Selected Computational Problems in Insurance

Andrew W.L. Fleck

A Dissertation submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

Graduate program in Applied Mathematics
York University
Toronto, Ontario

August 2024

Ship your grain across the sea;
after many days you may receive a return.
Invest in seven ventures, yes, in eight;
you do not know what disaster may come upon the land.

– Ecclesiastes 11:1–2

Abstract

The coming together of digital data sets, computational power and rigorous probability theory has transformed finance and insurance in the last century. Once the purview of heuristics and an almost artisanal knowledge, these fields have increasingly taken on a scientific sophistication in technique. Modern regulations even require firms to retain the mathematical skill necessary to perform complex risk analysis. Mechanical heuristics originally developed in the absence of probabilistic assumptions have been demystified and reworked for novel applications. Complex structured products can be simulated and statistical learning algorithms can be applied to gain insights where none existed before. This dissertation is concerned with such problems.

In the realm of property and casualty insurance, this thesis addresses challenges in risk estimation, quantification and allocation when the risk can be modelled by multivariate Stable distributions, which we will argue provide a suitable null model in the case of heavy-tailed losses. Traditionally the lack of means and distribution functions has rendered these distributions difficult to work with. We will sidestep this issue in estimation by using an integral transformation-based method of estimation. For risk quantification, we develop computationally simple and efficient representations of commonly used risk measures. Allocation then follows from our choice of dependence structure.

In the area of life contingencies, we will study a relatively new product, the fixed index annuity (FIA). The variety of annuity parameters and the complexity of the underlying index makes FIA comparisons very challenging. While still an insurance product, FIAs require sophisticated models of equity indices to analyze. We elect to use machine learning techniques to reproduce FIA-linked equity indices. In order to understand our often surprising results, we make use of a few stochastic volatility models.

Dedication

To my family: my dedicated parents Chris and Joane, my encouraging brother Sam, my super-heroic wife Manelle and of course my friends who know that they are family.

Acknowledgements

"A PhD can only be understood backwards; but must be lived forwards"

– N. Gold, paraphrasing A. Fleck, paraphrasing S. Kierkegaard

Reaching the end of a doctorate can feel anticlimactic. All the unfinished threads and potential future vistas, set against the tapestry of work that you have drawn from, makes your own contribution feel like a very small piece of the puzzle. But in writing the acknowledgements, the magnitude of the task strikes you when reviewing the sheer number of people who have had a hand in creating it. Truly it has been a journey.

No journey is complete without a beginning, and to start I should credit my high school teachers at St. Paul's, who showed a great deal of faith in my mathematical talent. From my first forays into academics at Carleton University I must thank Drs. Dave Amundsen and Lucy Campbell for being my first mathematical mentors and giving me the confidence to continue.

On arriving in Toronto I was lucky enough to become associated for my Masters with Dr. Sebastian Ferrando of TMU. His penetrating and curious mind, patience, wisdom, grace, and good counsel were and remain invaluable to me. Thank you Sebastian for always encouraging me to take the next step, and for helping me develop my critical faculty.

Coming to York University was another experience altogether: challenging and rewarding in equal parts. My chief supervisors and advisors were Drs. Edward Furman and Yang Shen. Dr. Shen's day-to-day encouragement, advice and brilliance were crucial. Dr. Furman is a giant of his field: the Canadian actuarial space is incomplete without his presence. His wealth of insights, suggestions, big-picture thinking and industry connections are the foundation of this thesis. Dr. Furman (and Nate Gold of course) being also responsible for my relationship with my most fruitful collaborator and friend Branislav Nikolić and the time we spent at CANNEX together. It would be remiss of me to neglect another influence through Dr. Furman, Ricardas Zitikis, whose mastery and good humour I treasure. I was also lucky with my academic "siblings," Yisub Kye, Nawarf Mohamed and Guanfu Qiao.

Special thanks also go to my office mates Gabriela Gonzalez and Justin Miles for tolerating my ups and downs.

A doctorate is a marathon, not a sprint. Without loved ones in your life to provide relief and perspective, it would be impossible. No acknowledgement is complete without my friends. At York, my comrades in the trenches provided essential moments of camaraderie. Allysa, I always looked forward to our MathLab chats. Nate & Nat, you have a storybook romance, unfortunately, balanced by your love of the Leafs (no one can be too perfect after all). Branislav, working and collaborating with you has been great fun, but most of all thanks for being a role model in more ways than one. Vishal and Neda, your support could always be counted on, and I always appreciated it.

Here I would like to thank the physics "club": Greg, Gabe, Anael, Matt & Meaghan, Patty and of course ... Andrew Macklin (Maria's membership is still under review, although henceforth is approved). Greg and Gabe's friendship survived the experience of being *my* roommate. Patty and Anael skateboarded with me once and I suppose that's the stuff of decades-long friendship. While Matt and Meaghan have now moved to Germany, they are never far from my thoughts. Andrew and Maria's absence in my life is a presence: there is far too little dancing without them.

