 4.10 is.

Appendix

For all the randomization schemes mentioned in Section 2, in addition to having a simple method to estimate the variance of the corresponding RQMC estimator, results giving exact expressions for this variance are also known and typically rely on using a well-chosen series expansion of the

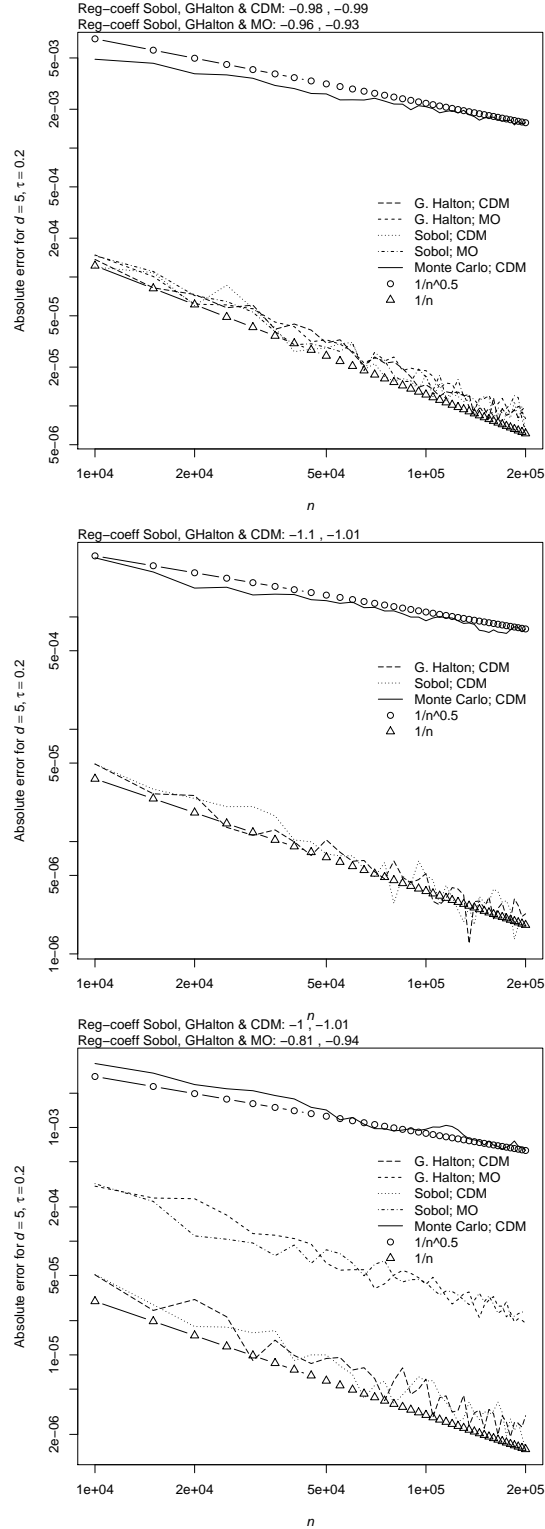


Figure 11: Average absolute errors for the test functions $\Psi_1(\mathbf{u}) = 3(u_1^2 + \dots + u_d^2)/d$ (top) and $\Psi_2(\mathbf{u}) = g_1((\phi^{\text{CDM}})^{-1}(\mathbf{u}))$ (bottom) for a Clayton copula with parameter such that Kendall's tau equals 0.2 based on $B = 25$ repetitions for $d = 5$; the middle plot shows results for $\Psi_1(\mathbf{u})$ and an exchangeable t copula with 3 degrees of freedom and Kendall's tau of 0.2.

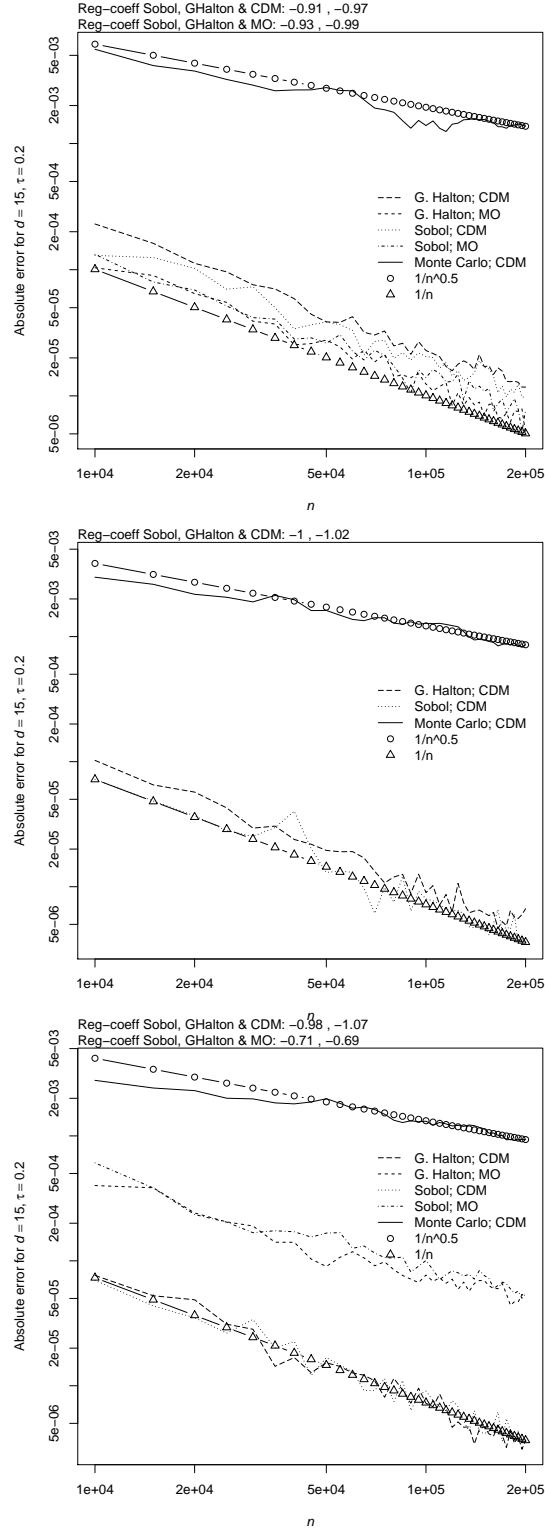


Figure 12: Average absolute errors for the test functions $\Psi_1(\mathbf{u}) = 3(u_1^2 + \dots + u_d^2)/d$ (top), $\Psi_2(\mathbf{u}) = K_C(C(\mathbf{u})) + 1/2$ (middle), and $\Psi_3(\mathbf{u}) = g_1((\phi^{\text{CDM}})^{-1}(\mathbf{u}))$ (bottom) for a Clayton copula with parameter such that Kendall's tau equals 0.2 based on $B = 25$ repetitions for $d = 15$; the middle plot shows results for $\Psi_1(\mathbf{u})$ and an exchangeable t copula with 3 degrees of freedom and Kendall's tau of 0.2.

function Ψ of interest. The following result recalls this variance expression in the case of randomly digitally shifted net; see [82] for a detailed proof. This result is used in the proof of Proposition 4.9 in Section 4.1.

Theorem 6.1 (Variance for randomly digitally shifted nets). *Let $\tilde{P}_n = \{\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_n\}$ be a randomly digitally shifted net in base b with corresponding RQMC estimator $\hat{\mu}_n$ given by*

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \Psi(\tilde{\mathbf{v}}_i)$$

and assume $\text{Var}(\Psi(\mathbf{U})) < \infty$ for $\mathbf{U} \sim U[0, 1]^d$. Then we have that

$$\text{Var}(\hat{\mu}_n) = \sum_{\mathbf{0} \neq \mathbf{h} \in \mathcal{L}_d^*} |\hat{\Psi}(\mathbf{h})|^2,$$

where $\hat{\Psi}(\mathbf{h})$ is the Walsh coefficient of Ψ at \mathbf{h} , given by

$$\hat{\Psi}(\mathbf{h}) = \int_{[0,1]^d} f(\mathbf{u}) e^{-2\pi i \langle \mathbf{h}, \mathbf{u} \rangle_b} d\mathbf{u}$$

where $\langle \mathbf{h}, \mathbf{u} \rangle_b = \frac{1}{b} \sum_{j=1}^d \sum_{l=0}^{\infty} h_{j,l} u_{j,l+1}$ with $h_{j,l}$ and $u_{j,l}$ obtained from the base b expansion of h_j and u_j , respectively, and $\mathcal{L}_d^* = \{\mathbf{h} \in \mathbb{Z}^d : \langle \mathbf{h}, \mathbf{v}_i \rangle_b \in \mathbb{Z}, \forall i = 1, \dots, n\}$ is the dual net of the deterministic net $P_n = \{\mathbf{v}_i, i = 1, \dots, n\}$ that has been shifted to get \tilde{P}_n .

Proofs

Proof of Proposition 3.4. Assume $P = \begin{pmatrix} P_{1:(j-1),1:(j-1)} & P_{1:(j-1),j} \\ P_{j,1:(j-1)} & P_{j,j} \end{pmatrix}$ and $P^{-1} = \begin{pmatrix} P_{1:(j-1),1:(j-1)}^{-1} & P_{1:(j-1),j}^{-1} \\ P_{j,1:(j-1)}^{-1} & P_{j,j}^{-1} \end{pmatrix}$ to be positive definite matrices. Corollary 3.3 implies that

$$C(u_j | u_1, \dots, u_{j-1}) = \frac{\int_{-\infty}^{x_j} h_{\nu,P}(x_1, \dots, x_{j-1}, z_j) dz_j}{h_{\nu,P_{1:(j-1),1:(j-1)}}(x_1, \dots, x_{j-1})},$$

where

$$h_{\nu,P}(x_1, \dots, x_j) = \frac{\Gamma(\frac{\nu+j}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{j/2} \sqrt{|P|}} \left(1 + \frac{\mathbf{x}^\top P^{-1} \mathbf{x}}{\nu} \right)^{-\frac{\nu+j}{2}} \quad (25)$$

is the density function of $t_{\nu,P}$. Using the block matrix equality

$$P_{1:(j-1),1:(j-1)}^{-1} - P_{1:(j-1),j}^{-1} (P_{j,j}^{-1})^{-1} P_{j,1:(j-1)}^{-1} = (P_{1:(j-1),1:(j-1)})^{-1},$$

we have that

$$\begin{aligned} & \mathbf{x}^\top P^{-1} \mathbf{x} \\ &= \mathbf{x}_{1:(j-1)}^\top P_{1:(j-1),1:(j-1)}^{-1} \mathbf{x}_{1:(j-1)} + x_j^2 P_{j,j}^{-1} + 2x_j \mathbf{x}_{1:(j-1)}^\top P_{1:(j-1),j}^{-1} \\ &= \mathbf{x}_{1:(j-1)}^\top (P_{1:(j-1),1:(j-1)})^{-1} \mathbf{x}_{1:(j-1)} + x_j^2 P_{j,j}^{-1} + 2x_j \mathbf{x}_{1:(j-1)}^\top P_{1:(j-1),j}^{-1} \\ & \quad + \mathbf{x}_{1:(j-1)}^\top P_{1:(j-1),j}^{-1} (P_{j,j}^{-1})^{-1} P_{j,1:(j-1)}^{-1} \mathbf{x}_{1:(j-1)} \\ &= g + k(x_j)^2, \end{aligned}$$

where

$$\begin{aligned} g &= \mathbf{x}_{1:(j-1)}^\top (P_{1:(j-1),1:(j-1)})^{-1} \mathbf{x}_{1:(j-1)}, \\ k(x_j)^2 &= \left(x_j \sqrt{P_{j,j}^{-1}} + s_2 \right)^2, \\ s_2 &= \mathbf{x}_{1:(j-1)}^\top P_{1:(j-1),j}^{-1} / \sqrt{P_{j,j}^{-1}}. \end{aligned}$$

We can thus rewrite (25) as

$$h_{\nu,P}(x_1, \dots, x_j) = a \left(1 + \frac{g}{\nu} \right)^{-\frac{\nu+j}{2}} h_{\nu+j-1}(l(x_j)),$$

where $h_{\nu+j-1}$ is the density of $t_{\nu+j-1}$ and

$$a = \frac{\Gamma(\frac{\nu+j-1}{2}) \sqrt{(\nu+j-1)\pi}}{\Gamma(\frac{\nu}{2})(\nu\pi)^{j/2} \sqrt{|P|}}, \quad l(x_j) = k(x_j) s_1, \quad s_1 = \sqrt{\frac{\nu+j-1}{\nu+g}}.$$

Using a change of variable argument, we compute

$$\int_{-\infty}^{x_j} h_{\nu,P}(x_1, \dots, x_{j-1}, z_j) dz_j = a \left(P_{j,j}^{-1} \frac{\nu+j-1}{\nu+g} \right)^{-1/2} t_{\nu+j-1}(l(x_j)).$$

Consequently,

$$C(u_j | u_1, \dots, u_{j-1}) = \sqrt{\frac{|P_{1:(j-1),1:(j-1)}|}{|P| P_{j,j}^{-1}}} t_{\nu+j-1}(l(x_j)) = t_{\nu+j-1}(l(x_j)),$$

where the last equality can be seen as follows. Let $P_{j|1:(j-1)}$ be as in (12). Since $|P| = |P_{1:(j-1),1:(j-1)}| |P_{j|1:(j-1)}|$, and $P_{j|1:(j-1)} = (P_{j,j}^{-1})^{-1}$ by [68, (0.7.3.1)], we then have

$$|P| = |P_{1:(j-1),1:(j-1)}| \cdot |P_{j|1:(j-1)}| = |P_{1:(j-1),1:(j-1)}| / P_{j,j}^{-1}.$$

□

Proof of Proposition 4.5. We start by providing more details on the expression (20), which is given by:

$$\begin{aligned} & \frac{\partial^l \Psi \circ \phi_C(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial v_{\alpha_1} \cdots \partial v_{\alpha_l}} = \\ & \sum_{1 \leq |\beta| \leq l} \frac{\partial^{|\beta|} \Psi}{\partial \beta^1 u_1 \cdots \partial \beta^d u_d} \sum_{s=1}^l \sum_{(\mathbf{k}, \gamma) \in p_s(\beta, \alpha)} c_\gamma \prod_{j=1}^s \frac{\partial^{|\gamma_j|} \phi_{C, k_j}(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial \gamma_{j,1} v_{\alpha_1} \cdots \partial \gamma_{j,l} v_{\alpha_l}} \end{aligned}$$

where $\beta \in \mathbb{N}_0^d$, $|\beta| = \sum_{j=1}^d \beta_j$, and the set $p_s(\beta, \alpha)$ includes pairs (\mathbf{k}, γ) such that \mathbf{k} is an s -dimensional vector $\mathbf{k} = (k_1, \dots, k_s)$ where each $k_j \in \{1, \dots, d\}$, and γ is an sl -dimensional vector $\gamma = (\gamma_1, \dots, \gamma_s)$ where each γ_j is an l -dimensional vector whose entries are either 0 or 1, and $\sum_{j=1}^s \gamma_{j,i} = 1$ for $i \in \{1, \dots, l\}$. Finally, the c_γ are constants, which are defined in detail in [27], along with further information on the precise definition of $p_s(\mathbf{k}, \gamma)$.

As mentioned in Section 4.1, a sufficient condition to show that $\|\Psi \circ \Phi_C\|_{d,1} < \infty$ is to establish that all products of the form (21) are in L_1 , which we recall is given by

$$\frac{\partial^{|\beta|} \Psi}{\partial^{\beta_1} u_1 \dots \partial^{\beta_d} u_d} \prod_{j=1}^s \frac{\partial^{|\gamma_j|} \phi_{C,k_j}(v_{\alpha_1}, \dots, v_{\alpha_l}, \mathbf{1})}{\partial^{\gamma_{j,1}} v_{\alpha_1} \dots \partial^{\gamma_{j,l}} v_{\alpha_l}},$$

for $s \in \{1, \dots, l\}$ and $(\mathbf{k}, \boldsymbol{\gamma}) \in p_s(\boldsymbol{\beta}, \boldsymbol{\alpha})$.

Recall also that for the MO algorithm, $\phi_{C,l}$ is a function of v_1 and v_{l+1} only, for $l = 1, \dots, d$. Hence the only partial derivatives of $\phi_{C,l}$ that are nonzero are those with respect to variables in $\{v_1, v_{l+1}\}$.

Now, since we assume that (22) holds, then it means we just need to show that the product found in (21) is in L_1 , under the conditions stated in the proposition. In turn, we first show that this holds if the following bounds hold for the mixed partial derivatives of ϕ_C :

$$\int_0^1 \left| \frac{\partial \phi_{C,l}(v_1 = 1, v_{l+1})}{\partial v_{l+1}} \right| dv_{l+1} < \infty, \quad (26)$$

$$\int_{[0,1]^l} \left| \frac{\partial^2 \phi_{C,1}(v_1, v_2)}{\partial v_1 \partial v_2} \prod_{j=2}^{l-1} \frac{\partial \phi_{C,j}(v_1, v_{j+1})}{\partial v_{j+1}} \right| dv_1 dv_2 \dots dv_l < \infty, \text{ and} \quad (27)$$

$$\int_{[0,1]^{l-1}} \left| \frac{\partial \phi_{C,r}(v_1, v_{r+1} = 1)}{\partial v_1} \left(\prod_{j=1, j \neq r}^{l-1} \left| \frac{\partial \phi_{C,j}(v_1, v_{j+1})}{\partial v_{j+1}} \right| dv_j \right) \right| dv_l < \infty \quad (28)$$

for all $l \leq d+1$.

We have three cases to consider.

Case 1: $1 \notin I$. Then the product in (21) is given by

$$\prod_{j=1}^l \left| \frac{\partial \phi_{C,j}(v_1 = 1, v_{j+1})}{\partial v_{j+1}} \right|,$$

where we assumed w.l.o.g. that $I = \{2, \dots, l+1\}$, $s = l$ and $k_j = j+1$ for $j \in \{1, \dots, s\}$. Since each term in the product depends on a distinct variable, the product is in L_1 if (26) holds.