Of course, that's not everyone. Lukas & Ashely maybe one day I'll actually enjoy ping pong without your company, but I doubt it. To Will, my oldest friend, I don't have the space here to express what your friendship has meant to me. It's an incredible journey we have been on, from nerdy teens discussing Jared Diamond and Sid Meiers Civ to our continuing affection today. Finally, Nikki gave me the greatest gift of all, Elephants, who am I kidding. No other answer would suffice.

Mere acknowledgement is insufficient for Manelle. Without my wife this doctorate simply would not exist. She has supported me through this process in every conceivable way. As was said on our wedding day, you try to fill each other's cup and give one another of your bread. Manelle has fulfilled this and then some. To my love, I will be forever thankful. You share all the fruits of this doctorate, part of an even longer journey I count myself incredibly lucky to be taking with you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABRM Additive Background Risk Model 5, 10

APT Arbitrage Pricing Theory 6

AY Accident Years 7

CAPM Capital Asset Pricing Model 6

CDF Cumulative Distribution Function 43

CF Characteristic Function 10, 12

CGMM Continuous Generalized Method of Moments 6, 10

CL Chain Ladder 7, 8

CLT Central Limit Theorem 11, 45

CR Cramér–Rao 19

DY Development Years 7

EDM Exponential Dispersion Model 13

ES Expected Shortfall 43

GCLT Generalized Central Limit Theorem 3, 45

GMM Generalized Method of Moments 17

MGF Moment Generating Function 29

MLE Maximum Likelihood Estimate 7, 8

PCPs Premium Calculation Principles 42

PDF Probability Density Function 6, 7

TCE Tail Conditional Expectation 43

VaR Value at Risk 43

WIPM Weighted Insurance Pricing Model 45

List of Symbols

$\pi[X]$ Premium Principle: The premium paid (in currency) for a risk represented by a random variable X 42

$H[X]$ Risk Measure: A functional over a suitable set of loss random variables mapping losses to *risk capital* 2

$\mathcal{F}\{f(x)\}(s)$ Fourier transform of f given by $\int_{-\infty}^{\infty} e^{-isx} f(x)dx$ 57

$\mathcal{L}\{f(x)\}(s)$ Laplace transform of f given by $\int_0^{\infty} e^{-sx} f(x)dx$ 58

$\mathcal{M}\{f(x)\}(s)$ Mellin transform of f given by $\int_0^{\infty} x^{s-1} f(x)dx$ 60, 157

For a function $f(x)$: $\mathcal{A}f$ or $(\mathcal{A}f)(x)$ Cursive capitals are reserved for integral operators. 20

$\mathbf{M}_{n\times m}$ Boldface quantities (with subscripts representing dimensions) are reserved for matrices and operators generally. For vectors, that is matrices with one column, we drop the dimension subscripts. 18

$\circ$ Outer product: Given two vectors $\mathbf{x}$ and $\mathbf{y}$ of appropriate length, we have $\mathbf{x} \circ \mathbf{y} = \mathbf{x}\mathbf{y}^\top$. 11

$\mathbf{M}_{n\times m}^\top$ Matrix Transpose 18

$\mathbf{X}$ or $\mathbf{x}$ Vectors: Random or deterministic vectors will be identified as boldface quantities. See Random Variables entry for more detail. 4

$F_X(x)$ Distribution Functions: The cdf of X at a specific value x i.e. $P(X \leq x)$ 2

X Random Variables: In general, we denote random quantities by capital letters and deterministic ones by lowercase. 2

$\mathbb{Cov}[X, Y]$ Covariance of random variables X and Y 44

σ_t,r_T **or** W_t **etc** Stochastic Processes: We denote families of random variables indexed by time (t or T) as either lower- or uppercase letters with time as the subscript. 81

$\mathbb{Var}[X]$ Variance of a random variable 44

$\phi_X(\tau) = \mathbb{E}[e^{i\tau X}]$ Characteristic function of a random variable X 12

$f_X(x|\theta)$ Density Functions: The pdf of X at a specific value x given parameters θ 7

Chapter 1

Introduction

> It is not possible to give a definition of insurance which is short and precise, and at the same time completely satisfactory.
>
> Karl H. Borch

The argument could be made that humans, as social creatures, invented the most basic insurance when they started living in groups that shared resources. Formally, insurance in some form or another has existed as long as human civilization has been sufficiently complex to support it. The first codified examples date back to ancient Babylon and Hammurabi. In a more modern, legal sense, we can think of insurance as a contract binding two parties together: the *insurer* and the *insured*. The insurer guarantees compensation to the insured in the case of specific uncertain events in the future in exchange for a certain premium paid today. This contracted bridge between certain amounts today and uncertain amounts tomorrow falls naturally within the purview of mathematics, specifically probability theory.

Indeed mathematics in the form of actuarial science has become essential to the understanding of insurance and the management of risk therein for insurers. At times the

contribution has flowed the other way, from the actuarial sciences to mathematics. Years before Kolmogorov's *Foundations of the Theory of Probability* was published, the great Harald Cramér was already working as an actuary in addition to his role as a professor. He would go on to make innumerable contributions to our understanding of probability and statistics as well as its application to more pure fields of mathematics (e.g. number theory). Today the Cramér–Lundberg model and risk theory can be considered the jewel in the crown of actuarial science.

The same can also be said of finance. Even earlier than Cramér, Bachelier had already gained practical trading experience running his family business before publishing his *Théorie de la Spéculation*. This text would go on to influence the program of probability theory and stochastic analysis for the next half century, with Einstein kicking off a parallel program in the physical sciences via his seminal paper on Brownian motion (see the Introduction of [Davis and Etheridge, 2006]). Today mathematical finance is a healthy and well-developed subfield of applied mathematics, primarily concerned with the valuation of contingent claims, which can themselves be viewed as a kind of insurance contract.

Specifically, in both finance and insurance most tasks can be formulated as follows. Given a random liability of interest (e.g. a capital loss, lifetime earning etc.), we want to model it by a random variable X described by a distribution function $F_X(x)$. Assuming our model is sufficient, we need to produce a real number $H[X]$ representing some ordinal measure of risk; $H[X]$ can be used to derive premiums directly or for management or marketing purposes to compare products and contracts. Nowadays regulatory bodies require that firms retain the mathematical sophistication necessary to perform complex risk calculations, so $H[X]$ may represent regulated and required metrics of risk to maintain solvency and financial stability.

As in Cramér's and Bachelier's time, these questions of a computational and mathematical nature in finance and insurance have often predated their formal mathematical study. This dissertation is concerned with three such examples:

- In Chapter 2 we focus on the basic question of how to model insurance losses and how to estimate quantities associated with such a model. Specifically, we will study

the practical question of how to estimate heavy-tailed losses for the purposes of loss reserving.

- In Chapter 3 we will expand on a special case of joint losses from Chapter 2. We believe our model can serve as a suitable null model in the case of heavy-tailed exposure via the Generalized Central Limit Theorem (GCLT) (see A.3). We then pair this with a model of risks and capital allocation ($H[X]$ above) that makes for easy and useful calculations.

- In Chapter 3 we will switch from property and causalty–style insurance questions to life contingencies, specifically indexed annuities. We will investigate the systematic comparison of these complex, financially linked products.

In each chapter, we will provide some of the general background required for the problem in hand, and will also elucidate any connections that may exist in the mathematics between them.

Chapter 2

Stochastic Loss Reserving: Dependence and Estimation

Earlier we claimed that much of insurance modelling can be abstracted to modelling losses X and some associated risk $H[X]$. However, this is an oversimplification. Nowadays insurers have to account for not just stand-alone risks but also potentially complex dependencies between risks. In that case X is now a vector $\mathbf{X}$ and the total risk capital is $H[\sum_i X^{(i)}]$. We study possible decomposition between business lines in Chapter 3. For now we focus on the losses alone. Another aspect not often considered is the role of *time*. In reality there is no single $\mathbf{X}$ that we need to model to calculate premiums but several, as claims are incurred through time. The total sum of these losses over time is the so-called "loss reserve," the portion of premiums set aside to meet future claims and maintain an insurer's solvency.

2.1 Introduction

Highly regulated and vitally necessary, the loss reserve is typically the largest liability on an insurance company's balance sheet. Proper estimation of future claims is therefore paramount for financial stability. In fact, with the introduction of regimes like Solvency

II actuaries are now sometimes required not just to estimate reserves but model potential shortfalls and risks of insolvency. This makes a *stochastic* model of loss reserves necessary (see [Wüthrich and Merz, 2008] or [Fröhlich and Weng, 2018] and references therein). This is complicated by the fact that delays between an incurred claim and proper reporting can take some time, often years. There may also be ongoing or renewed liability at a later date for many reasons, such as legal proceedings and lengthy investigations. In the recent past, there have been several high profile examples of these kinds of "tort liabilities" such as asbestos and other environmental pollutants ([Carmean, 1995] and [Madigan and Metzner, 2003]). Newer concerns such as the health risks of engineered materials may present similar issues ([McAlea et al., 2016]). Complicating estimation even further is the potential dependence among claims between business lines. One example that could induce such a dependence may be industry-specific inflationary trends. Medical costs can often rise faster than economy-wide price levels; accident business lines especially may need to incorporate this into their reserves. Similarly, auto repair techniques may incur increased costs for both commercial and personal lines. Such a dependence can represent a potential for diversification or an increased risk to the insurer (see e.g. [De Jong, 2012]).