Case 2: $1 \in I$ and j such that $\gamma_{j,1} = 1$ has $k_j + 1 \notin I$. This case can be analyzed w.l.o.g. by assuming I is of the form $I = \{1, \dots, r, r+2, \dots, l+1\}$ for some $r \geq 1$. In that case, the products in (21) are of the form

$$\left| \frac{\partial \phi_{C,r}(v_1, v_{r+1} = 1)}{\partial v_1} \prod_{j=1, j \neq r}^{l-1} \frac{\partial \phi_{C,j}(v_1, v_{j+1})}{\partial v_{j+1}} \right|$$

and is thus in L_1 as long as (28) holds.

Case 3: $1 \in I$ and j such that $\gamma_{j,1} = 1$ has $k_j + 1 \in I$. In this case, we can assume w.l.o.g. that $I = \{1, \dots, l\}$ and therefore the products in (21) are of the form

$$\left| \frac{\partial^2 \phi_{C,r}(v_1, v_{r+1})}{\partial v_1 \partial v_{r+1}} \prod_{j=1, j \neq r}^{l-1} \frac{\partial \phi_{C,j}(v_1, v_{j+1})}{\partial v_{j+1}} \right|$$

and is thus in L_1 as long as (27) holds.

The last part of the proof is to show that (26), (27), and (28) hold. First we study the partial derivatives involved in these expressions and find they are given by:

$$\begin{aligned}\frac{\partial \phi_{C,1}(v_1, v_2)}{\partial v_1} &= \psi' \left(-\frac{\log(v_2)}{x_1} \right) \frac{\log v_2}{x_1^2} \frac{\partial x_1}{\partial v_1}, \\ \frac{\partial \phi_{C,1}(v_1, v_2)}{\partial v_2} &= -\psi' \left(-\frac{\log(v_2)}{x_1} \right) \frac{1}{x_1 v_2}, \\ \frac{\partial^2 \phi_{C,1}(v_1, v_2)}{\partial v_1 \partial v_2} &= \frac{\partial x_1}{\partial v_1} \frac{1}{v_2 x_1^2} \left[\psi' \left(-\frac{\log v_2}{x_1} \right) - \frac{\log v_2}{x_1} \psi'' \left(-\frac{\log v_2}{x_1} \right) \right],\end{aligned}$$

where $x_1 = F^{-1}(v_1)$ and $\frac{\partial x_1}{\partial v_1} = 1/f(x_1)$, where f is the pdf corresponding to F , which exists since we assumed F was continuous. Now, the partial derivatives with respect to either v_1 or v_2 are clearly non-negative for all v_1 and v_2 . Hence it is easy to see that (26) and (28) hold, because we can remove the absolute value inside the integrals and therefore, these integrals amount to take differences/sums of $\phi_{C,r}(\cdot, \cdot)$ at different values over its domain, which obviously yields a finite value since $\phi_{C,r}(\cdot, \cdot)$ always takes values in $[0, 1]$.

As for the mixed partial derivative with respect to v_1 and v_2 , our assumption on $\psi'(t) + t\psi''(t)$ implies we have at most one sign change over the domain of the integral. If there is no sign change, the argument used in the previous paragraph to handle (26) and (28) can be used to show (27) is bounded. If there is one sign change, then we let t^* be such that

$$\psi'(t) + t\psi''(t) \leq 0 \quad \text{for } 0 \leq t \leq t^* \quad \text{and} \quad \psi'(t) + t\psi''(t) \geq 0 \quad \text{for } t^* \leq t.$$

Then let $q(v)$ be a function such that $-\log q(v)/F^{-1}(v) = t^*$. For instance, one can verify that for the Clayton copula, $q(v) = e^{-\theta F^{-1}(v)}$. When integrating the absolute value of the mixed partial derivative $\partial^2 \phi_{C,1}(v_1, v_2)/\partial v_1 \partial v_2$, we get

$$\begin{aligned}& \int_0^1 \frac{\partial x_1}{\partial v_1} \frac{1}{x_1^2} \left[\int_0^{q(v_1)} \frac{1}{v_2} \left(\psi' \left(-\frac{\log v_2}{x_1} \right) - \frac{\log v_2}{x_1} \psi'' \left(-\frac{\log v_2}{x_1} \right) \right) dv_2 \right. \\ & \quad \left. + \int_{q(v_1)}^1 \frac{1}{v_2} \left(-\psi' \left(-\frac{\log v_2}{x_1} \right) + \frac{\log v_2}{x_1} \psi'' \left(-\frac{\log v_2}{x_1} \right) \right) dv_2 \right] dv_1 \\ &= 2 \int_0^1 \frac{\partial x_1}{\partial v_1} \frac{1}{x_1^2} [\psi'(-\log q(v_1)/x_1) \log q(v_1)] dv_1 \\ &= -2t^* \psi'(t^*) \int_0^1 \frac{1}{F^{-1}(v_1)} \frac{\partial F^{-1}(v_1)'}{\partial v_1} dv_1 = -2t^* \psi'(t^*) \log F^{-1}(v_1)|_0^1.\end{aligned}\tag{29}$$

Now, in most cases $F^{-1}(1)$ is not bounded, and thus we cannot prove that $\Psi \circ \phi_C$ has bounded variation. However, from there we can still get the upper bound on the error given in the result, by using a technique initially developed by [119] to handle improper integrals, and later by [60] to deal with unbounded integration problems taken w.r.t. to a measure that is not necessarily uniform (as studied in Section 4.2). Note that to apply their approach more easily, we need to make a small change and assume that rather than generating V as $F^{-1}(v_1)$, we use $F^{-1}(1 - v_1)$, so that in our study of the variation above (via the integral of the absolute value of the mixed partial derivatives),

the boundedness condition fails at $v_1 = 0$ instead of $v_1 = 1$. Following the approach in [60] (see their Equation (24)) and taking $\mathbf{c} = (1/pn, 0, \dots, 0)$, the integration error satisfies

$$\begin{aligned} & \left| \frac{1}{n} \sum_{i=1}^n \Psi(\mathbf{u}_i) - \mathbb{E}[\Psi(\mathbf{U})] \right| \\ & \leq \frac{1}{pn} \Psi(1, \dots, 1) + D^*(P_n) V_{[\mathbf{c}, \mathbf{1}]}(\Psi \circ \phi_C) + I_{rest} \end{aligned}$$

where $V_{[\mathbf{c}, \mathbf{1}]}(\Psi \circ \phi_C)$ denotes the variation of $\Psi \circ \phi_C$ over $[\mathbf{c}, \mathbf{1}]$ and

$$I_{rest} = \left| \int_{\mathbf{0}}^{\mathbf{1}} \Psi \circ \phi_C(\mathbf{v}) d\mathbf{v} - \int_{\mathbf{c}}^{\mathbf{1}} \Psi \circ \phi_C(\mathbf{v}) d\mathbf{v} \right| \leq \frac{M}{pn} \text{ for some } M > 0,$$

since we assumed $|\psi(\mathbf{u})|$ was bounded. As for $V_{[\mathbf{c}, \mathbf{1}]}(\Psi \circ \phi_C)$, we can infer from the steps that led to (29) that it is bounded by a constant times $\log F^{-1}(1 - 1/pn) \leq a \log n + \log c$ by assumption. Therefore there exists a constant $K^{(d)}$ such that $V_{[\mathbf{c}, \mathbf{1}]}(\Psi \circ \phi_C) \leq K^{(d)} \log n$. \square

Proof of Proposition 4.7. Let p_l be such that $P(V = l) = p_l$, for $l \geq 1$. Let $P_l = \sum_{k=1}^l p_k$ for $l \geq 1$ and $P_0 = 0$. We also let $\phi_C^l(v_2, \dots, v_{d+1}) = \phi_C(P_{l-1}, v_2, \dots, v_{d+1})$ for $l \geq 1$ (transformation ϕ_C when v_1 generates the value l for V). Consider a given value of n and low-discrepancy point set P_n . If we use inversion to generate V , then we have that the subset $P_n^l = \{\mathbf{v}_i : P_{l-1} < v_{i,1} \leq P_l\}$ will be used to produce copula samples with $V = l$. Let $\tilde{n}_l = |P_n^l|$ and $n_l = np_l$. It is clear that if l becomes too large, then \tilde{n}_l will eventually be 0. Let $L(n)$ be the largest value of l such that $\tilde{n}_l > 0$, and let $\tilde{p}_l = \tilde{n}_l/n$. Then we can write

$$\begin{aligned} & \left| \int_{[0,1]^{d+1}} \Psi \circ \phi_C(\mathbf{v}) d\mathbf{v} - \frac{1}{n} \sum_{i=1}^n \Psi \circ \phi_C(\mathbf{v}_i) \right| \\ & \leq \left| \sum_{l=1}^{L(n)} p_l \left(\int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} - \frac{1}{n_l} \sum_{P_n^l} \Psi \circ \phi_C(\mathbf{v}_i) \right) \right| \\ & \quad + \sum_{l=L(n)+1}^{\infty} p_l \left| \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} \right| \\ & \leq \sum_{l=1}^{L(n)} \tilde{p}_l \left| \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} - \frac{1}{\tilde{n}_l} \sum_{P_n^l} \Psi \circ \phi_C(\mathbf{v}_i) \right| \\ & \quad + \sum_{l=L(n)+1}^{\infty} p_l \left| \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} \right| \\ & \quad + \sum_{l=1}^{L(n)} \left| (p_l - \tilde{p}_l) \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} \right| \\ & \leq \sum_{l=1}^{L(n)} (\tilde{p}_l A(n, d)) + B(n, d) + C(n, d), \end{aligned}$$

where $A(n, d)$, $B(n, d)$, and $C(n, d)$ are bounds such that

$$\begin{aligned} \left| \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} - \frac{1}{\tilde{n}_l} \sum_{P_n^l} \Psi \circ \phi_C(\mathbf{v}_i) \right| &\leq A(n, d) \\ \sum_{l=L(n)+1}^{\infty} p_l \left| \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} \right| &\leq B(n, d) \\ \sum_{l=1}^{L(n)} \left| (p_l - \tilde{p}_l) \int_{[0,1]^d} \Psi \circ \phi_C^l(\mathbf{v}) dv_2 \dots dv_{d+1} \right| &\leq C(n, d). \end{aligned}$$

First, by definition of $D^*(P_n)$ we have $|\tilde{n}_l - n_l| \leq 2nD^*(P_n)$ and thus $|\tilde{p}_l - p_l| \leq 2D^*(P_n)$. Hence we can take $C(n, d) = 2\mathbb{E}(|\Psi(\mathbf{U})|)D^*(P_n)$. Similarly, we can show that $\sum_{l=L(n)+1}^{\infty} p_l \leq D^*(P_n)$ and can therefore take $B(n, d) = \mathbb{E}(|\Psi(\mathbf{U})|)D^*(P_n)$. The analysis of the expression to be bounded by $A(n, d)$ is more complicated. First, we note that under the assumption we have on Ψ and its partial derivatives, we need to show that the product in (21) is in L_1 , but where each ϕ_{C,k_j} is replaced by ϕ_{C,k_j}^l for a given l . Since ϕ_{C,k_j}^l is solely a function of v_{k_j+1} , then it means that the only relevant products to consider are of the form

$$\prod_{j=1}^r \frac{\partial \phi_{C,k_j}^l(v_{k_j+1})}{\partial v_{k_j+1}} \quad (30)$$

in which each term is of the form $-\psi' \left(\frac{-\log v_{k_j+1}}{l} \right) \frac{1}{lv_{k_j+1}}$ which is non-negative for any v_{k_j+1} . Using a similar reasoning to the one used in the proof of Proposition 4.5 (to conclude that (26) and (28) hold), it is easy to see that (30) is in L_1 .

What remains to be done is to analyze the discrepancy of P_n^l . That is, here we are looking for a bound on $\sup_{\mathbf{z} \in \mathcal{J}^*} |E(\mathbf{z}; P_n^l)|$, where we recall that \mathcal{J}^* is the set of intervals of $[0, 1]^d$ of the form $\mathbf{z} = \prod_{j=1}^d [0, z_j]$, where $0 < z_j \leq 1$. So consider a given $\mathbf{z} \in [0, 1]^d$. Then $E(\mathbf{z}; P_n^l) = A(\mathbf{z}; P_n^l)/\tilde{n}_l - \lambda(\mathbf{z})$. Let $\mathbf{z}_1 = (P_l, \mathbf{z})$ and $\mathbf{z}_2 = (P_{l-1}, \mathbf{z})$, which are both in $[0, 1]^{d+1}$. Note that $A(\mathbf{z}_1; P_n) - A(\mathbf{z}_2; P_n) = A(\mathbf{z}; P_n^l)$. By definition of $D^*(P_n)$, it is not hard to see that

$$\left| \frac{(A(\mathbf{z}_1; P_n) - A(\mathbf{z}_2; P_n))}{n} - p_l \lambda(\mathbf{z}) \right| \leq 2D^*(P_n)$$

and therefore

$$\left| \frac{A(\mathbf{z}; P_n^l)}{\tilde{n}_l} - \frac{n_l}{\tilde{n}_l} \lambda(\mathbf{z}) \right| \leq 2D^*(P_n) \frac{n}{\tilde{n}_l}.$$

Using the fact that $|\tilde{n}_l - n_l| \leq 2nD^*(P_n)$, after some further simplifications we get that

$$\left| \frac{A(\mathbf{z}; P_n^l)}{\tilde{n}_l} - \lambda(\mathbf{z}) \right| \leq 4D^*(P_n) \frac{n}{\tilde{n}_l}.$$

Hence we can take $A(n, d) = 4D^*(P_n) \frac{n}{\tilde{n}_l}$ and then get $\sum_{l=1}^{L(n)} \tilde{p}_l A(n, d) \leq 4L(n)D^*(P_n)$. To show that the overall bound for the integration error is of the form $(\log n)D^*(P_n)$ times a constant, we

simply need to show that $L(n) \in O(\log n)$. But this follows from our assumptions on P_n and F , since by definition, $L(n)$ is the largest integer such that $1 - F(L(n)) > 1/pn$ but we also have $1 - F(L(n)) \leq cq^{L(n)}$, hence

$$1/pn < cq^{L(n)} \Leftrightarrow L(n) \log(1/q) - \log c < \log p + \log n$$

and thus $L(n) \leq (\log n + \log p + \log c) / \log(1/q)$, as required. \square

Online supplement

Additional numerical results

Here we provide a few additional results for the experimental setup described in Section 5, namely for the finance and insurance examples (see Figures 13 and 14) and then for the test functions (see Figures 15 and 16).

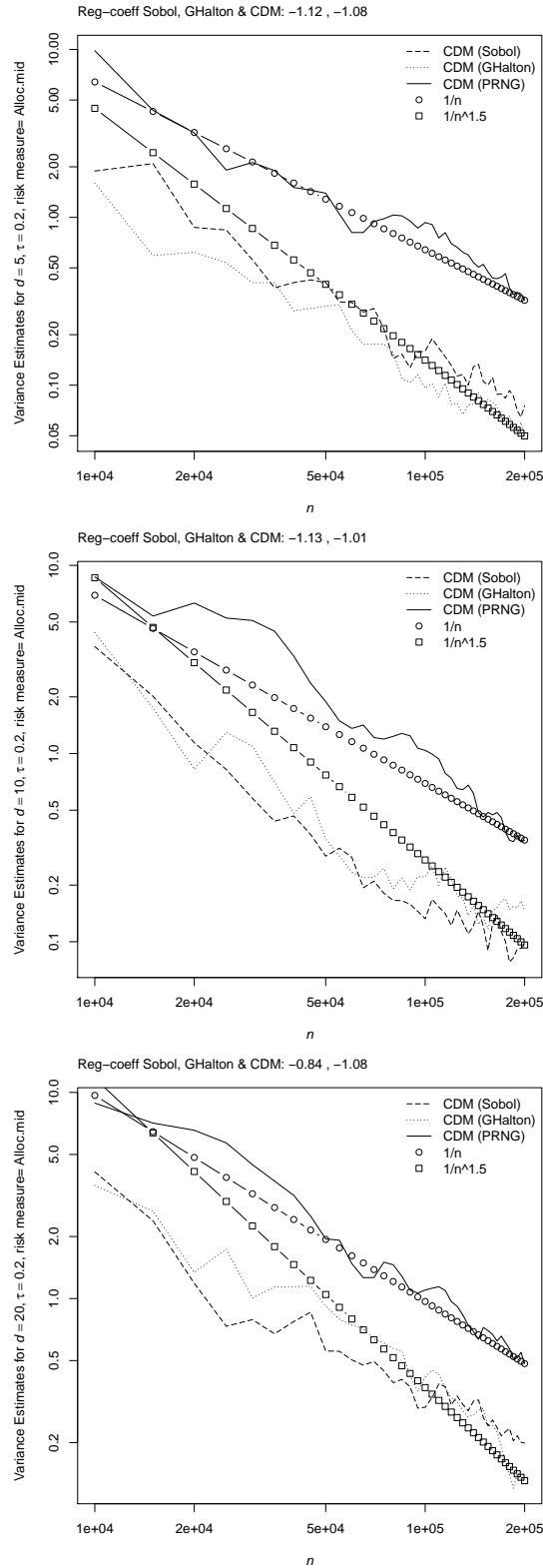


Figure 13: Variance estimates for the functionals Allocation First for lognormal margins and an exchangeable t copula with three degrees of freedom such that Kendall's tau equals 0.2 based on $B = 25$ repetitions for $d = 5$ (top), $d = 10$ (middle) and $d = 20$ (bottom).