Given the importance of proper loss reserving it is unsurprising that there are as many forms of reserve estimation as techniques in statistics (see [Wüthrich and Merz, 2008] and references therein). In this paper we focus on the model popularized by [Merz et al., 2013] and model claims parametrically and dependence "cell-wise" across business lines. Within this framework, it is popular to model severity and dependence separately via copulas (e.g. [Zhang and Dukic, 2013]). We instead take a multivariate modelling approach as in [Avanzi et al., 2016]. The benefits and drawbacks of these two approaches (copulas vs multivariate models) are essentially the same as in traditional statistics. Copulas provide a great deal of model flexibility but at the cost of increased numbers of parameters and decreased interpretability. Multivariate models are much more parsimonious but restrict the available marginals.

In order to negotiate this trade-off we construct incremental loss models across business lines via an Additive Background Risk Model (ABRM). ABRMs provide an easily interpretable and flexible dependence through the use of a common shock structure across

business lines. The technique can be easily extended to a variety of marginal distributions leading to many possible multivariate models. By way of example, this chapter makes use of the multivariate gamma and Tweedie model of, respectively, [Furman and Landsman, 2005] and [Furman and Landsman, 2010] as in [Avanzi et al., 2016] as well as introduces a particular multivariate Stable distribution. The idea of additive risk models is not new in economics and finance. The most famous examples are of course the Capital Asset Pricing Model (CAPM) [Fama and French, 2004] and Arbitrage Pricing Theory (APT) [Ross, 2013]. More recently the potential for applications in insurance – especially enterprise risk management – has been explored (see e.g. [Furman et al., 2018] and [Zhou et al., 2018]). Other ways to introduce dependence in loss reserves are the Multiplicative Background Risk Models (MBRMs) (e.g., [Furman et al., 2021], [Asimit et al., 2016], [Semenikhine et al., 2018] and [Marri and Moutanabbir, 2022]), minimum-based background risk models (e.g., [Asimit et al., 2010] and [Pai and Ravishanker, 2020]), and background risk models that allow for multiple types of risk factors (e.g., [Su and Furman, 2017a] and [Su and Furman, 2017b]).

The main contribution of our work is not just applying ABRMs but also model estimation. Useful loss models frequently lack a closed form or computationally simple Probability Density Function (PDF), making classical estimation difficult (e.g. compound Poisson, NIG in the Tweedie case or most non-normal stables). The small sample sizes and many parameters in reserve models naturally lead many to rely on a Bayesian analysis (see for example [Zhang and Dukic, 2013]). There have been attempts to study estimation in the multivariate Tweedie using the method of moments ([Alai et al., 2016]) but for the reasons already stated this seems inappropriate. Continuous Generalized Method of Moments (CGMM) of [Carrasco and Florens, 2000] offers a hope of success where the basic GMM may fail. Incorporating an infinite number of moment conditions makes the CGMM maximally statistically efficient. In this work we outline a novel use of the CGMM that is relatively computationally inexpensive, especially for "larger" multivariate models. By using the CGMM in conjunction with ABRMs we open up the practical application of a variety of models without the need to resort to highly uncertain Bayesian estimates.

This chapter is organized as follows. In Section 2.2, we give a more detailed account of

cell-wise loss modelling and basic estimation. A discussion of what makes a useful model and the introduction of our Tweedie and Stable ABRM examples takes place in Section 2.3. The CGMM and our novel approach are outlined in Section 2.4. Finally, some simulation results and an illustration using real data are given in Sections 2.5 and 2.6.

2.2 Parametric Loss-Reserving Models

In this section we briefly review parametric loss-reserving models. Let us consider the following situation. We are in the mth accident year since writing a particular non-life policy. We have been able to observe incremental claims $x_{i,j}$ for each of $i = 1, \dots, n$ Accident Years (AY)[1] and $j = 1, \dots, m - i + 1$ Development Years (DY). We assume that all these are individual samples from random variables $X_{i,j}$ that are stochastically independent. We add some necessary structure by enforcing shared parameters in a typical way (see [Wüthrich and Merz, 2008]):

$$\mathbb{E}[X_{i,j}] = \mu_{i,j} = \eta_i \nu_j \quad \text{and} \quad \mathbb{V}\text{ar}[X_{i,j}] = \sigma_{i,j} = w_{i,j}\gamma_j \tag{2.1}$$

where $w_{i,j}$ is an appropriate weight and γ_j a scale parameter. For some $X_{i,j}$ with a PDF of the form $f_{X_{i,j}}(x_{i,j}|\mu_{i,j}, \sigma_{i,j})$, such as Tweedie, Stable, and so on, we can easily construct a Maximum Likelihood Estimate (MLE) for the model parameters:

$$\{(\hat{\eta}_i, \hat{\nu}_j, \hat{\gamma}_j)\}_{i,j}^{m} = \underset{\eta_i, \nu_j, \gamma_j}{\operatorname{argmin}} \left\{ \prod_{i,j}^{i+j<m+1} f_{X_{i,j}}(x_{i,j}|\mu_{i,j}, \sigma_{i,j})|_{\mu_{i,j}=\eta_i\nu_j, \sigma_{i,j}=w_{i,j}\gamma_j} \right\}. \tag{2.2}$$

In fact, for Tweedie-distributed incremental losses, estimators of the form (2.2) are numerically similar to the well-known Chain Ladder (CL) estimators of [Mack, 1993] (see [Mack, 1991] or especially [Taylor, 2009] for more details). Notably, for a Tweedie power parameter of $p = 1$ (the overdispersed Poisson) the correspondence is exact. For values

[1] Year losses were incurred but not necessarily paid.