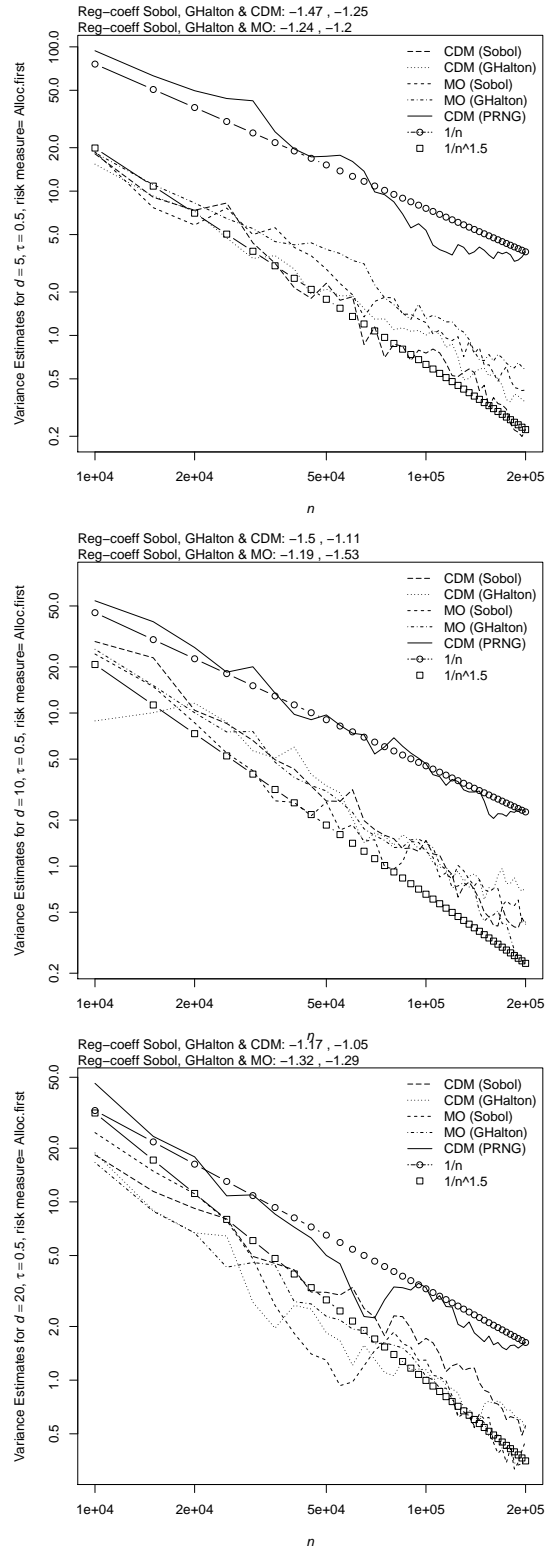


Figure 14: Variance estimates for the functionals Allocation Middle with Pareto margins and for a Clayton copula with parameter such that Kendall's tau equals 0.5 based on $B = 25$ repetitions for $d = 5$ (top), $d = 10$ (middle) and $d = 20$ (bottom).

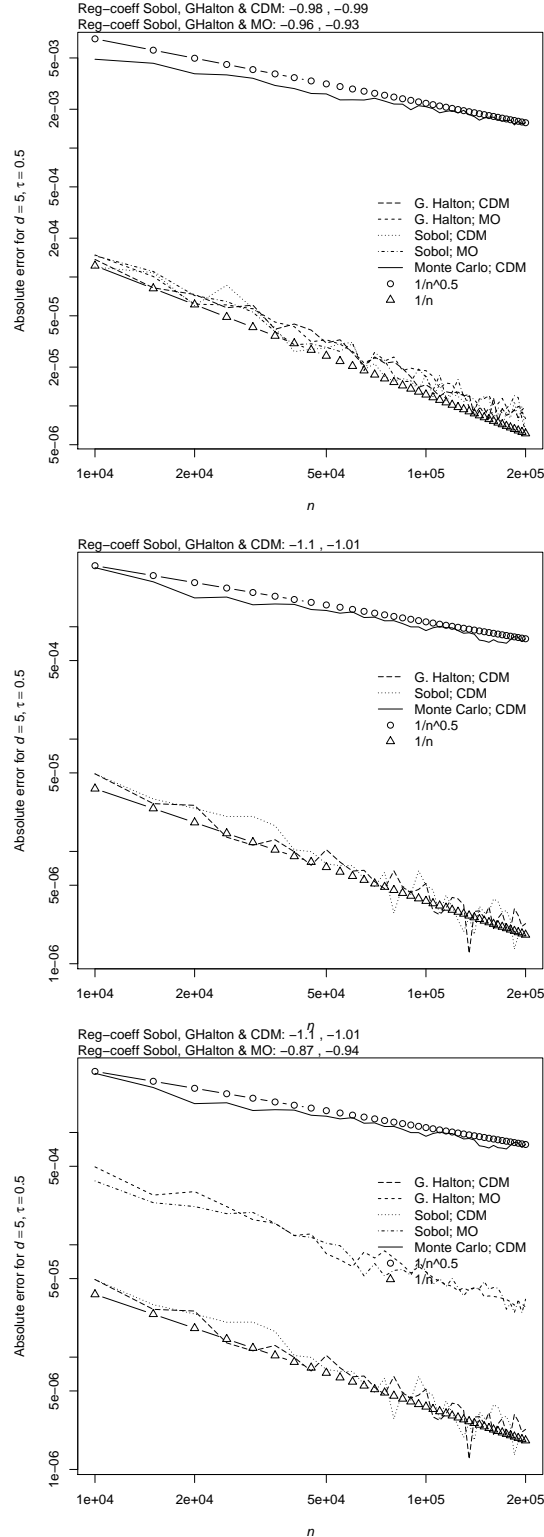


Figure 15: Average absolute errors for the test functions $\Psi_1(\mathbf{u}) = 3(u_1^2 + \dots + u_d^2)/d$ (top) and $\Psi_2(\mathbf{u}) = g_1((\phi^{\text{CDM}})^{-1}(\mathbf{u}))$ (bottom) for a Clayton copula with parameter such that Kendall's tau equals 0.5 based on $B = 25$ repetitions for $d = 5$: the middle plot shows results for $\Psi_1(\mathbf{u})$ and an exchangeable t copula with three degrees of freedom and Kendall's tau of 0.2.

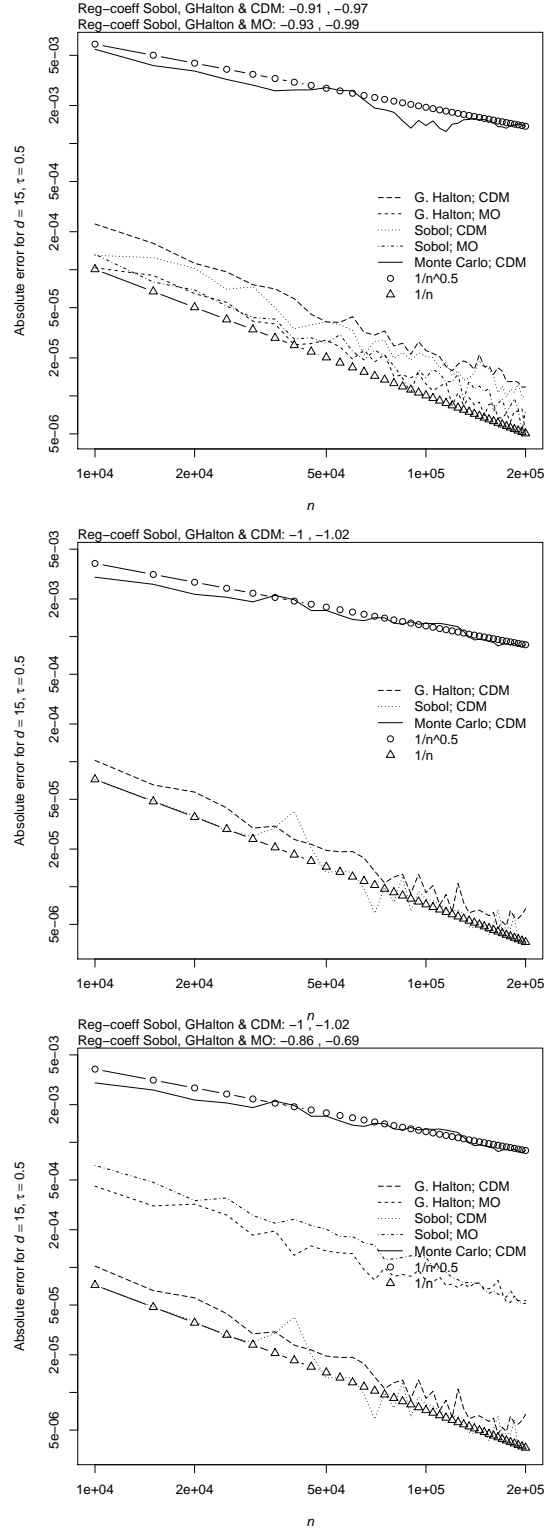


Figure 16: Average absolute errors for the test functions $\Psi_1(\mathbf{u}) = 3(u_1^2 + \dots + u_d^2)/d$ (top), $\Psi_2(\mathbf{u}) = g_1((\phi^{\text{CDM}})^{-1}(\mathbf{u}))$ (bottom) for a Clayton copula with parameter such that Kendall's tau equals 0.5 based on $B = 25$ repetitions for $d = 15$: the middle plot shows results for $\Psi_1(\mathbf{u})$ and an exchangeable t copula with three degrees of freedom and Kendall's tau of 0.2.

Paper D

Replicating Portfolio Approach to Capital Calculation

Replicating Portfolio Approach to Capital Calculation*

Mathieu Cambou[†] Damir Filipović[‡]

August 24, 2016

Abstract

The replicating portfolio (RP) approach to the calculation of capital for life insurance portfolios is an industry standard. The RP is obtained from projecting the terminal loss of discounted asset-liability cash flows on a set of factors generated by a family of financial instruments that can be efficiently simulated. We provide the mathematical foundations and a novel dynamic and path-dependent RP approach for real-world and risk-neutral sampling. We show that the RP approach yields asymptotically consistent capital estimators. We illustrate the tractability of the RP approach by two numerical examples.

Key words: asset-liability portfolio, chaos expansion, replicating portfolio, solvency capital

1 Introduction

The calculation of solvency and economic capital for life insurance portfolios is a challenging task, which has not gained much attention in the finance literature yet. The new European regulatory framework Solvency II and the Swiss Solvency Test require the modeling of the profit and loss distribution of the asset-liability portfolio on a one-year time horizon, see [24] and [53]. Solvency capital is determined as 99.5%-value at risk for Solvency II, and 99%-expected shortfall for the Swiss Solvency Test, of this one-year profit and loss.

The value of insurance liabilities is defined as (conditional) risk-neutral expectation of the accumulated discounted liability cash flows. For life insurance this requires simulations of cash flows up to 40 years and beyond. These simulations are computationally costly, the modeling of the one-year discounted loss on the asset-liability portfolio L cannot be done via nested simulations. The replicating portfolio (RP) approach consists in projecting the terminal time- T discounted loss of the asset-liability portfolio Z onto a set of factors generated by a family of m financial instruments, whose discounted gains processes can efficiently be simulated. This results in a dynamic and path-dependent self-financing portfolio, the RP, with discounted value process V_t that replicates Z in a

*We thank Matthias Aellig, Valérie Chavez, Michel Dacorogna, Anthony Davison, Guido Grützner, Stephan Morgenthaler, Antoon Pelsser, Johan Segers, Sonja Sterki, Ralf Werner, and participants at the Oberwolfach Workshop on The Mathematics and Statistics of Quantitative Risk Management 2015 for comments. An early version of this paper was implemented in the Master thesis of Haobo Jia, “New Aspects of the Replicating Portfolio for Group Life Insurance”, in the Financial Engineering program at EPFL and carried out at Swiss Life in 2013. The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 307465-POLYTE.

[†]EPFL, Institute of Mathematics, Lausanne, Switzerland, email: mathieucambou@gmail.com

[‡]EPFL and Swiss Finance Institute, Lausanne, Switzerland, email: damir.filipovic@epfl.ch

least-squares sense. We use $L_1 = V_1$ and alternatively $L_2 = V_1 + Z - V_T$ as proxies for L . We show that the RP approach is asymptotically consistent in the sense that the value at risk and expected shortfall of L_1 and L_2 converge to the true solvency capital, the value at risk and expected shortfall of L . Our results hold for least-squares projections under both the real-world and risk-neutral measures.

We also study the estimation errors that result from the finite sampling of Z . We show that the simulation based estimator of the RP is unbiased and satisfies the law of large numbers (LLN) and the central limit theorem (CLT) as the number n of simulations tends to infinity. As a consequence we find that the Monte-Carlo estimate is asymptotically consistent. Again, this is done under both the real-world and risk-neutral sampling measures.

The insurance market is incomplete under static hedging with the underlying financial instruments for two reasons. First, there are more factors driving insurance cash flows than there are traded financial instruments for their replication. Second, insurance cash flows are nonlinear in the underlying financial instruments. We illustrate both effects and the performance of the RP approach in two numerical examples. We find that our dynamic and path-dependent RP significantly outperforms the industry standard static RP.

Related literature is the following. [96] and [97] elaborate on the equivalence of optimal cash flow matching and optimal terminal value matching, based on the setups of [7] and [105]. In [98] they analyse bounds on the capital approximation error under a combination of real-world and risk-neutral measures. A discussion of least-squares Monte Carlo methods (so-called regress-now and regress-later) applied to insurance liability valuation is given in [12] and further elaborated in [111] and [13]. We add to this literature by providing a novel dynamic and path-dependent construction of the RP and establishing asymptotic consistency of the capital estimators under both the real-world and risk-neutral sampling measures. Our RP is multilinear in the running gains of the financial instruments such that it applies to generic insurance cash flows. In contrast, least-squares Monte Carlo assumes that insurance cash flows are defined as specific functions of the underlying factors. A technical report of the German Actuarial Society [32] compares various proxy methods for capital calculation in life insurance. They discuss pros and cons for the nested simulation, curve fitting, least-squares Monte Carlo, and RP approach. The advantage of RP over curve fitting and least-squares Monte Carlo is that the latter two use abstract function classes and underlying factors for the projection, while RP uses financial instruments as special functions with a clear model-independent interpretation. It is also advised to match terminal discounted value rather than cash flows because the scope is on the change of market value (profit and loss) of the portfolio, which is what we do in our paper. [32] also reports as typical numbers in practice $n = 1000$ to 5000 samples and $m = 10$ to 50 financial instruments.

The remainder of the paper is as follows. In Section 2 we introduce the formal setup. In Section 3 we develop the RP theory and analyze the capital approximation error. In Section 4 we analyze the Monte-Carlo error and prove asymptotic consistency of the capital proxy estimators. In Section 5 we provide two numerical examples, and in Section 6 we conclude. Appendix A contains the proofs of all lemmas and theorems. Appendix B contains all figures.

2 General setup

We consider an economy with a fixed time horizon T . Randomness is modeled on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, $t \leq T$, where \mathbb{P} denotes the real-world measure. All financial values are

discounted by some numeraire. The corresponding risk-neutral pricing measure $\mathbb{Q} \sim \mathbb{P}$ has density process $D_t = \frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t}$. For $p \geq 1$, we denote by $L^p(\mathbb{P})$ the Lebesgue space of random variables X with norm $\|X\|_{L^p(\mathbb{P})} = \mathbb{E}^{\mathbb{P}}[|X|^p]^{1/p}$ and similarly for $L^p(\mathbb{Q})$. Throughout, we assume that the following constants are finite

$$\begin{aligned} C_1^{\mathbb{P}} &= \|D_T\|_{L^2(\mathbb{P})}, & C_2^{\mathbb{P}} &= \|D_T/D_1\|_{L^2(\mathbb{P})}, \\ C_1^{\mathbb{Q}} &= \|1/D_T\|_{L^2(\mathbb{Q})}, & C_2^{\mathbb{Q}} &= \|1/D_1\|_{L^2(\mathbb{Q})}. \end{aligned}$$

Because $1/D_t$ is a \mathbb{Q} -martingale, we obviously have $C_2^{\mathbb{Q}} \leq C_1^{\mathbb{Q}}$.¹

Henceforth we let \mathbb{M} be a placeholder for either \mathbb{P} or \mathbb{Q} . Here is a basic lemma, which we will apply throughout the paper without further notice.

Lemma 2.1. *For a random variable X we have*

$$\|X\|_{L^1(\mathbb{Q})} \leq C_1^{\mathbb{P}}\|X\|_{L^2(\mathbb{P})}, \quad \|X\|_{L^1(\mathbb{P})} \leq C_1^{\mathbb{Q}}\|X\|_{L^2(\mathbb{Q})}.$$

Hence $L^2(\mathbb{P}) \subset L^1(\mathbb{Q})$ and $L^2(\mathbb{Q}) \subset L^1(\mathbb{P})$. If $X \in L^2(\mathbb{M})$ then

$$\left\| \mathbb{E}^{\mathbb{Q}}[X \mid \mathcal{F}_1] \right\|_{L^1(\mathbb{P})} \leq C_2^{\mathbb{M}}\|X\|_{L^2(\mathbb{M})}.$$

We denote by $\text{VaR}_\alpha[X]$ the \mathbb{P} -value at risk and by $\text{ES}_\alpha[X]$ the \mathbb{P} -expected shortfall of X for a fixed confidence level $\alpha \in (0, 1)$. For Solvency II the risk measure is $\text{VaR}_{99.5\%}$ and for the Swiss Solvency Test it is $\text{ES}_{99\%}$. For the definition and basic properties of value at risk and expected shortfall we refer the reader to [51]. For example, for any sequence of random variables X_n converging in law to a random variable X whose left and right α -quantiles coincide, $q_\alpha^-[X] = q_\alpha^+[X]$, we have

$$\lim_{n \rightarrow \infty} \text{VaR}_\alpha[X_n] = \text{VaR}_\alpha[X]. \quad (1)$$

The condition that left and right α -quantiles of the limit X have to coincide for the continuity (1) of value at risk to hold is not to be underestimated. Indeed, the situation $q_\alpha^-[X] \ll q_\alpha^+[X]$ is likely to occur in simulation-based models, which are commonly used in practice. For a quantification of this effect see [20, Example 6.3]. Value at risk also fails to be convex, which may have unpleasant consequences as illustrated in Example 3.7. Expected shortfall, on the other hand, is convex and Lipschitz continuous on $L^1(\mathbb{P})$,

$$|\text{ES}_\alpha[X] - \text{ES}_\alpha[Y]| \leq \frac{1}{1-\alpha} \|X - Y\|_{L^1(\mathbb{P})}. \quad (2)$$

Henceforth we let ρ be a placeholder for either VaR_α or ES_α .