"close" to $p = 1$, these estimates are *very* similar conditional on the scale. For example, if we consider a Gamma model of the form

$$X_{i,j} \sim \text{Gamma}\left(\frac{1}{\gamma_j}, \eta_i \nu_j \gamma_j\right) \sim Tw_2(\eta_i \nu_j, \gamma_j) \tag{2.3}$$

where $Tw_p(\mu, \sigma^2)$ denotes the reproductive Tweedie density (see A.2 for definitions and details), we can simulate a Gamma-distributed loss development triangle and compare the MLE-derived development factors to those from the CL estimates in Figure 2.1. It can be seen that the CL and MLE estimates are virtually identical and both of them are fairly good estimates for the true model.

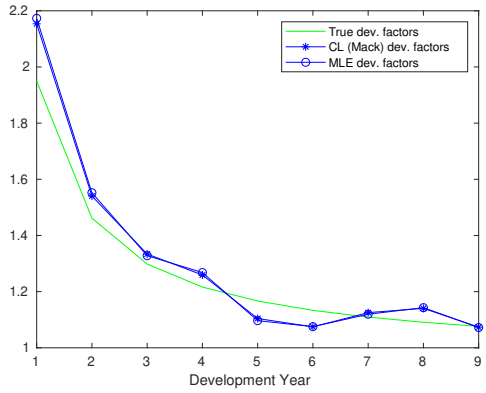
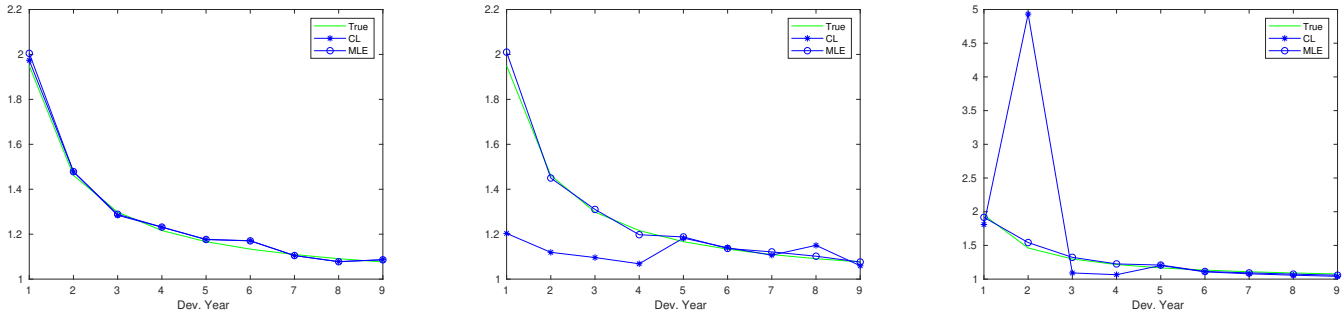


Figure 2.1: Comparison of the true, MLE- and CL-estimated development factors in a Gamma-distributed development triangle, specifically $p = 2$ with $\gamma = \gamma_j = 0.2$, $\eta_i = \eta = 5$ and ν ranging over 1 to 0.55

While the Tweedie class of distributions is quite large, there are some losses for which it is inappropriate. For example, fire and automobile insurance coverage frequently exhibit heavy Pareto-style tails, suggesting an infinite or undefined variance (see [Seal, 1980]). The behaviour of such heavy-tailed distributions is qualitatively different from the typical Tweedie models near $p = 1$ (or thin-tailed models generally). One extremely useful and well-motivated model is the Stable distribution. If we repeat our experiment from Figure 2.1 with a stable loss model we can see in Figure 2.2 how the CL estimates quickly break

down.



(a) (b) (c)

Figure 2.2: Examples of draws from the same development triangle ($\eta = 5$ and $\nu = 1-0.55$) featuring Stable-distributed losses with tail parameter $\alpha = 1.8$

In the stable case, many draws would exhibit fairly typical behaviour (see Figure 2.2a). This is not too surprising as the CL estimates are unbiased provided the mean exists. In the typical case, the sample is mostly represented by the fairly well-behaved centre of the distribution. However, successive draws can reveal that even one or two losses reported out on the tail of the distribution can quickly contaminate the CL estimates, rendering them useless (refer to Figure 2.2b and 2.2c). It is worth pointing out that in Figure 2.2b a major loss early in the development pattern had the unexpected effect of *underestimating* the development factors. This may not be so obvious in a less marked example when a sample from the tail is creating similar issues. Though the examples chosen are extreme by design, it is easy to see that blind application of non-parametric estimates is ill-advised.