3 Replicating portfolio theory

We consider an asset-liability portfolio whose discounted cash flows accumulate to the discounted terminal loss $Z \in L^2(\mathbb{M})$ at $t = T$. We assume the portfolio is fairly priced at $t = 0$ such that

¹There is no such relation between $C_1^{\mathbb{P}}$ and $C_2^{\mathbb{P}}$. For example, let D_T/D_1 be such that $\mathbb{E}^{\mathbb{P}}[D_T^2/D_1^2 \mid \mathcal{F}_1] = 1/D_1^2$ and that assume $\mathbb{E}^{\mathbb{P}}[1/D_1^2] = \infty$. Then $C_1^{\mathbb{P}} = 1$ but $C_2^{\mathbb{P}} = \infty$. Conversely, assume that $D_T = D_1$ and $\mathbb{E}^{\mathbb{P}}[D_1^2] = \infty$. Then $C_2^{\mathbb{P}} = 1$ but $C_1^{\mathbb{P}} = \infty$.

$\mathbb{E}^{\mathbb{Q}}[Z] = 0$.² The discounted one-year loss is given by

$$L = \mathbb{E}^{\mathbb{Q}}[Z \mid \mathcal{F}_1].$$

Solvency capital calculation boils down to compute the value at risk or expected shortfall,

$$K = \rho[L].$$

The \mathbb{P} -distribution of L is not known in closed form in general and has to be computed by simulating Z and its conditional \mathbb{Q} -expectation L . But the simulation of Z is costly in practice for several reasons. The time horizon T can be rather large of the order of 40 years and more. The liability cash flows are strongly path-dependent because of embedded options such as minimum rate guarantees, management and regulatory rules such as policyholder participation, and policyholder behaviour such as lapsing. Moreover, the number of economic factors underlying the liability cash flows is large and can be of the order of thousands. This renders nested simulation or least-square Monte Carlo methods infeasible.

The RP approach builds on m financial instruments, in addition to the numeraire, with discounted gains processes

$$\mathbf{G}_t = (G_{1t}, \dots, G_{mt})^\top$$

that can be efficiently simulated. Absence of arbitrage requires that \mathbf{G}_t is a \mathbb{Q} -martingale. We aim at approximating Z , and thus L , by a portfolio invested in the financial instruments \mathbf{G}_t . There to we consider dynamic portfolio strategies of the following kind. Fix a time partition $0 = t_0 < t_1 < \dots < t_N = T$, containing $t_j = 1$ for some j , and write

$$\Delta G_{ij} = G_{it_j} - G_{it_{j-1}}$$

for the gain of instrument i on the interval $(t_{j-1}, t_j]$, for $j = 1 \dots N$. Our portfolio strategies are multilinear in the running gains ΔG_{ij} ; that is, linear in the running product of such gains. In stochastic analysis this is known as chaos expansion and is formalized as follows. Let \mathcal{P} be a family of pairs $(\mathcal{I}, \mathcal{J})$ where \mathcal{J} is a subset of $\{1, \dots, N\}$ and $\mathcal{I} : \mathcal{J} \rightarrow \{1, \dots, m\}$ is a mapping. For any $(\mathcal{I}, \mathcal{J}) \in \mathcal{P}$ we define the corresponding product of gains

$$\Delta \mathbf{G}_{(\mathcal{I}, \mathcal{J})} = \prod_{j \in \mathcal{J}} \Delta G_{\mathcal{I}(j)j},$$

and we assume that $\Delta \mathbf{G}_{(\mathcal{I}, \mathcal{J})} \in L^2(\mathbb{M})$. The \mathbb{Q} -martingale property of \mathbf{G}_t then implies

$$\mathbb{E}^{\mathbb{Q}}[\Delta \mathbf{G}_{(\mathcal{I}, \mathcal{J})} \mid \mathcal{F}_{t_j}] = 0 \quad \text{for all } j < \min \mathcal{J}. \quad (3)$$

Any choice of deterministic coefficients $\phi = \{\phi_{(\mathcal{I}, \mathcal{J})} \mid (\mathcal{I}, \mathcal{J}) \in \mathcal{P}\} \in \mathbb{R}^{|\mathcal{P}|}$ and initial wealth v gives rise to a self-financing portfolio with value process

$$V_t^{v, \phi} = v + \sum_{(\mathcal{I}, \mathcal{J}) \in \mathcal{P} \mid t_{\max \mathcal{J}} \leq t} \phi_{(\mathcal{I}, \mathcal{J})} \Delta \mathbf{G}_{(\mathcal{I}, \mathcal{J})}.$$

²The computation of the initial asset-liability value is not the subject of this paper. But it could be estimated by the same methods.

Property (3) implies that $V_t^{v,\phi}$ is a \mathbb{Q} -martingale. Note that, while the coefficients ϕ are deterministic and exogenous, the positions in the instruments \mathbf{G}_t can be path-dependent. Specifically, the gain $\phi_{(\mathcal{I},\mathcal{J})}\Delta\mathbf{G}_{(\mathcal{I},\mathcal{J})}$ results from holding $\phi_{(\mathcal{I},\mathcal{J})} \times \prod_{j \in \mathcal{J} \setminus \{\max \mathcal{J}\}} \Delta G_{\mathcal{I}(j)j}$ units of instrument $\mathcal{I}(\max \mathcal{J})$ during the interval $(t_{\max \mathcal{J}-1}, t_{\max \mathcal{J}}]$.

The following two examples clarify important special cases.

Example 3.1. Assuming that $|\mathcal{J}| = 1$ for all $(\mathcal{I}, \mathcal{J}) \in \mathcal{P}$, we obtain the first-order portfolio with value process

$$V_t^{v,\phi} = v + \sum_{t_j \leq t} \sum_{i=1}^m \phi_{ij} \Delta G_{ij}$$

for the components $\phi_{ij} = \phi_{(\mathcal{I},\mathcal{J})}$ where $\mathcal{J} = \{j\}$ and $\mathcal{I}(j) = i$. This example is further illustrated in Section 5.1.

Example 3.2. Assuming only one risky asset $m = 1$, all formulas above simplify as we can omit the trivial mapping $\mathcal{I}(j) \equiv 1$ and write $\phi_{\mathcal{J}}^{\mathcal{I}} = \phi_{\mathcal{J}}$. Hence \mathcal{P} becomes a family of subsets \mathcal{J} of $\{1, \dots, N\}$ and we write

$$\Delta \mathbf{G}_{\mathcal{J}} = \prod_{j \in \mathcal{J}} \Delta G_j$$

and

$$V_t^{v,\phi} = v + \sum_{\mathcal{J} \in \mathcal{P} | t_{\max \mathcal{J}} \leq t} \phi_{\mathcal{J}} \Delta \mathbf{G}_{\mathcal{J}}. \quad (4)$$

This example is further illustrated in Section 5.2.

For further use, we write

$$\mathbf{A} = (\Delta \mathbf{G}_{(\mathcal{I},\mathcal{J})} | t_{\max \mathcal{J}} \leq 1)^\top, \quad \mathbf{B} = (\Delta \mathbf{G}_{(\mathcal{I},\mathcal{J})} | t_{\max \mathcal{J}} > 1)^\top$$

for the random vectors of portfolio gains up to one year and after one year, respectively. Accordingly, we decompose the portfolio coefficient vector $\phi^\top = (\phi_A^\top, \phi_B^\top)$. With this convention, the portfolio values of $V_t^{v,\phi}$ at $t = 0, 1, T$ become simple expressions,

$$V_0^{v,\phi} = v, \quad V_1^{v,\phi} = v + \phi_A^\top \mathbf{A}, \quad V_T^{v,\phi} = v + \phi_A^\top \mathbf{A} + \phi_B^\top \mathbf{B}.$$

Lemma 2.1 and (1)–(2) suggest to choose a strategy (v, ϕ) which solves the $L^2(\mathbb{M})$ -norm minimization problem

$$\min_{(v,\phi) \in \mathbb{R}^{1+|\mathcal{P}|}} \left\| Z - V_T^{v,\phi} \right\|_{L^2(\mathbb{M})}. \quad (5)$$

The respective self-financing portfolio $V_t^{v,\phi}$ is called *replicating portfolio (RP)*, though the replication of Z is not perfect in general.

Remark 3.3. *The family of instruments \mathbf{G}_t is said to have the chaotic representation property (CRP) if the terminal values of all self-financing portfolios $\{V_T^{v,\phi}\}$ form a basis of $L^2(\mathbb{Q})$. The replication of Z would thus be perfect. However, the CRP is a strong property that is not shared by all martingales, see e.g. [44] for a discrete-time approach. The prototype of a martingale with the CRP in continuous time is the Brownian motion, see e.g. [45] and Section 5.2 below.*

Geometrically speaking, the terminal value of the RP $V_T^{v,\phi}$ is the $L^2(\mathbb{M})$ -projection of Z onto $\{1, \mathbf{A}, \mathbf{B}\}$. For $\mathbb{M} = \mathbb{P}$ the formal solution of (5) is given by the general closed-form expression

$$\begin{pmatrix} v^{\mathbb{P}} \\ \phi_A^{\mathbb{P}} \\ \phi_B^{\mathbb{P}} \end{pmatrix} = \mathcal{M}^{-1} \mathbb{E}^{\mathbb{P}} \left[\begin{pmatrix} Z \\ \mathbf{A}Z \\ \mathbf{B}Z \end{pmatrix} \right] \quad (6)$$

with Gram matrix

$$\mathcal{M} = \mathbb{E}^{\mathbb{P}} \left[\begin{pmatrix} 1 & \mathbf{A}^\top & \mathbf{B}^\top \\ \mathbf{A} & \mathbf{A}\mathbf{A}^\top & \mathbf{A}\mathbf{B}^\top \\ \mathbf{B} & \mathbf{B}\mathbf{A}^\top & \mathbf{B}\mathbf{B}^\top \end{pmatrix} \right].$$

For $\mathbb{M} = \mathbb{Q}$ the above expressions simplify. In view of the martingale property (3) we have $\mathbb{E}^{\mathbb{Q}}[\mathbf{A}] = 0$, $\mathbb{E}^{\mathbb{Q}}[\mathbf{B}] = 0$, and $\mathbb{E}^{\mathbb{Q}}[\mathbf{A}\mathbf{B}^\top] = 0$. The solution of (5) is thus given by

$$v^{\mathbb{Q}} = 0, \quad \begin{pmatrix} \phi_A^{\mathbb{Q}} \\ \phi_B^{\mathbb{Q}} \end{pmatrix} = \mathcal{N}^{-1} \mathbb{E}^{\mathbb{Q}} \left[\begin{pmatrix} \mathbf{A}Z \\ \mathbf{B}Z \end{pmatrix} \right] \quad (7)$$

with block-diagonal reduced Gram matrix

$$\mathcal{N} = \mathbb{E}^{\mathbb{Q}} \left[\begin{pmatrix} \mathbf{A}\mathbf{A}^\top & 0 \\ 0 & \mathbf{B}\mathbf{B}^\top \end{pmatrix} \right].$$

In practice, for large dimensions, the Gram matrices \mathcal{M} and \mathcal{N} can be close to singular due to possible strong correlation between the instruments \mathbf{G}_t . This can cause numerical problems for their inverse. Instruments for which there exist closed-form expressions for \mathcal{M} and \mathcal{N} are thus a preferred choice. Fortunately, there is a large class of so-called polynomial models in finance that have this feature, see [50].

Denote by

$$\epsilon^{\mathbb{M}} = Z - v^{\mathbb{M}} - \phi_A^{\mathbb{M}\top} \mathbf{A} - \phi_B^{\mathbb{M}\top} \mathbf{B}$$

the residual from the $L^2(\mathbb{M})$ -projection. Then we can represent the one-year loss as

$$L = v^{\mathbb{M}} + \phi_A^{\mathbb{M}\top} \mathbf{A} + \mathbb{E}^{\mathbb{Q}} [\epsilon^{\mathbb{M}} | \mathcal{F}_1].$$

As the computation of $\mathbb{E}^{\mathbb{Q}}[\epsilon^{\mathbb{M}} | \mathcal{F}_1]$ is infeasible for the same reasons as the direct computation of L , we now study the following two proxies for L ,

$$\begin{aligned} L_1^{\mathbb{M}} &= v^{\mathbb{M}} + \phi_A^{\mathbb{M}\top} \mathbf{A}, \\ L_2^{\mathbb{M}} &= v^{\mathbb{M}} + \phi_A^{\mathbb{M}\top} \mathbf{A} + \epsilon^{\mathbb{M}} = Z - \phi_B^{\mathbb{M}\top} \mathbf{B}. \end{aligned}$$

Accordingly, we approximate the capital requirement K by

$$\begin{aligned} K_1^{\mathbb{M}} &= \rho[L_1^{\mathbb{M}}] = v^{\mathbb{M}} + \rho[\phi_A^{\mathbb{M}\top} \mathbf{A}], \\ K_2^{\mathbb{M}} &= \rho[L_2^{\mathbb{M}}] = \rho[Z - \phi_B^{\mathbb{M}\top} \mathbf{B}]. \end{aligned}$$

Unless $\mathbb{E}^{\mathbb{Q}}[\epsilon^{\mathbb{M}} | \mathcal{F}_1]$ and L are negatively correlated in the tail of L , the proxy $K_1^{\mathbb{M}}$ will tend to underestimate K , while $K_2^{\mathbb{M}}$ is likely to overshoot. Note that $L_2^{\mathbb{M}}$ is a good proxy for L under any of the conditions stated in the following lemma, whose proof is elementary and thus omitted.

Lemma 3.4. *The following are equivalent:*

- (i) $\mathbb{E}^{\mathbb{Q}}[\epsilon^{\mathbb{M}} \mid \mathcal{F}_1] = \epsilon^{\mathbb{M}}$;
- (ii) $\epsilon^{\mathbb{M}}$ is \mathcal{F}_1 -measurable;
- (iii) cash flows beyond $t = 1$ are spanned by the instruments \mathbf{G}_t ,

$$Z - L = \phi_B^{\mathbb{M}\top} \mathbf{B}.$$

Our first main result provides an upper bound on the $L^1(\mathbb{P})$ -approximation error for either proxy.

Theorem 3.5. *The $L^1(\mathbb{P})$ -approximation errors are bounded by*

$$\|L - L_1^{\mathbb{M}}\|_{L^1(\mathbb{P})} \leq C_2^{\mathbb{M}} \|\epsilon^{\mathbb{M}}\|_{L^2(\mathbb{M})} \quad (8)$$

and

$$\begin{aligned} \|L - L_2^{\mathbb{M}}\|_{L^1(\mathbb{P})} &\leq \|\mathbb{E}^{\mathbb{Q}}[\epsilon^{\mathbb{M}} \mid \mathcal{F}_1] - \epsilon^{\mathbb{M}}\|_{L^1(\mathbb{P})} \\ &\leq \begin{cases} (C_1^{\mathbb{P}} + C_2^{\mathbb{P}}) \|\epsilon^{\mathbb{M}}\|_{L^2(\mathbb{P})}, \\ C_1^{\mathbb{Q}} \|\mathbb{E}^{\mathbb{Q}}[\epsilon^{\mathbb{M}} \mid \mathcal{F}_1] - \epsilon^{\mathbb{M}}\|_{L^2(\mathbb{Q})}, \end{cases} \end{aligned} \quad (9)$$

where the last factor is bounded by $\|\epsilon^{\mathbb{M}}\|_{L^2(\mathbb{Q})}$.

As a meta corollary of Theorem 3.5 we conclude that the RP approach to capital calculation, based on either $L_i^{\mathbb{M}}$, for $i = 1, 2$, $\mathbb{M} = \mathbb{P}, \mathbb{Q}$, and $\rho = \text{VaR}_\alpha, \text{ES}_\alpha$, is asymptotically consistent as we increase the number $|\mathcal{P}|$ of factors such that, asymptotically, $\{1, \mathbf{A}, \mathbf{B}\}$ forms a basis of $L^2(\mathbb{M})$. Indeed, if the residual $\epsilon^{\mathbb{M}} \rightarrow 0$ in $L^2(\mathbb{M})$ then the proxies $L_i^{\mathbb{M}} \rightarrow L$ in $L^1(\mathbb{P})$. In view of (1) and (2), this implies that the *capital approximation error* converges to zero,

$$K_i^{\mathbb{M}} \rightarrow K,$$

where for value at risk, $\rho = \text{VaR}_\alpha$, we have to assume that $q_\alpha^- [L] = q_\alpha^+ [L]$. Note that increasing the number $|\mathcal{P}|$ of factors can be through increasing the number m of instruments \mathbf{G}_t (see Section 5.1) or the number N of time steps and/or the degree of path-dependence $|\mathcal{J}|$ (see Section 5.2)

For expected shortfall, $\rho = \text{ES}_\alpha$, we can combine (8)–(9) with the Lipschitz property (2) to obtain upper bounds on the capital approximation errors $|K - K_i^{\mathbb{M}}|$. However, while asymptotically powerful, these bounds are not sharp. A more useful upper range of the true capital requirement is given by the following result.

Lemma 3.6. *For expected shortfall, $\rho = \text{ES}_\alpha$, the capital requirement K is dominated by*

$$K \leq \text{ES}_\alpha [ZD_T/D_1], \quad (10)$$

$$K \leq K_1^{\mathbb{M}} + \text{ES}_\alpha [\epsilon^{\mathbb{M}} D_T/D_1] \quad (11)$$

and we have

$$K_2^{\mathbb{M}} \leq K_1^{\mathbb{M}} + \text{ES}_\alpha [\epsilon^{\mathbb{M}}].$$

Note that the upper bounds in Lemma 3.6 do not hold for value at risk, $\rho = \text{VaR}_\alpha$, because it is not convex as the following example shows.

Example 3.7. Let $D_T = 1$ such that $\mathbb{P} = \mathbb{Q}$ and \mathcal{F}_1 be the trivial σ -algebra. Assume $\mathbb{P}[Z = 0] = \alpha$ and $\mathbb{P}[Z = 1] = 1 - \alpha$, and hence $L = \mathbb{E}^\mathbb{P}[Z] = 1 - \alpha$. Then $K = \text{VaR}_\alpha[L] = 1 - \alpha > 0 = \text{VaR}_\alpha[Z]$.

The following example contrasts our approach with the current industry standard.