To fully appreciate the implications of heavy tails, we consider that in Figure 2.2c a single cell incurred about 50% of all losses in the triangle. In multiple business lines sharing systematic risk, this can be even more consequential. A single draw from a significantly heavy-tailed risk shared across a portfolio could potentially be greater than the reserve estimated in a thin-tailed model. Such counterintuitive behaviour for heavy tails cannot be ignored.

While a dramatic level of risk exists in heavy-tailed losses, there is hope of overcoming

this issue. Unsurprisingly, the MLE estimates in Figure 2.2 are adequate, but this is not without major qualification. First, in order to compute the MLEs one needs a PDF. In the stable case (excepting the Normal distribution), there is no closed-form PDF with finite variance. Calculating the PDF requires the numerical inversion of the Characteristic Function (CF) or the evaluation of a truncated infinite series to some precision. This numerical quadrature can be very expensive and in the case of multiple business lines is not practical. Additionally, the level of precision required is very high. In the MLEs we studied, we found that the scale parameter was often poorly reported. Furthermore, the method is very sensitive to the choice of initial points used in the optimization. In the case of Figure 2.2b, a few attempts had to be made before producing the results shown.

In this chapter we motivate the use of stable loss models. We also extend them to the case of multiple lines of business with a stable ABRM. The ABRM has a flexible and easily interpretable dependence structure modelling cell-wise dependence with a multivariate stable distribution. In order to deal with the aforementioned estimation challenges, we make use of the CGMM of [Carrasco and Florens, 2000]. These estimates are comparable to MLEs in a more typical case with an order of magnitude fewer calculations. This is achieved by constructing estimators from the CF directly as opposed to reconstructing a PDF. For example, in the examples above with identical optimization parameters, the CGMM was about thirty times faster than the MLE (60 vs 2000 seconds on a standard IBM ThinkPad). We show that the CGMM makes the multivariate estimation of multiple business lines a practical reality for stable ABRMs.

2.3 An Additive Background Risk Model

While a single loss development triangle can be used to estimate the required loss reserve facing that line of business, an insurance company typically manages a portfolio of several lines. Given the importance of reserve estimates, dependence between reserves must be modelled. To that end, we identify a few desirable properties of any cell-wise dependent model:

- Marginal flexibility: Any model must capture a large class of possible distributions.
- Closure under marginal convolution: To model sub-portfolios accurately, we must be able to add incremental cells together easily.
- Model confidence: Our model should exist in the limit of a set of reasonable stochastic models, as for example the Normal arises from the Central Limit Theorem (CLT). This allows us to feel confident applying the model in general situations.
- Simple and flexible dependence: Any dependence must be transparent and easily interpretable.

To establish our last desired property we introduce an ABRM. The assumptions behind this model are simple. For the kth line of business (LoB) of a total of n, we assume a cell-wise model of the form

$$X_{i,j}^{(k)} = a_{i,j}^{(k)} Y_{i,j}^{(k)} + b_{i,j} Z_{i,j}. \tag{2.4}$$

That is, we have an independent idiosyncratic component $Y_{i,j}^{(k)}$ and a "common shock" component $Z_{i,j}$ across business lines. If we denote

$$\mathbf{a}_{i,j} = (a_{i,j}^{(1)}, ..., a_{i,j}^{(n)})' \quad \text{and} \quad \mathbf{b}_{i,j} = (b_{i,j}^{(1)}, ..., b_{i,j}^{(n)})'$$

then we can arrange the same cells across business lines into a vector form:

$$\mathbf{X}_{i,j} = \mathbf{a}_{i,j} \circ \mathbf{Y_{i,j}} + \mathbf{b}_{i,j} Z_{i,j} \tag{2.5}$$

where "$\circ$" denotes the outer product. In this way, we can reduce dependence to the $\mathbf{b}_{i,j}$ parameters alone. In fact, we can show that for some models (e.g. Normal) this is related to the typical Pearson correlation structure.

The density of such a model can be found by integrating the univariate PDFs à la equation (2.9) in [Avanzi et al., 2016] or equation (6.1) of [Furman and Landsman, 2010].