Example 3.8. Industry standard is a *static* first-order RP based on Example 3.1. This nests formally in our framework by setting $N = 2$, $t_1 = 1$, $\mathbf{A} = \mathbf{G}_1 - \mathbf{G}_0$, $\mathbf{B} = \mathbf{G}_T - \mathbf{G}_1$, and $\phi_A = \phi_B$. The constrained $L^2(\mathbb{M})$ -norm minimization problem (5) can be rewritten as the unconstrained $L^2(\mathbb{M})$ -norm minimization problem

$$\min_{(v, \psi) \in \mathbb{R}^{1+m}} \left\| Z - v - \psi^\top (\mathbf{A} + \mathbf{B}) \right\|_{L^2(\mathbb{M})}$$

and we set $\phi_A = \phi_B = \psi$. For $\mathbb{M} = \mathbb{P}$ the formal solution is given by the general closed-form expression

$$\begin{pmatrix} \tilde{v}^\mathbb{P} \\ \tilde{\phi}_A^\mathbb{P} \end{pmatrix} = \tilde{\mathcal{M}}^{-1} \mathbb{E}^\mathbb{P} \left[\begin{pmatrix} Z \\ (\mathbf{A} + \mathbf{B})Z \end{pmatrix} \right], \quad \tilde{\phi}_B^\mathbb{P} = \tilde{\phi}_A^\mathbb{P},$$

with Gram matrix

$$\tilde{\mathcal{M}} = \mathbb{E}^\mathbb{P} \left[\begin{pmatrix} 1 & (\mathbf{A} + \mathbf{B})^\top \\ (\mathbf{A} + \mathbf{B}) & (\mathbf{A} + \mathbf{B})(\mathbf{A} + \mathbf{B})^\top \end{pmatrix} \right].$$

For $\mathbb{M} = \mathbb{Q}$ the formal solution is given by

$$\tilde{v}^\mathbb{Q} = 0, \quad \tilde{\phi}_A^\mathbb{Q} = \tilde{\phi}_B^\mathbb{Q} = \tilde{\mathcal{N}}^{-1} \mathbb{E}^\mathbb{Q} [(\mathbf{A} + \mathbf{B})Z]$$

with reduced Gram matrix

$$\tilde{\mathcal{N}} = \mathbb{E}^\mathbb{Q} [(\mathbf{A} + \mathbf{B})(\mathbf{A} + \mathbf{B})^\top].$$

It is understood that our dynamic and path-dependent RP approach should outperform the static industry approach. The main improvement comes from the fact that we can easily generate a large number $|\mathcal{P}|$ of factors based on a given family of m financial instruments by increasing the number N of time steps and/or the degree of path-dependence $|\mathcal{J}|$. Numerical illustrations are given in Sections 5.1 and 5.2.

Remark 3.9. *While we have developed the RP theory for both projection measures $\mathbb{M} = \mathbb{P}$ and $\mathbb{M} = \mathbb{Q}$, there are several reasons to stick to $\mathbb{M} = \mathbb{Q}$ in practice. First, the RP expressions simplify and we do not need to estimate $v^\mathbb{Q} = 0$. Second, simulating \mathbb{Q} -dynamics of the financial instruments \mathbf{G}_t on a large time horizon T seems preferable to simulating \mathbb{P} -dynamics. The former are martingales, while the latter are subject to possible misspecification of the risk premiums, which may propagate and accumulate to large model errors over long time horizons. The specification of insurance risks under \mathbb{Q} is not problematic, because one can usually assume that insurance risks and the density process D_t are independent. Third, along with Z and \mathbf{G}_t one can always simulate the martingale $1/D_t$ under \mathbb{Q} and obtain samples of the Radon–Nikodym likelihood ratio $d\mathbb{P}/d\mathbb{Q} = 1/D_T$ in order to quantify real-world likelihoods of the risk-neutral scenarios.*

We have also tested other projection norms, different from $L^2(\mathbb{M})$, in (5). This included $L^1(\mathbb{M})$ projections. We did not find that such would improve the approximation errors. For the sake of analytical tractability we thus stick to $L^2(\mathbb{M})$.

4 Monte-Carlo analysis

In practice we have to solve the $L^2(\mathbb{M})$ -norm minimization problem (5) by empirical regression. We thus henceforth assume that $\mathbf{A}Z, \mathbf{B}Z \in L^2(\mathbb{M})$. Simulate n i.i.d. copies

$$\left(\mathbf{A}^{(j)}, \mathbf{B}^{(j)}, Z^{(j)} \right), \quad j = 1 \dots n, \quad (12)$$

of $(\mathbf{A}, \mathbf{B}, Z)$ under \mathbb{M} . We denote by \mathcal{G} the σ -algebra generated by the sample (12).³ For $\mathbb{M} = \mathbb{P}$ in view of (6) we obtain unbiased estimators

$$\begin{pmatrix} \widehat{v}^{\mathbb{P}} \\ \widehat{\phi}_A^{\mathbb{P}} \\ \widehat{\phi}_B^{\mathbb{P}} \end{pmatrix} = \mathcal{M}^{-1} \frac{1}{n} \sum_{j=1}^n \begin{pmatrix} Z^{(j)} \\ \mathbf{A}^{(j)} Z^{(j)} \\ \mathbf{B}^{(j)} Z^{(j)} \end{pmatrix}$$

of the strategy $(v^{\mathbb{P}}, \phi_A^{\mathbb{P}}, \phi_B^{\mathbb{P}})$. The LLN implies that

$$\left(\widehat{v}^{\mathbb{P}}, \widehat{\phi}_A^{\mathbb{P}}, \widehat{\phi}_B^{\mathbb{P}} \right) \rightarrow \left(v^{\mathbb{P}}, \phi_A^{\mathbb{P}}, \phi_B^{\mathbb{P}} \right) \text{ a.s. as } n \rightarrow \infty. \quad (13)$$

The CLT implies that, asymptotically for large n , $\left(\widehat{v}^{\mathbb{P}}, \widehat{\phi}_A^{\mathbb{P}}, \widehat{\phi}_B^{\mathbb{P}} \right)$ is normal with mean $(v^{\mathbb{P}}, \phi_A^{\mathbb{P}}, \phi_B^{\mathbb{P}})$ and covariance given by

$$\text{cov}^{\mathbb{P}} \left[\begin{pmatrix} \widehat{v}^{\mathbb{P}} \\ \widehat{\phi}_A^{\mathbb{P}} \\ \widehat{\phi}_B^{\mathbb{P}} \end{pmatrix} \right] = \frac{1}{n} \mathcal{M}^{-1} \text{cov}^{\mathbb{P}} \left[\begin{pmatrix} Z \\ \mathbf{A}Z \\ \mathbf{B}Z \end{pmatrix} \right] \mathcal{M}^{-1} = \frac{1}{n} \mathbf{C}^{\mathbb{P}}. \quad (14)$$

For $\mathbb{M} = \mathbb{Q}$ the above expressions simplify. In view of (7) we obtain unbiased estimators

$$\widehat{v}^{\mathbb{Q}} = 0, \quad \begin{pmatrix} \widehat{\phi}_A^{\mathbb{Q}} \\ \widehat{\phi}_B^{\mathbb{Q}} \end{pmatrix} = \mathcal{N}^{-1} \frac{1}{n} \sum_{j=1}^n \begin{pmatrix} \mathbf{A}^{(j)} Z^{(j)} \\ \mathbf{B}^{(j)} Z^{(j)} \end{pmatrix}$$

of the strategy $v^{\mathbb{Q}} = 0$ and $(\phi_A^{\mathbb{Q}}, \phi_B^{\mathbb{Q}})$. The LLN implies that

$$\left(\widehat{\phi}_A^{\mathbb{Q}}, \widehat{\phi}_B^{\mathbb{Q}} \right) \rightarrow \left(\phi_A^{\mathbb{Q}}, \phi_B^{\mathbb{Q}} \right) \text{ a.s. as } n \rightarrow \infty. \quad (15)$$

The CLT implies that, asymptotically for large n , $\left(\widehat{\phi}_A^{\mathbb{Q}}, \widehat{\phi}_B^{\mathbb{Q}} \right)$ is normal with mean $(\phi_A^{\mathbb{Q}}, \phi_B^{\mathbb{Q}})$ and covariance given by

$$\text{cov}^{\mathbb{Q}} \left[\begin{pmatrix} \widehat{\phi}_A^{\mathbb{Q}} \\ \widehat{\phi}_B^{\mathbb{Q}} \end{pmatrix} \right] = \frac{1}{n} \mathcal{N}^{-1} \text{cov}^{\mathbb{Q}} \left[\begin{pmatrix} \mathbf{A}Z \\ \mathbf{B}Z \end{pmatrix} \right] \mathcal{N}^{-1} = \frac{1}{n} \mathbf{C}^{\mathbb{Q}}. \quad (16)$$

³Formally, we assume that primary and sample random variables $\mathbf{A}(\omega) = \mathbf{A}(\omega_1)$ and $\mathbf{A}^{(j)}(\omega) = \mathbf{A}^{(j)}(\omega_2)$, etc. with $\omega = (\omega_1, \omega_2)$ are modeled on a product space $\Omega = \Omega' \times \Omega'$, $\mathcal{F} = \mathcal{F}' \otimes \mathcal{F}'$ equipped with product probability measures $\mathbb{M} = \mathbb{M}' \otimes \mathbb{M}'$.

We obtain estimators of the solvency capital proxies $K_i^{\mathbb{M}}$,

$$\begin{aligned}\widehat{K}_1^{\mathbb{M}} &= \widehat{v}^{\mathbb{M}} + \rho \left[\widehat{\phi}_A^{\mathbb{M}\top} \mathbf{A} \mid \mathcal{G} \right], \\ \widehat{K}_2^{\mathbb{M}} &= \rho \left[Z - \widehat{\phi}_B^{\mathbb{M}\top} \mathbf{B} \mid \mathcal{G} \right],\end{aligned}\tag{17}$$

where $\rho[\cdot \mid \mathcal{G}]$ denotes the conditional value at risk or expected shortfall given \mathcal{G} generated by the sample (12). Our second main result shows that the Monte-Carlo estimate is asymptotically consistent as the sample size n tends to infinity.

Theorem 4.1. *We have $\widehat{K}_i^{\mathbb{M}} \rightarrow K_i^{\mathbb{M}}$ a.s. as $n \rightarrow \infty$, where for value at risk, $\rho = \text{VaR}_\alpha$, we assume that $q_\alpha^- [L_i^{\mathbb{M}}] = q_\alpha^+ [L_i^{\mathbb{M}}]$.*

For expected shortfall, $\rho = \text{ES}_\alpha$, we can further quantify this result. The *total capital estimation error* amounts to

$$\left\| K - \widehat{K}_i^{\mathbb{M}} \right\|_{L^2(\mathbb{M})} \leq \left| K - K_i^{\mathbb{M}} \right| + \left\| K_i^{\mathbb{M}} - \widehat{K}_i^{\mathbb{M}} \right\|_{L^2(\mathbb{M})}.\tag{18}$$

The first term on the right hand side of (18) is the approximation error, which can be bounded by Theorem 3.5 in conjunction with (2). The second term on the right hand side of (18) is the *Monte-Carlo error*. Here is an upper bound.

Theorem 4.2. *For expected shortfall, $\rho = \text{ES}_\alpha$, asymptotically for large n , the Monte-Carlo error is bounded by*

$$\left\| K_i^{\mathbb{M}} - \widehat{K}_i^{\mathbb{M}} \right\|_{L^2(\mathbb{M})} \leq \sqrt{\frac{1}{n}} \text{MCE}_i^{\mathbb{M}}$$

with constants

$$\text{MCE}_i^{\mathbb{M}} = \begin{cases} \sqrt{C_{vv}^{\mathbb{P}} + \frac{1}{1-\alpha} \text{tr}(\mathbf{C}_{\mathbf{A}\mathbf{A}}^{\mathbb{P}} \mathbb{E}^{\mathbb{P}}[\mathbf{A}\mathbf{A}^\top])}, & \text{if } i = 1 \text{ and } \mathbb{M} = \mathbb{P}, \\ \frac{1}{1-\alpha} \sqrt{\text{tr}(\mathbf{C}_{\mathbf{A}\mathbf{A}}^{\mathbb{Q}} \mathbb{E}^{\mathbb{P}}[\mathbf{A}\mathbf{A}^\top])}, & \text{if } i = 1 \text{ and } \mathbb{M} = \mathbb{Q}, \\ \frac{1}{1-\alpha} \sqrt{\text{tr}(\mathbf{C}_{\mathbf{B}\mathbf{B}}^{\mathbb{M}} \mathbb{E}^{\mathbb{P}}[\mathbf{B}\mathbf{B}^\top])}, & \text{if } i = 2 \end{cases}$$

where $\mathbf{C}_{vv}^{\mathbb{P}}$, $\mathbf{C}_{\mathbf{A}\mathbf{A}}^{\mathbb{M}}$, and $\mathbf{C}_{\mathbf{B}\mathbf{B}}^{\mathbb{M}}$ denote the respective diagonal 1×1 -, $|\mathbf{A}| \times |\mathbf{A}|$ -, and $|\mathbf{B}| \times |\mathbf{B}|$ -blocks of matrix $\mathbf{C}^{\mathbb{M}}$.

Remark 4.3. *One would expect that, while the capital approximation error is decreasing in the number $|\mathcal{P}|$ of factors, the Monte-Carlo error is decreasing in the sample size n but increasing in $|\mathcal{P}|$. Hence for a given computer time budget n there is likely an optimal choice of $|\mathcal{P}|$ which minimizes the total capital estimation error. However, in the numerical examples below the Monte-Carlo error is always dominated by the corresponding approximation error. This stipulates that in applications one would prefer more factors over less factors.*

We finally address the computation of value at risk or expected shortfall $\rho[X]$ of a random variable X , such as $X = \phi_A^\top \mathbf{A}$ and $X = Z - \phi_B^\top \mathbf{B}$ in (17) for given $\phi_A = \widehat{\phi}_A^{\mathbb{M}}$ and $\phi_B = \widehat{\phi}_B^{\mathbb{M}}$. Recall that, due to law-invariance of value at risk and expected shortfall, $\rho[X] = \rho[\mu]$ is a function of the \mathbb{P} -distribution μ of X . We state and prove a general consistency result whose application to the computation of the expected shortfall in (17) is straightforward.

Theorem 4.4. *Let X be a random variable satisfying*

$$\begin{cases} q_{\alpha}^{-}[X] = q_{\alpha}^{+}[X], & \text{if } \rho = \text{VaR}_{\alpha}, \\ X \in L^1(\mathbb{P}), & \text{if } \rho = \text{ES}_{\alpha}. \end{cases}$$

Let $(X^{(j)}, d\mathbb{P}/d\mathbb{M}^{(j)})$, $j \geq 1$, be an i.i.d. sequence of random variables with the same \mathbb{M} -law as $(X, d\mathbb{P}/d\mathbb{M})$. For any $n \geq 1$, define the weights

$$w^{(j)} = \frac{d\mathbb{P}/d\mathbb{M}^{(j)}}{\sum_{k=1}^n d\mathbb{P}/d\mathbb{M}^{(k)}} \quad (= 1/n \text{ if } \mathbb{M} = \mathbb{P})$$

and the empirical \mathbb{P} -distribution of $X^{(1)}, \dots, X^{(n)}$,

$$\hat{\mu}_n = \sum_{j=1}^n w^{(j)} \delta_{X^{(j)}}.$$

Then $\rho[\hat{\mu}_n] \rightarrow \rho[X]$ a.s. as $n \rightarrow \infty$.

Theorem 4.4 is well known for $\mathbb{M} = \mathbb{P}$. This case can be proved using methods from [123], see also [1, Proposition 4.1] and [76, Theorem 2.6]. What is less known is the statement for $\mathbb{M} = \mathbb{Q}$ where the empirical \mathbb{P} -distribution of X is sampled under \mathbb{Q} .

Remark 4.5. *While \mathbf{A} and \mathbf{B} can be efficiently simulated such that their \mathbb{P} -distributions can be assumed to be known, this might be less so for Z . In practice, one would reuse the \mathbb{M} -sample (12) and apply Theorem 4.4, assuming that the sample size n is large enough for an accurate estimation of the value at risk or expected shortfall in (17). If $\mathbb{M} = \mathbb{Q}$, this requires an extension of the \mathbb{Q} -sample (12) by the Radon–Nikodym density $d\mathbb{P}/d\mathbb{Q}^{(j)} = 1/D_T^{(j)}$.*

5 Examples

There are two sources for incompleteness of the insurance market under static hedging with the underlying financial instruments. First, there are more factors driving the insurance cash flows than there are traded financial instruments for their replication. Second, insurance liability cash flows are nonlinear functions of the financial instruments. Both effects superpose in practice. In the following two examples, we disentangle these effects. We consider first an economy where the insurance losses are linear functions of a multidimensional Brownian motion and financial instruments form a subset of the driving Brownian motions. In the second example the loss is a nonlinear (exponential) function of a scalar Brownian motion which constitutes the only available financial instrument.

5.1 Arithmetic Brownian motion

Consider a d -dimensional \mathbb{P} -Brownian motion $W_t = (W_{1t}, \dots, W_{dt})^{\top}$ with $d \gg 1$. The market price of risk is a constant $\gamma \in \mathbb{R}^d$, such that

$$D_t = \exp\left(-\gamma^{\top} W_t - \frac{\|\gamma\|^2}{2} t\right)$$

is the Radon–Nikodym density process for the pricing measure \mathbb{Q} . We assume a time partition $t_0 = 0, t_1 = 1, t_2 = T$, and two volatility regimes $\lambda_A, \lambda_B \in \mathbb{R}^d$, such that the one-year and terminal losses are

$$\begin{aligned} L &= \lambda_A^\top (W_1 + \gamma), \\ Z &= L + \lambda_B^\top (W_T - W_1 + \gamma(T - 1)), \end{aligned}$$

respectively.