As the PDFs may not always exist, we are more concerned with the CF. Letting $\mathbf{t} = (t^{(1)}, ..., t^{(n)})'$, we compute the multivariate CF in terms of the univariate CFs:

$$\phi_{\mathbf{X}_{i,j}}(\mathbf{t}) = \mathbb{E}\left[\exp\left\{i\mathbf{t}^\top \mathbf{X}_{i,j}\right\}\right] \tag{2.6}$$

$$= \mathbb{E}\left[\exp\left\{\sum_{k=1}^{n} it^{(k)} a_{i,j}^{(k)} Y_{i,j}^{(k)} + i\mathbf{t}^\top \mathbf{b}_{i,j} Z_{i,j}\right\}\right] \tag{2.7}$$

$$= \prod_{k=1}^{n} \phi_{Y_{i,j}^{(k)}}(t^{(k)} a_{i,j}^{(k)}) \phi_{Z_{i,j}}(\mathbf{t}^\top \mathbf{b}_{i,j}) \tag{2.8}$$

where $\mathbf{x}^\top \mathbf{y}$ denotes the inner product. We can see how the value $\mathbf{t}^\top \mathbf{b}_{i,j}$ will control the dependence structure of the distribution. We now move on to some specific examples to illustrate why this representation may be more appropriate.

2.3.1 Example 1: A Multivariate Tweedie Approach

In our first example we discuss the multivariate Tweedie's desirable properties. It satisfies the requirements of marginal flexibility and closure under marginals. As discussed, special cases of the Tweedie are the Poisson and Gamma distributions; a more exhaustive list would include the compound Gamma, Normal, and Normal Inverse Gaussian distributions. Additionally, there is a Tweedie class generated by Stable distributions. Being an exponential dispersion model, the multivariate Tweedie has many desirable properties for inference, computation, and interpretation.

We can construct the Tweedie ABRM model in the form of Eq. (2.4), as follows:

$$Y_{i,j}^{(k)} \sim Tw_p(\eta_i^{(k)} \nu_j^{(k)}, \gamma^{(k)}), \tag{2.9}$$

$$Z_{i,j} \sim Tw_p(\alpha, \beta) \tag{2.10}$$

where the additional model parameters are

$$a_{i,j}^{(k)} = 1, \tag{2.11}$$

$$b_{i,j}^{(k)} = \left(\frac{\alpha}{\eta_i^{(k)} \nu_j^{(k)}} \right)^{1-p} \frac{\gamma^{(k)}}{\beta}. \tag{2.12}$$

This gives us a cell-wise distribution of

$$X_{i,j}^{(k)} = Tw_p \left(\eta_i^{(k)} \nu_j^{(k)} \left[\left(\frac{\alpha}{\eta_i^{(k)} \nu_j^{(k)}} \right)^{2-p} \frac{\gamma^{(k)}}{\beta} + 1 \right], \gamma^{(k)} \left[\left(\frac{\alpha}{\eta_i^{(k)} \nu_j^{(k)}} \right)^{2-p} \frac{\gamma^{(k)}}{\beta} + 1 \right] \right). \tag{2.13}$$

For various values of the Tweedie power parameter p, we lack a closed-form PDF. We can however easily construct the univariate CF via the cumulant function of the Tweedie distribution (see sections (A.1) and (A.2) for details). If $Y \sim Tw_p(\mu, \sigma^2)$, then

$$\mathbb{E}[e^{itY}] = \exp\left\{ \frac{1}{\sigma^2} \left[\kappa_p(g(\mu) + it\sigma^2) - \kappa_p(g(\mu)) \right] \right\}$$

where $\kappa_p(\cdot)$ is given by equation A.10.

The universality properties of the Tweedie are perhaps what makes it most applicable to insurance loss modelling. Tweedie distributions are the only EDMs that are closed under rescaling. Take any Exponential Dispersion Model (EDM) whose variance is asymptotically a function of the mean such that $\mathbb{V}\text{ar}[Y] \sim C \cdot \mathbb{E}[Y]^p$. This model "converges" in some sense with rescaling to a Tweedie distribution ([Jørgensen, 1987]).

Abusing notation slightly, suppose $ED(\mu, \sigma)$ is a random variable with an exponential dispersion distribution function (A.8). Reading the following equality in the sense of distribution, if

$$cED(\mu, \sigma) = ED(c\mu, c^{2-p}\sigma^2)$$

then $ED(\mu, \sigma)$ is Tweedie. EDMs converge under rescaling to Tweedies, in much the same way that Stable variables remain stable under addition and the limit of normalized sums. For any $\mu > 0$ and $\sigma^2 > 0$, suppose $\mathbb{V}\text{ar}[Y] = \sigma^2 V(\mu)$ (A.9); if $V(\mu) \sim c_0 \mu^p$ as $c \to 0$ (resp.

∞), then we have

$$\frac{ED(c\mu, \sigma^2 c^{2-p})}{c} \to Tw_p(\mu, c_0\sigma^2)$$

in distribution as $c \to 0$ (resp. ∞). Consider some hypothetical losses incurred according to an EDM with s, μ and $s\sigma$ where s is the scale (\$1,000's, \$1M's etc.) as long as the variance remains constant at scale (i.e. $p = 0$). In this case,

$$\frac{ED(s\mu, (s\sigma)^2)}{s} \to Tw_2(\mu, s_0) = \mathcal{N}(\mu, s_0).$$

Another example would be for discrete data. Imagine that we are modelling the number of claims N in some risk model with distribution $N \sim ED(\mu, 1)$, where $V(\mu) \sim \mu$. Then

$$\frac{ED(c\mu, c)}{c} \to Tw_1(\mu, 1) = Poi(\mu).$$

A trivial example of this would be if claims were binomially distributed and the number of claims were aggregated or rescaled. The compound Poisson–Gamma is a Tweedie distribution (a compound distribution of two other Tweedies) that may be obtained using this mechanism, which may explain its popularity as a loss model.