Now let $m \leq d$. For any vector $x \in \mathbb{R}^d$ we write $x_{[m]} = (x_1, \dots, x_m)^\top$ and $x_{\setminus[m]} = (x_{m+1}, \dots, x_d)^\top$. The financial instruments, next to the numeraire, have gains processes $\mathbf{G}_t = W_{[m]t} + \gamma_{[m]}t$. We consider first-order portfolios as in Example 3.1 such that

$$V_1^{v,\phi} = v + \phi_A^\top \mathbf{A}, \quad V_T^{v,\phi} = v + \phi_A^\top \mathbf{A} + \phi_B^\top \mathbf{B}$$

for $\mathbf{A} = \mathbf{G}_1 - \mathbf{G}_0 = \mathbf{G}_1$ and $\mathbf{B} = \mathbf{G}_T - \mathbf{G}_1$. It is readily seen that the solution of the $L^2(\mathbb{M})$ -projection (5) is given by

$$\begin{aligned} \phi_A^\mathbb{P} &= \phi_A^\mathbb{Q} = \lambda_{A[m]}, \\ \phi_B^\mathbb{P} &= \phi_B^\mathbb{Q} = \lambda_{B[m]}, \\ v^\mathbb{P} &= \mathbb{E}^\mathbb{P}[Z] - \phi_A^\mathbb{P}{}^\top \mathbb{E}^\mathbb{P}[\mathbf{A}] - \phi_B^\mathbb{P}{}^\top \mathbb{E}^\mathbb{P}[\mathbf{B}] \\ &= \lambda_{A \setminus [m]}^\top \gamma_{\setminus [m]} + \lambda_{B \setminus [m]}^\top \gamma_{\setminus [m]}(T - 1), \end{aligned} \tag{19}$$

and $v^\mathbb{Q} = 0$. The proxies become

$$\begin{aligned} L_1^\mathbb{P} &= \lambda_A^\top \gamma + \lambda_{B \setminus [m]}^\top \gamma_{\setminus [m]}(T - 1) + \lambda_{A[m]}^\top W_{[m]1}, \\ L_1^\mathbb{Q} &= \lambda_{A[m]}^\top \gamma_{[m]} + \lambda_{A[m]}^\top W_{[m]1}, \end{aligned}$$

and

$$L_2^\mathbb{P} = L_2^\mathbb{Q} = \lambda_A^\top (W_1 + \gamma) + \lambda_{B \setminus [m]}^\top (W_{\setminus [m]T} - W_{\setminus [m]1} + \gamma_{\setminus [m]}(T - 1)),$$

such that for $m = d$ we obtain perfect replication, $L_i^\mathbb{M} = L$. The \mathbb{P} -distribution of $L_i^\mathbb{M}$ is normal with mean $\mu_i^\mathbb{M}$ and variance $c_i^\mathbb{M}$ where

$$\begin{aligned} \mu_1^\mathbb{P} &= \lambda_A^\top \gamma + \lambda_{B \setminus [m]}^\top \gamma_{\setminus [m]}(T - 1), \\ \mu_1^\mathbb{Q} &= \lambda_{A[m]}^\top \gamma_{[m]}, \\ c_1^\mathbb{P} &= c_1^\mathbb{Q} = \|\lambda_{A[m]}\|^2 \end{aligned}$$

and

$$\begin{aligned} \mu_2^\mathbb{M} &= \mu_1^\mathbb{P}, \\ c_2^\mathbb{M} &= \|\lambda_A\|^2 + \|\lambda_{B \setminus [m]}\|^2 (T - 1). \end{aligned}$$

The solvency capital proxies are then given by the following well know formula for value at risk and expected shortfall of normal random variables:

$$K_i^\mathbb{M} = \rho[L_i^\mathbb{M}] = \mu_i^\mathbb{M} + \sqrt{c_i^\mathbb{M}} \rho[\mathcal{N}(0, 1)]$$

where

$$\rho[\mathcal{N}(0, 1)] = \begin{cases} q_\alpha[\mathcal{N}(0, 1)], & \text{for } \rho = \text{VaR}_\alpha, \\ \frac{1}{1-\alpha} \frac{1}{\sqrt{2\pi}} e^{-q_\alpha[\mathcal{N}(0,1)]^2/2}, & \text{for } \rho = \text{ES}_\alpha. \end{cases} \quad (20)$$

and $q_\alpha[\mathcal{N}(0, 1)]$ is the α -quantile of the standard normal distribution. We have

$$q_{99.5\%}[\mathcal{N}(0, 1)] \approx 2.5758, \quad q_{99\%}[\mathcal{N}(0, 1)] \approx 2.3263, \quad (21)$$

and $\text{ES}_{99\%}[\mathcal{N}(0, 1)] \approx 2.6652$. The true solvency capital equals the proxies, $K = K_i^{\mathbb{M}}$, for $m = d$.

As for the Gram matrices we obtain

$$\mathcal{M} = \begin{pmatrix} 1 & \gamma_{[m]}^\top & (T-1)\gamma_{[m]}^\top \\ \gamma_{[m]} & \text{Id}_m + \gamma_{[m]}\gamma_{[m]}^\top & (T-1)\gamma_{[m]}\gamma_{[m]}^\top \\ (T-1)\gamma_{[m]} & (T-1)\gamma_{[m]}\gamma_{[m]}^\top & (T-1)\text{Id}_m + (T-1)^2\gamma_{[m]}\gamma_{[m]}^\top \end{pmatrix}$$

and

$$\mathcal{N} = \begin{pmatrix} \text{Id}_m & 0 \\ 0 & (T-1)\text{Id}_m \end{pmatrix}.$$

For a numerical implementation we choose the time horizon $T = 5$, number of Brownian motions $d = 5$, market price of risk $\gamma = 0.1 \times \mathbf{1}$, where we write $\mathbf{1} = (1, \dots, 1)^\top$. As risk measure we consider expected shortfall, $\rho = \text{ES}_{99\%}$. Note that λ_A and λ_B are nonpositive because economically we expect a negative risk premium on the liabilities such that the \mathbb{P} -expected loss is negative, $\mathbb{E}^\mathbb{P}[Z] < 0$. We consider two cases:

- (i) $\lambda_A = \lambda_B = -0.2 \times \mathbf{1}/\sqrt{d}$: constant volatility. In this case, in view of (19), the industry standard static proxies from Example 3.8 coincide with our dynamic and path-dependent proxies, $\tilde{L}_i^{\mathbb{M}} = L_i^{\mathbb{M}}$.
- (ii) $\lambda_A = -0.2 \times \mathbf{1}/\sqrt{d}$, $\lambda_B = 0$: no cash flows beyond $t = 1$. In this case $L_2^{\mathbb{P}} = L_2^{\mathbb{Q}} = L$ are exact proxies.

The following capital requirements and proxies are normalised, using the homogeneity of expected shortfall, such that always $K = 1$. Figure 1 shows the capital requirement and proxies as functions of m for case (i). The proxies $K_1^{\mathbb{P}}$ and $K_1^{\mathbb{Q}}$ underestimate the true capital requirement by 80% and 50% for $m = 1$ and converge monotonically to K as m increases to $d = 5$. The proxies $K_2^{\mathbb{P}} = K_2^{\mathbb{Q}}$ overestimate the true capital requirement by 80% for $m = 1$ and converge monotonically to K as m increases to $d = 5$. Figure 2 illustrates case (ii) and reveals that the industry standard static proxies from Example 3.8 are vastly outperformed by our dynamic and path-dependent proxies.

Figure 3 shows the upper bounds (10) and (11) from Lemma 3.6 on the capital requirement as functions of m . In case (i) (left panel) the bound (10) is almost six times greater than K . The bounds (11) for both $\mathbb{M} = \mathbb{P}$ and \mathbb{Q} are tighter but still more than three times greater than K for $m \leq 4$, which has to be compared to the relative approximation errors of 80% and less reported above. In case (ii) (right panel) the bounds are economically more reasonable and useful as their overshoot of K by 80% and less compares to the relative approximation errors shown in Figure 2.

Figures 4 and 5 show the corresponding Monte-Carlo errors, see (18), for our proxies with a sample size of $n = 1000$. They have been computed by repeated sampling both with exact computation of the expected shortfall in (17) either using (20) or a very large independent \mathbb{P} -sample

(top panels) and with reusing the corresponding \mathbb{M} -sample (12) as outlined in Remark 4.5 (bottom panels). The latter Monte-Carlo errors are larger than the former because they are composed of the variances (14) and (16), as described in Theorem 4.2, and the variances implicit in Theorem 4.4. All Monte-Carlo errors are dominated by the corresponding approximation errors shown in Figures 1 and 2, except when the proxies are exact. This stipulates that a larger number m should be preferred over smaller m in applications.

5.2 Geometric Brownian motion

This example works only for the risk-neutral projection measure $\mathbb{M} = \mathbb{Q}$. Consider a scalar \mathbb{P} -Brownian motion W_t . The market price of risk is a constant $\gamma \in \mathbb{R}$, such that

$$D_t = \exp\left(-\gamma W_t - \frac{\gamma^2}{2}t\right)$$

is the Radon–Nikodym density process for the pricing measure \mathbb{Q} . There is one financial instrument, next to the numeraire, with gains process $G_t = W_t + \gamma t$. We define the \mathbb{Q} -martingale

$$M_t = \exp\left(\lambda G_t - \frac{\lambda^2}{2}t\right)$$

and assume that one-year and terminal losses are $L = M_1 - 1$ and $Z = M_T - 1$.

The Wiener chaos expansion theory tells us that M_t can be expanded as orthogonal series in $L^2(\mathbb{Q})$ as

$$M_t = 1 + \sum_{k=1}^{\infty} \int_{0 < s_1 < \dots < s_k \leq t} \lambda^k dG_{s_1} dG_{s_2} \dots dG_{s_k}, \quad (22)$$

see e.g. [104, Exercise 1.1.1]. On the other hand, the following representation holds

$$M_t = 1 + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} t^{k/2} H_k\left(\frac{G_t}{\sqrt{t}}\right),$$

for the Hermite polynomials $H_k(x) = (-1)^k \exp(x^2/2) \frac{d^k}{dx^k} \exp(-x^2/2)$. Note that $t^{k/2} H_k(G_t/\sqrt{t})$ are orthogonal \mathbb{Q} -martingales for all $k \geq 1$.

Comparing the chaos expansion (22) with the expression (4) suggests that in discrete time we have

$$v = 0, \quad \phi_{\mathcal{J}} = \lambda^{|\mathcal{J}|}.$$

Asymptotically, for $N \rightarrow \infty$ with $\max |t_j - t_{j-1}| \rightarrow 0$, we thus obtain

$$L_1^{\mathbb{Q}} = \sum_{k=1}^J \frac{\lambda^k}{k!} H_k(G_1), \quad L_2^{\mathbb{Q}} = L_1^{\mathbb{Q}} + \epsilon^{\mathbb{Q}}$$

with

$$\epsilon^{\mathbb{Q}} = M_T - 1 - \sum_{k=1}^J \frac{\lambda^k}{k!} T^{k/2} H_k\left(\frac{G_T}{\sqrt{T}}\right),$$

for various approximation degrees $J = 1, 2, \dots$, related to the maximal path dependence degree $|\mathcal{J}| \leq J$.

As $\log M_1$ is normal with mean $\lambda\gamma - \frac{\lambda^2}{2}$ and variance λ^2 , we obtain for the solvency capital

$$\begin{aligned} K &= \rho[M_1 - 1] = \rho[M_1] - 1 \\ &= \begin{cases} \exp\left[\lambda\gamma - \frac{\lambda^2}{2} + |\lambda|q_\alpha[\mathcal{N}(0, 1)]\right] - 1, & \text{for } \rho = \text{VaR}_\alpha, \\ \frac{1}{1-\alpha} \exp[\lambda\gamma] \Phi(|\lambda| - q_\alpha[\mathcal{N}(0, 1)]) - 1, & \text{for } \rho = \text{ES}_\alpha, \end{cases} \end{aligned}$$

where $q_\alpha[\mathcal{N}(0, 1)]$ is the α -quantile and Φ the cumulative distribution function of the standard normal distribution, see (21). The capital proxies $K_i^{\mathbb{Q}}$ have to be computed by simulation, see Theorem 4.4.

For a numerical implementation we choose the time horizon $T = 5$, varying approximation degrees $J = 1, \dots, 5$, market price of risk $\gamma = 0.1$, and volatility $\lambda = -0.2$. As in the previous example, λ is negative because economically we expect a negative risk premium on the liabilities such that the \mathbb{P} -expected loss is negative, $\mathbb{E}^{\mathbb{P}}[Z] < 0$. As risk measure we consider expected shortfall, $\rho = \text{ES}_{99\%}$.

The following capital requirement and proxies are normalised, using the homogeneity of expected shortfall, such that $K = 1$. Figure 6 shows the capital requirement and proxies as functions of J . The relative approximation errors for the proxies $K_1^{\mathbb{Q}}$ and $K_2^{\mathbb{Q}}$ are less than 0.5% for $J \geq 2$. The industry standard static proxies from Example 3.8 correspond to $J = 1$. The relative approximation error for $K_1^{\mathbb{Q}}$ is ten times larger for $J = 1$ than for $J \geq 2$.

Figure 7 shows the upper bounds (10) and (11) from Lemma 3.6 on the capital requirement as functions of J . While the bound (10) is 3.5 times larger than K , the bound (11) is much tighter and compares to the relative approximation errors for $J \geq 2$.

These findings suggest that the second and higher order RPs capture the nonlinearities of the insurance liability cash flows significantly better than the first-order industry standard static RP.

6 Conclusion

We provide the mathematical foundation and a dynamic and path-dependent extension of the replicating portfolio (RP) approach to capital calculation for life insurance or other long-term asset-liability portfolios. We show that value at risk and expected shortfall-based capital estimates are asymptotically consistent under real-world and risk-neutral sampling. Two numerical examples illustrate that the dynamic and path-dependent RP outperforms the industry standard static RP. The dynamic and path-dependent extensions could be readily built into existing projection tools used in practice. Our findings also suggest that the approximation error dominates the Monte-Carlo error. This stipulates that more factors, that is, large $|\mathcal{P}|$, should be preferred over few factors in applications. This calls for a real-world study, which is left for future research.

A Proofs

This appendix contains all proofs.

Proof of Lemma 2.1

The first statement follows from the Cauchy–Schwarz inequality. As for the second statement we note that

$$\left\| \mathbb{E}^{\mathbb{Q}}[X \mid \mathcal{F}_1] \right\|_{L^1(\mathbb{P})} \leq \|XD_T/D_1\|_{L^1(\mathbb{P})} = \left\| \sqrt{D_T}X\sqrt{D_T}/D_1 \right\|_{L^1(\mathbb{P})}.$$

Applying the Cauchy–Schwarz inequality to the second and third expression yields the claim for $\mathbb{M} = \mathbb{P}$ and $\mathbb{M} = \mathbb{Q}$, respectively.

Proof of Theorem 3.5

The bound (8) follows from Lemma 2.1. The bound (9) follows similarly, where we write

$$\left\| \mathbb{E}^{\mathbb{Q}}[\epsilon^{\mathbb{M}} \mid \mathcal{F}_1] - \epsilon^{\mathbb{M}} \right\|_{L^1(\mathbb{P})} \leq \left\| \epsilon^{\mathbb{M}} \right\|_{L^1(\mathbb{P})} + \left\| \mathbb{E}^{\mathbb{Q}}[\epsilon^{\mathbb{M}} \mid \mathcal{F}_1] \right\|_{L^1(\mathbb{P})}$$

for the \mathbb{P} -bound.

Proof of Lemma 3.6

We have $K = \text{ES}_\alpha \left[\mathbb{E}^{\mathbb{P}}[ZD_T/D_1 \mid \mathcal{F}_1] \right]$. Monotonicity with respect to stochastic order of expected shortfall yields (10), see [51, Corollary 4.59]. Using subadditivity of expected shortfall the other inequalities follow similarly.

Proof of Theorem 4.1

The theorem follows from combining (1) and (2) with the LLN (13) and (15).

Proof of Theorem 4.2

Assume $i = 1$. Using Lipschitz property (2) of expected shortfall we derive

$$\begin{aligned} \left\| K_1^{\mathbb{M}} - \widehat{K}_1^{\mathbb{M}} \right\|_{L^2(\mathbb{M})} &= \left\| v^{\mathbb{M}} - \widehat{v}^{\mathbb{M}} + \text{ES}_\alpha[\phi_A^{\mathbb{M}\top} \mathbf{A}] - \text{ES}_\alpha \left[\widehat{\phi}_A^{\mathbb{M}\top} \mathbf{A} \mid \mathcal{G} \right] \right\|_{L^2(\mathbb{M})} \\ &\leq \left\| v^{\mathbb{M}} - \widehat{v}^{\mathbb{M}} \right\|_{L^2(\mathbb{M})} + \frac{1}{1-\alpha} \sqrt{\mathbb{E}^{\mathbb{M}} \left[\mathbb{E}^{\mathbb{P}} \left[\left((\phi_A^{\mathbb{M}} - \widehat{\phi}_A^{\mathbb{M}})^\top \mathbf{A} \right)^2 \mid \mathcal{G} \right] \right]} \\ &= \left\| v^{\mathbb{M}} - \widehat{v}^{\mathbb{M}} \right\|_{L^2(\mathbb{M})} + \frac{1}{1-\alpha} \sqrt{\mathbb{E}^{\mathbb{M}} \left[(\phi_A^{\mathbb{M}} - \widehat{\phi}_A^{\mathbb{M}})^\top \mathbb{E}^{\mathbb{P}}[\mathbf{A}\mathbf{A}^\top] (\phi_A^{\mathbb{M}} - \widehat{\phi}_A^{\mathbb{M}}) \right]}. \end{aligned}$$

The CLT (14) and (16) now yields the claim. The case $i = 2$ follows similarly.

Proof of Theorem 4.4

Denote by μ the \mathbb{P} -distribution of X . We claim that

$$\widehat{\mu}_n \rightarrow \mu \text{ weakly a.s.} \tag{23}$$

Let $f(x)$ be a continuous function with compact support. Write

$$\int f(x) d\widehat{\mu}_n = \frac{\frac{1}{n} \sum_{j=1}^n f(X^{(j)}) d\mathbb{P}/d\mathbb{M}^{(j)}}{\frac{1}{n} \sum_{k=1}^n d\mathbb{P}/d\mathbb{M}^{(k)}}.$$

By the law of large numbers the numerator and denominator converge,

$$\frac{1}{n} \sum_{j=1}^n f(X^{(j)}) d\mathbb{P}/d\mathbb{M}^{(j)} \rightarrow \mathbb{E}_{\mathbb{M}} [f(X) d\mathbb{P}/d\mathbb{M}] = \mathbb{E}_{\mathbb{P}} [f(X)] \quad \text{a.s.}$$

and

$$\frac{1}{n} \sum_{k=1}^n d\mathbb{P}/d\mathbb{M}^{(k)} \rightarrow \mathbb{E}_{\mathbb{M}} [d\mathbb{P}/d\mathbb{M}] = 1 \quad \text{a.s.}$$

Hence

$$\int f(x) d\tilde{\mu}_n \rightarrow \int f(x) d\mu \quad \text{a.s.}$$

As by assumption we have $X d\mathbb{P}/d\mathbb{M} \in L^1(\mathbb{M})$ for expected shortfall, $\rho = \text{ES}_\alpha$, the same holds for $f(x) = |x|$,

$$\int |x| d\hat{\mu}_n \rightarrow \int |x| d\mu \quad \text{a.s.} \tag{24}$$

Because the space of continuous functions with compact support is separable, there exists a measurable $\Omega_0 \in \mathcal{F}$ such that $\mathbb{M}[\Omega_0] = 1$ and such that for each $\omega \in \Omega_0$ the desired properties (23) and (24), respectively, hold. The theorem now follows directly from (1) for $\rho = \text{VaR}_\alpha$ and using the arguments as in the proof of [76, Theorem 2.6] for $\rho = \text{ES}_\alpha$.

B Figures

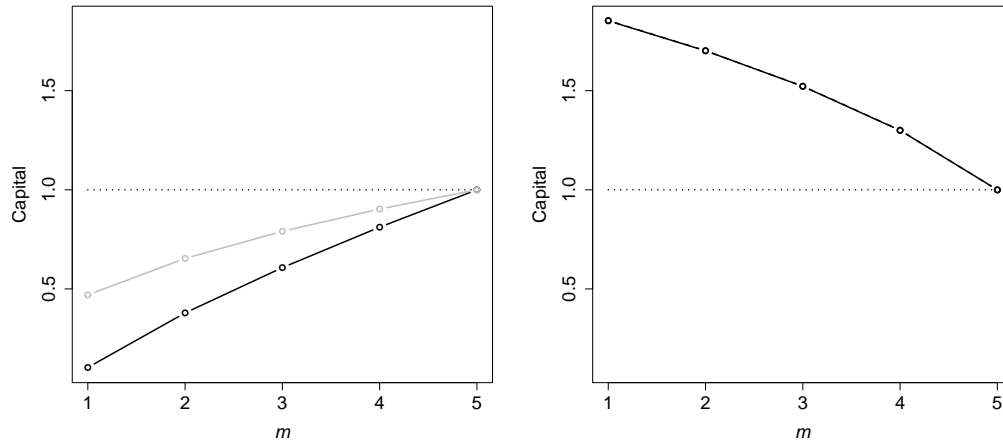


Figure 1: Capital requirement and proxies as functions of number of instruments m for case (i) in the example with arithmetic Brownian motion given in Section 5.1. Risk measure is expected shortfall, $\rho = \text{ES}_{99\%}$. The left panel shows the true capital requirement $K = 1$ (dotted line) and the proxies $K_1^{\mathbb{P}}$ (black line) and $K_1^{\mathbb{Q}}$ (grey line). They coincide with the corresponding industry standard static proxies $\tilde{K}_1^{\mathbb{P}}$ and $\tilde{K}_1^{\mathbb{Q}}$ from Example 3.8. The right panel shows the true capital requirement $K = 1$ (dotted line) and the proxies $K_2^{\mathbb{P}} = K_2^{\mathbb{Q}} = \tilde{K}_2^{\mathbb{P}} = \tilde{K}_2^{\mathbb{Q}}$ (black line).

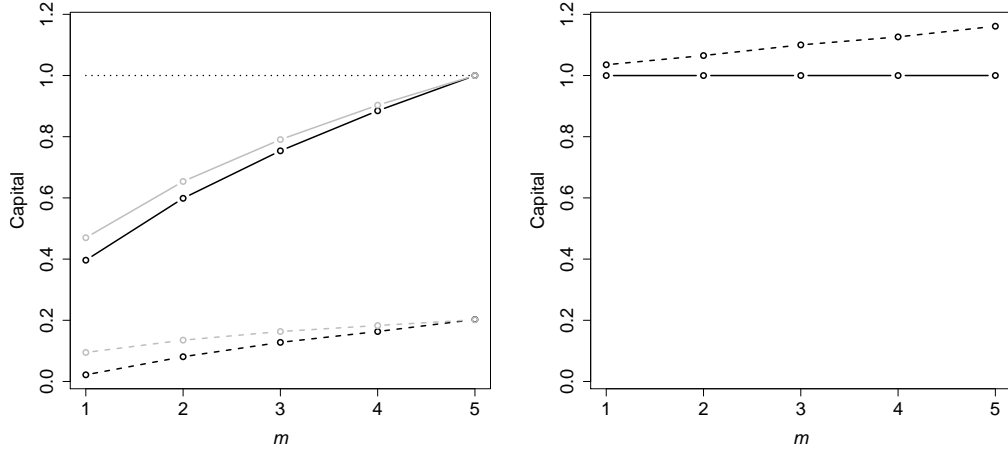


Figure 2: Capital requirement and proxies as functions of number of instruments m for case (ii) in the example with arithmetic Brownian motion given in Section 5.1. Risk measure is expected shortfall, $\rho = \text{ES}_{99\%}$. The left panel shows the true capital requirement $K = 1$ (dotted line), the proxies $K_1^{\mathbb{P}}$ (black solid line) and $K_1^{\mathbb{Q}}$ (grey solid line), and the corresponding industry standard static proxies $\tilde{K}_1^{\mathbb{P}}$ (black dashed line) and $\tilde{K}_1^{\mathbb{Q}}$ (grey dashed line) from Example 3.8. The right panel shows the true capital requirement $K = 1$, which coincides with the proxies $K_2^{\mathbb{P}} = K_2^{\mathbb{Q}}$ (black solid line), and the proxies $\tilde{K}_2^{\mathbb{P}} = \tilde{K}_2^{\mathbb{Q}}$ (black dashed line).

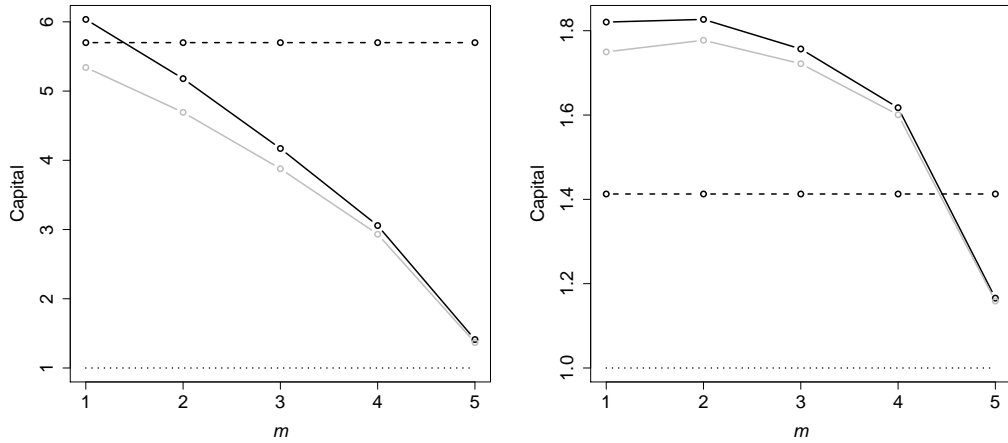


Figure 3: Upper bounds from Lemma 3.6 on the capital requirement as functions of number of instruments m for the example with arithmetic Brownian motion given in Section 5.1. Risk measure is expected shortfall, $\rho = \text{ES}_{99\%}$. The left and right panels correspond to cases (i) and (ii). Dotted lines show the capital requirement $K = 1$, dashed lines show the bound (10), solid lines show the bound (11) for $\mathbb{M} = \mathbb{P}$ (black) and $\mathbb{M} = \mathbb{Q}$ (grey).

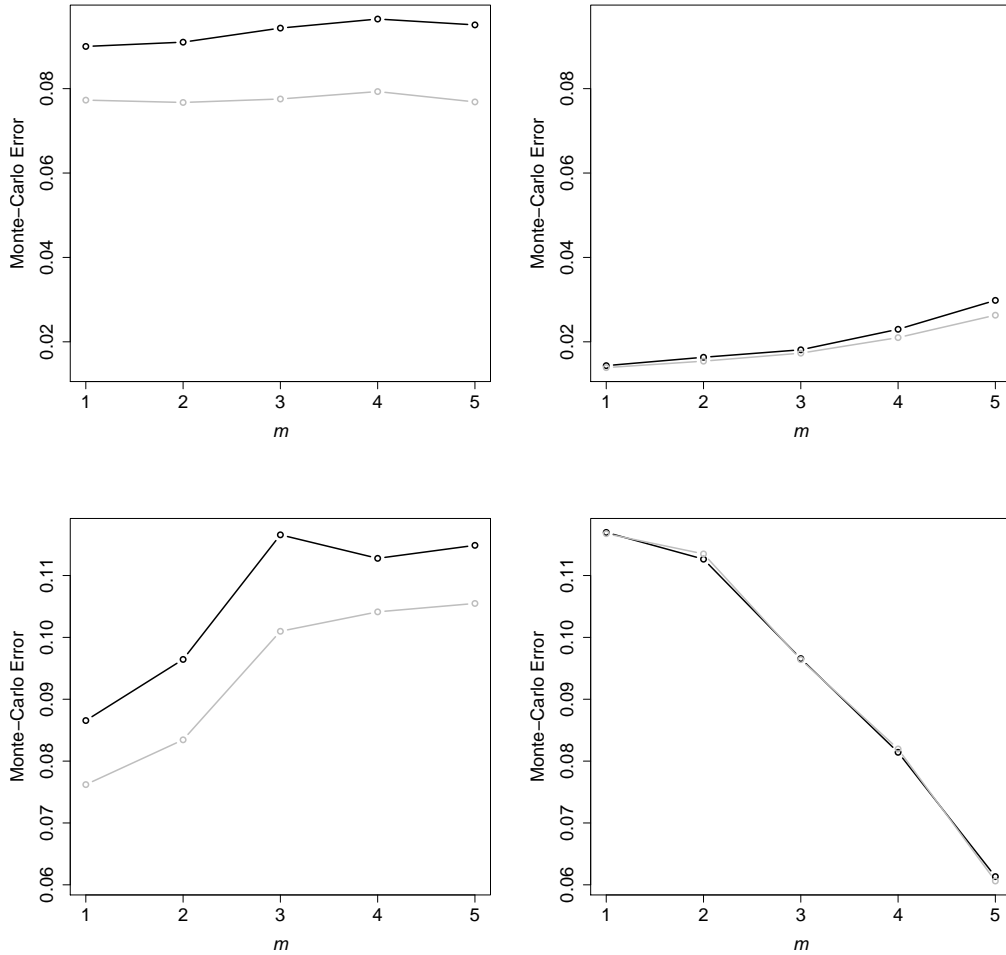


Figure 4: Monte-Carlo errors as functions of number of instruments m for case (i) in the example with arithmetic Brownian motion given in Section 5.1. Risk measure is expected shortfall, $\rho = \text{ES}_{99\%}$. The top panels show the Monte-Carlo errors for $K_1^{\mathbb{P}}$ (left panel, black line), $K_1^{\mathbb{Q}}$ (left panel, grey line), $K_2^{\mathbb{P}}$ (right panel, black line), and $K_2^{\mathbb{Q}}$ (right panel, grey line). The bottom panels show the corresponding Monte-Carlo errors when expected shortfall in (17) is estimated reusing the corresponding M-sample (12) as outlined in Remark 4.5.

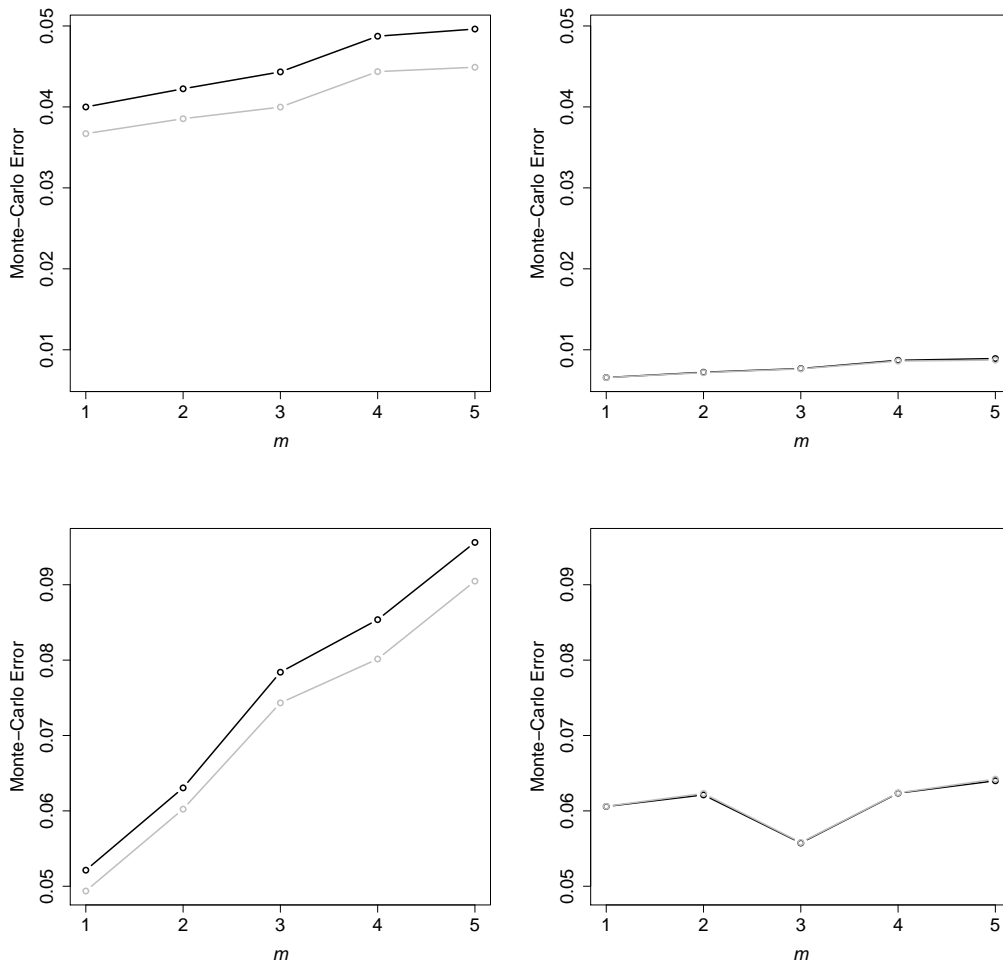


Figure 5: Monte-Carlo errors as functions of number of instruments m for case (ii) in the example with arithmetic Brownian motion given in Section 5.1. Risk measure is expected shortfall, $\rho = \text{ES}_{99\%}$. The top panels show the Monte-Carlo errors for $K_1^{\mathbb{P}}$ (left panel, black line), $K_1^{\mathbb{Q}}$ (left panel, grey line), $K_2^{\mathbb{P}}$ (right panel, black line), and $K_2^{\mathbb{Q}}$ (right panel, grey line). The bottom panels show the corresponding Monte-Carlo errors when expected shortfall in (17) is estimated reusing the corresponding M-sample (12) as outlined in Remark 4.5.

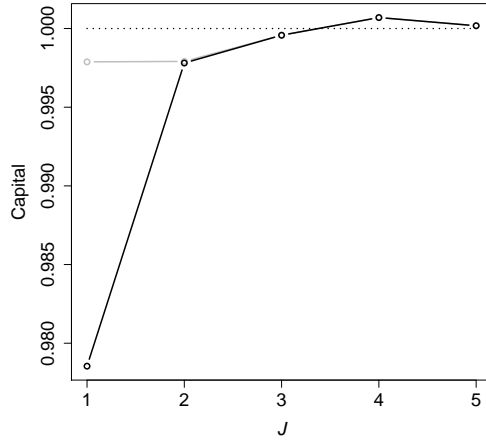


Figure 6: Capital requirement and proxies as functions of approximation degree J for the example with geometric Brownian motion given in Section 5.2. Risk measure is expected shortfall, $\rho = \text{ES}_{99\%}$. Dotted line shows the true capital requirement $K = 1$, solid lines show the proxies $K_1^{\mathbb{Q}}$ (black) and $K_2^{\mathbb{Q}}$ (grey). The industry standard static proxies $\tilde{K}_1^{\mathbb{Q}}$ and $\tilde{K}_2^{\mathbb{Q}}$ from Example 3.8 correspond to $J = 1$.

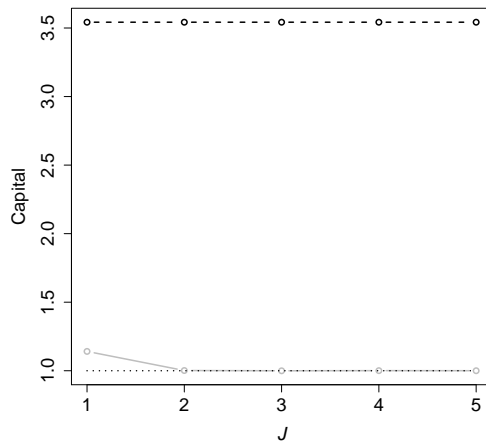


Figure 7: Upper bounds from Lemma 3.6 on the capital requirement as functions of approximation degree J for the example with geometric Brownian motion given in Section 5.2. Risk measure is expected shortfall, $\rho = \text{ES}_{99\%}$. Dotted line shows the capital requirement $K = 1$, dashed line shows the bound (10), grey solid line shows the bound (11) for $\mathbb{M} = \mathbb{Q}$.

Cumulative Bibliography

Cumulative Bibliography

- [1] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002.
- [2] C. Aistleitner and J. Dick. Functions of bounded variation, signed measures, and a general Koksma–Hlawka inequality. *Acta Arithmetica*, 167:143–171, 2015.
- [3] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society. Series B (Methodological)*, 28(1):131–142, 1966.
- [4] P. Arbenz, M. Cambou, and M. Hofert. An importance sampling approach for copula models in insurance. *Submitted*, <https://arxiv.org/abs/1403.4291v3>, 2015.
- [5] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.
- [6] S. Asmussen and P. W. Glynn. *Stochastic simulation: algorithms and analysis*. Stochastic Modelling and Applied Probability. Springer, New York, 2007.
- [7] D. Bauer, D. Bergmann, and R. Kiesel. On the risk-neutral valuation of life insurance contracts with numerical methods in view. *Astin Bulletin*, 40(1):65–95, 2010.
- [8] BCBS. International Convergence of Capital Measurement and Capital Standards, A Revised Framework, 2005.
- [9] BCBS. Guidelines for Computing Capital for Incremental Risk in the Trading Book, 2008.
- [10] M. Bee. Adaptive importance sampling for simulating copula-based distributions. *Insurance: Mathematics and Economics*, 48(2):237–245, 2011.
- [11] J. Berkowitz. A coherent framework for stress-testing. *Journal of Risk*, 2:1–11, 2000.
- [12] E. Beutner, A. Pelsser, and J. Schweizer. Fast convergence of regress-later estimates in least squares Monte Carlo. *SSRN*: <http://ssrn.com/abstract=2328709>, 2013.
- [13] E. Beutner, A. Pelsser, and J. Schweizer. Theory and validation of replicating portfolios in insurance risk management. *SSRN*: <http://ssrn.com/abstract=2557368>, 2016.
- [14] J.-P. Bouchaud, F. Caccioli, and D. Farmer. Impact-adjusted valuation and the criticality of leverage. *Insurance Risk*, 2013.
- [15] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.
- [16] T. Breuer and I. Csizsár. Systematic stress tests with entropic plausibility constraints. *Journal of Banking and Finance*, 37:1552–1559, 2013.
- [17] T. Breuer, M. Jandacka, J. Mencía, and M. Summer. A systematic approach to multi-period stress testing of portfolio credit risk. *Journal of Banking and Finance*, 360:332–340, 2012.

- [18] R. Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, 7:1–49, 1998.
- [19] S. Cambanis, S. Huang, and G. Simons. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385, 1981.
- [20] M. Cambou and D. Filipović. Model uncertainty and scenario aggregation. *Mathematical Finance*, 2015.
- [21] M. Cambou and D. Filipović. Replicating portfolio approach to capital calculation. *Submitted*, 2016.
- [22] M. Cambou, M. Hofert, and C. Lemieux. Quasi-random numbers for copula models. *Statistics and Computing*, pages 1–23, 2016.
- [23] CEIOPS. CEIOPS Advice for Level 2 Implementing Measures on Solvency II: Articles 120 to 126, Tests and Standards for Internal Model Approval . Technical report, Committee of European Insurance and Occupational Pensions Supervisors, 2009.
- [24] CEIOPS. QIS5 Technical Specifications. Technical report, Committee of European Insurance and Occupational Pensions Supervisors, 2010.
- [25] C. P. Chambers. An axiomatization of quantiles on the domain of distribution of distribution functions. *Mathematical Finance*, 19(2):335–342, 2009.
- [26] A. Charpentier and A. Juri. Limiting dependence structures for tail events, with applications to credit derivatives. *Journal of Applied Probability*, 43(2):563–586, 2006.
- [27] G. Constantine and T. Savits. A multivariate Faa Di Bruno formula with applications. *Transactions of the American Mathematical Society*, 348(2):503–520, 1996.
- [28] R. Cont. Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16(3):519–547, 2006.
- [29] R. Cont, R. Deguest, and G. Scandolo. Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 10(6):593–606, 2010.
- [30] R. Cranley and T. Patterson. Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis*, 13(6):904–914, 1976.
- [31] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318, 1967.
- [32] DAV. Proxy-Modelle für die Risikokapitalberechnung. Technical report, Ausschuss Investment der Deutschen Aktuarvereinigung (DAV), 2015.
- [33] A. C. Davison. *Statistical models*, volume 11 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2003.
- [34] L. Devroye. *A course in density estimation*, volume 14 of *Progress in Probability and Statistics*. Birkhäuser Boston Inc., Boston, MA, 1987.

- [35] J. Dick and F. Pillichshammer. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, UK, 2010.
- [36] F. Durante and P. Jaworski. Invariant dependence structure under univariate truncation. *Statistics*, 46(2):263–277, 2012.
- [37] R. Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.
- [38] K. K. Dutta and D. F. Babbel. Scenario analysis in the measurement of operational risk capital: A change of measure approach. *Journal of Risk and Insurance*, 81(2):303–334, 2014.
- [39] P. Embrechts, F. Lindskog, and A. J. McNeil. Modelling dependence with copulas and applications to risk management. In S. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, pages 329–384. Elsevier, 2003.
- [40] P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependence in risk management: Properties and pitfalls. *Risk Management: Value at Risk and Beyond*, pages 176–223, 2002.
- [41] P. Embrechts, G. Puccetti, and L. Rüschendorf. Model Uncertainty and VaR Aggregation. *Journal of Banking and Finance*, 37(8):2750–2764, 2013.
- [42] P. Embrechts, G. Puccetti, L. Rüschendorf, R. Wang, and A. Belaraj. An academic response to basel 3.5. *Risks*, 2(1):25–48, 2014.
- [43] P. Embrechts, B. Wang, and R. Wang. Aggregation-robustness and model uncertainty of regulatory risk measures. *Preprint*, 2014.
- [44] M. Émery. A discrete approach to the chaotic representation property. In *Séminaire de Probabilités, XXXV*, volume 1755 of *Lecture Notes in Math.*, pages 123–138. Springer, Berlin, 2001.
- [45] M. Émery. Chaotic representation property of certain Azéma martingales. *Illinois J. Math.*, 50(1-4):395–411, 2006.
- [46] S. Emmer, M. Kratz, and D. Tasche. What is the best risk measure in practice? a comparison of standard measures. *Journal of Risk*, 18(2):31–60, 2015.
- [47] K. Fang, S. Kotz, and K. Ng. *Symmetric Multivariate and Related Distributions*. Chapman & Hall/CRC, 1990.
- [48] H. Faure. Discrépance des suites associées à un système de numération (en dimension s). *Acta Arithmetica*, 41:337–351, 1982.
- [49] H. Faure and C. Lemieux. Generalized Halton sequence in 2008: A comparative study. *ACM Transactions on Modeling and Computer Simulation*, 19(15), 2009.
- [50] D. Filipović and M. Larsson. Polynomial diffusions and applications in finance. *Finance and Stochastics*, forthcoming, 2016.

- [51] H. Föllmer and A. Schied. *Stochastic finance: An introduction in discrete time*. Walter de Gruyter & Co., Berlin, third edition, 2011.
- [52] FOPI. Technical document on the Swiss Solvency Test, 2006.
- [53] FOPI. Technical document on the Swiss Solvency Test. Technical report, Swiss Federal Office of Private Insurance, 2006.
- [54] C. Genest, J. Neslehová, and N. Ben Ghorbal. Estimators based on Kendall’s tau in multivariate copula models. *Australian & New Zealand Journal of Statistics*, 53(2):157–177, 2011.
- [55] J. Geweke. *Contemporary Bayesian econometrics and statistics*. Wiley Series in Probability and Statistics. Wiley-Interscience John Wiley & Sons, Hoboken, NJ, 2005.
- [56] P. Glasserman and J. Li. Importance sampling for portfolio credit risk. *Management Science*, 51(11):1643–1656, 2005.
- [57] A. Haier and T. Pfeiffer. Scenarios and their aggregation in the regulatory risk measurement environment, 2012.
- [58] J. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2:84–90, 1960.
- [59] F. R. Hampel. A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 42(6):1887–1896, 1971.
- [60] J. Hartinger, R. Kainhofer, and R. Tichy. Quasi-Monte Carlo algorithms for unbounded, weighted integration problems. *Journal of Complexity*, 20:654–668, 2004.
- [61] E. Hlawka. Über die diskrepanz mehrdimensionaler folgen mod 1. *Mathematische Zeitschrift*, 77:273–284, 1961.
- [62] E. Hlawka and R. Mück. Über eine transformation von gleichverteilten folgen ii. *Computing*, 9(2):127–138, 1972.
- [63] M. Hofert. *Sampling Nested Archimedean Copulas with Applications to CDO Pricing*. PhD thesis, University of Ulm, 2010.
- [64] M. Hofert, M. Mächler, and A. McNeil. Likelihood inference for archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis*, 110:133–150, 2012.
- [65] M. Hofert, M. Mächler, and A. McNeil. Archimedean copulas in high dimensions: Estimators and numerical challenges motivated by financial applications. *Journal de la Société Française de Statistique*, 154(1):25–63, 2013.
- [66] M. Hofert and F. Vriens. Sibuya copulas. *Journal of Multivariate Analysis*, 114:318–337, 2013.
- [67] H. Hong and F. Hickernell. Algorithm 823: Implementing scrambled digital sequences. *ACM Transactions on Mathematical Software*, 29:95–109, 2003.
- [68] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 2012.

- [69] P. Huang, D. Subramanian, and J. Xu. An importance sampling method for portfolio CVaR estimation with Gaussian copula models. In *Proceedings of the 2010 Winter Simulation Conference (WSC)*, pages 2790 – 2800, 2010.
- [70] P. J. Huber and E. M. Ronchetti. *Robust Statistics*. John Wiley & Sons, Inc, New Jersey, 2009.
- [71] A. A. Javid. Copulas with truncation-invariance property. *Communications in Statistics - Theory and Methods*, 38(20):3756–3771, 2009.
- [72] P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik, editors. *Copula Theory and Its Applications*, volume 198 of *Lecture Notes in Statistics – Proceedings*. Springer, 2010.
- [73] H. Joe. *Dependence Modeling with Copulas*. Chapman & Hall/CRC, Boca Raton, 2014.
- [74] J. H. B. Kemperman. On the optimum rate of transmitting information. *The Annals of Mathematical Statistics*, 40(6):2156–2177, 1969.
- [75] F. Knight. *Risk, Uncertainty and Profit*. Reprints of Economic Classics. Augustus M. Kelley, New York, 1921.
- [76] V. Krätschmer, A. Schied, and H. Zähle. Comparative and qualitative robustness for law-invariant risk measures. *Finance Stoch.*, 18(2):271–295, 2014.
- [77] S. Kullback. *Information theory and statistics*. Dover Publications Inc., Mineola, NY, USA, 1968.
- [78] P. Kumaraswamy. A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46:79–88, 1980.
- [79] P. H. Kupiec. Stress testing in a value at risk framework. *Journal of Derivatives*, 6:7–24, 1998.
- [80] D. Kurowicka and R. Cooke. Sampling algorithms for generating joint uniform distributions using the vine-copula method. *Computational Statistics & Data Analysis*, 51:2889–2906, 2007.
- [81] J. Kwiatkowski and R. Rebonato. A coherent aggregation framework for stress testing and scenario analysis. *Applied Mathematical Finance*, 18(2):139–154, 2011.
- [82] C. Lemieux. *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer Series in Statistics, Springer, New York, 2009.
- [83] F. Liese and I. Vajda. On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412, 2006.
- [84] R. Litterman. Hot spots and hedges. *Journal of Portfolio Management*, 22:52–75, 1996.
- [85] J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. Springer, New York, 2008.
- [86] A. Marshall and I. Olkin. Families of multivariate distributions. *Journal of the American Statistical Association*, 83(403):834–841, 1988.

- [87] A. W. Marshall and I. Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62:30–44, 1967.
- [88] J. Matoušek. On the L_2 -discrepancy for anchored boxes. *Journal of Complexity*, 14:527–556, 1998.
- [89] A. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press, Princeton, 2005.
- [90] A. McNeil and N. J. Multivariate Archimedean copulas, d -monotone functions and l_1 -norm symmetric distributions. *The Annals of Statistics*, 37(5b):3059–3097, 2009.
- [91] A. J. McNeil and A. D. Smith. Multivariate Stress Scenarios and Solvency. *Insurance: Mathematics and Economics*, 50:299–308, 2012.
- [92] M. Mesfioui and J.-F. Quessy. Dependence structure of conditional Archimedean copulas. *Journal of Multivariate Analysis*, 99(3):372–385, 2008.
- [93] A. Meucci. Fully Flexible Views: Theory and Practice. *Risk*, 21(10):97–102, 2008.
- [94] A. Meucci. The Black–Litterman Approach: Original Model and Extensions. In *Encyclopedia of Quantitative Finance*. Wiley, New York, 2010.
- [95] W. Morokoff and R. Caflisch. Quasi-random sequences and their discrepancies. *SIAM Journal of Science Computing*, 15(6):1251–1279, 1994.
- [96] J. Natolski and R. Werner. Mathematical analysis of different approaches for replicating portfolios. *European Actuarial Journal*, 4(2):411–435, 2014.
- [97] J. Natolski and R. Werner. *Innovations in Quantitative Risk Management: TU München, September 2013*, chapter Improving Optimal Terminal Value Replicating Portfolios, pages 289–301. Springer International Publishing, Cham, 2015.
- [98] J. Natolski and R. Werner. Mathematical foundation of the replicating portfolio approach. *SSRN*: <http://ssrn.com/abstract=2771254>, 2016.
- [99] R. Nelsen. *An Introduction to Copulas*. Springer, New York, second edition, 2006.
- [100] H. Niederreiter. Point sets and sequences with small discrepancy. *Monatshefte für Mathematik*, 104:273–337, 1987.
- [101] H. Niederreiter. *Random number generation and quasi-Monte Carlo methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.
- [102] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer Series in Operations Research. Springer-Verlag, New York, 1999.
- [103] J. Nolan. Stable distributions – models for heavy tailed data. Technical report, 2014.
- [104] D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.

- [105] J. Oechslin, O. Aubry, M. Aellig, A. Kappeli, D. Bronnimann, A. Tandonnet, and G. Valois. Replicating embedded options. *Life and Pensions Risk*, pages 47–52, 2007.
- [106] A. Owen. Randomly permuted (t, m, s) -nets and (t, s) -sequences. In: *Niederreiter H, Shiue PJS (eds) Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Lecture Notes in Statistics*, 106:299–317, 1995.
- [107] A. Owen. Monte Carlo variance of scrambled equidistribution quadrature. *SIAM Journal on Numerical Analysis*, 34(5):1884–1910, 1997.
- [108] A. Owen. Scrambled net variance for integrals of smooth functions. *Annals of Statistics*, 25(4):1541–1562, 1997.
- [109] A. Owen. Variance and discrepancy with alternative scramblings. *ACM Transactions on Modeling and Computer Simulation*, 13:363–378, 2003.
- [110] A. Owen. Multidimensional variation for quasi-Monte Carlo. In: *Fan J, Li G (eds) International Conference on Statistics in honour of Professor Kai-Tai Fang's 65th birthday*, pages 49–74, 2005.
- [111] A. Pelsser and J. Schweizer. The difference between LSMC and replicating portfolio in insurance liability modeling. *SSRN: <http://ssrn.com/abstract=2557383>*, 2015.
- [112] T. Pillards and R. Cools. Using box-muller with low discrepancy points. In: *et al MG (ed) ICCSA 2006, Lecture Notes in Computer Science*, 3984:780–788, 2006.
- [113] R. Rebonato. *Coherent Stress Testing: A Bayesian Approach to the Analysis of Financial Stress*. The Wiley Finance Series. John Wiley & Sons, Ltd, Chichester, UK, 2010.
- [114] B. Remillard. *Statistical Methods for Financial Engineering*. Chapman and Hall/CRC, 2013.
- [115] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.
- [116] R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.
- [117] V. Schmitz. *Copulas and Stochastic Processes*. PhD thesis, Institute of Statistics, Aachen University, 2003.
- [118] I. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7:86–112, 1967.
- [119] I. Sobol'. Calculation of improper integrals using uniformly distributed sequences. *Soviet Mathematics Doklady*, 14:734–738, 1973.
- [120] G. Stahl, J. Zheng, R. Kiesel, and R. Rühlicke. Conceptualizing robustness in risk management. *Preprint*, 2012.
- [121] G. Studer. *Maximum loss for measurement of market risk*. PhD thesis, ETHZ, 1997.

- [122] D. Tasche. Capital Allocation to Business Units and Sub-Portfolios: the Euler Principle. In A. Resti, editor, *Pillar II in the New Basel Accord: The Challenge of Economic Capital*, pages 423–453. Risk Books, London, 2008.
- [123] W. R. van Zwet. A strong law for linear functions of order statistics. *The Annals of Probability*, 8(5):986–990, 1980.
- [124] D. Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.
- [125] F. Wu, E. Valdez, and M. Sheris. Simulating exchangeable multivariate archimedean copulas and its applications. *Communications in Statistics – Simulation and Computation*, 36(5):1019–1034, 2006.

Curriculum Vitae

MATHIEU CAMBOU

29/12/1988 ◊ French
mathieucambou@gmail.com

WORK EXPERIENCE

Edge Laboratories AG September 2015 - Present
Co-Founder *Lausanne & Zürich, CH*

Ecole Polytechnique Fédérale de Lausanne (EPFL) September 2011 - Exp. 2016
Teaching Assistant *Lausanne, CH*

SCOR Switzerland AG September 2011 - August 2015
Actuarial Research Consultant *Zürich, CH*

Bank of America Merrill Lynch July 2010 - August 2011
Structured Credit Trading *London, UK*

Cantor Fitzgerald June 2009 - June 2010
Quantitative Developer *London, UK*

EDUCATION

Ecole Polytechnique Fédérale de Lausanne (EPFL), CH September 2011-Exp. 2016
Ph.D. in Mathematics
Supervisors: Prof. Damir Filipovic & Prof. Anthony Davison.

University of Cambridge, UK September 2009-June 2010
M.Phil. in Statistical Science

Ecole Polytechnique Fédérale de Lausanne (EPFL), CH September 2006-June 2009
B.Sc. in Mathematics