2.3.2 Example 2: A Multivariate Stable Approach

We now consider the most "natural" model that would meet all our required criteria: the multivariate Normal distribution. As the basin of attraction in the Central Limit Theorem, it has the required closure under convolutions and an interpretable linear dependence structure. In fact, for Tweedie power parameter $p = 0$ we recover this model.

Since we lack marginal flexibility, the symmetric nature of the Normal makes it awkward for insurance applications. We can however generalize the Normal to the *Stable* family of distributions. Furthermore, we can consider the class of totally skewed Stables for a more realistic application to insurance. Stables exist as the basin of attraction of sums of

heavy-tailed random variables. Thus, given the data compiled by [Eaton et al., 1971] and [Embrechts et al., 2013], Stable models are a good candidate for heavy-tailed insurance portfolios, which *cannot* be captured by the Tweedie ARBM. Indeed, we will expand on this model for insurance losses in Section 3.3.

Proceeding in the same way as before,

$$Y_{i,j}^{(k)} \sim S_\alpha(\eta_i^{(k)}\nu_j^{(k)}, \gamma^{(k)}, 1), \tag{2.14}$$

$$Z_{i,j} \sim S_\alpha(\mu, \sigma, 1), \tag{2.15}$$

we can set

$$a_{i,j}^{(k)} = b_{i,j}^{(k)} = 1. \tag{2.16}$$

In fact we can show that for these parameter values $\mathbf{X_{i,j}}$ is a true multivariate Stable vector with marginals

$$X_{i,j}^{(k)} = S_\alpha(\eta_i^{(k)}\nu_j^{(k)} + \mu, ((\gamma^{(k)})^\alpha + \sigma^\alpha)^{\frac{1}{\alpha}}, 1), \tag{2.17}$$

where the mean, if it exists, is μ, and we specify scale σ and importantly the skewness parameter β. We consider the special case of maximum skewness i.e. $\beta = 1$ for the rest of the chapter as it is the most appropriate for loss modelling. The parameter α controls the heavy-tailedness of the distribution.

A Stable model for losses could arise simply through aggregation of smaller losses. Whenever there is an incremental loss, it can be written as the sum of smaller i.i.d. losses L_i:

$$X_n = \frac{L_1 + \cdots + L_n}{p_n} - q_n. \tag{2.18}$$

Then by the Generalized Central Limit Theorem (Theorem A.3.2), we have that $f_{X_n} \to f_X$ weakly where f_X is a standardized Stable distribution:

$$X \xrightarrow{dist.} S_\alpha(1, \beta, 0).$$

In both the ABRMs just discussed, we often lack a closed-form PDF of the kind we had in the earlier Gamma model (Eq. (2.3)). In the Stable case, for instance, we only have a closed-form PDF for tail parameters $\alpha = 0, 5, 1, 2$. Unlike our simple example earlier we also have many more parameters across n business lines. This further compounds our computational issues.

And yet we have shown such models are desirable. This raises the question: Is there a way of estimating reserves given the observed part of a loss triangle using only what we are guaranteed to have, for example the CF given by Eq. (2.8)? Also, what sacrifices in terms of efficiency would this imply, if any? We explore these issues in the following section.

2.4 Estimation via Continuous Generalized Method of Moments

2.4.1 Motivation

As seen in Section 2.2, given the cell-wise PDF of our claims it is fairly easy to construct maximum likelihood estimators for our model. Unfortunately, as pointed out in the previous section, the ABRMs can lack such a PDF for many parameter values. Indeed, this may make even simple cases (single loss triangles, not too many AY/DYs) computationally expensive. As a point of comparison, one estimation from Table 2.3 in Section 2.5.1 with identical optimization parameters is about thirty times faster using the continuous generalized method of moments (CGMM) objective than with the MLE objective. This is due to the fact that every evaluation of the PDF/likelihood requires a numerical quadrature of a characteristic or moment function. In multiple lines of business, this is made even worse by the need to add another quadrature for a convolution of univariate PDFs.

In such a case, it may be preferable to seek out alternatives to likelihood estimation. This is the approach adopted by [Avanzi et al., 2016] in relation to [Furman and Landsman, 2010], employing a Markov chain Monte Carlo approach to the completion of a Bayesian analysis. While experience rating can doubtless help the estimation procedure along, we

would hope this is not the only recourse available, especially given the often counterintuitive hidden risks of heavy-tailed models specified as in Eq. (2.17).

The only other method of inference introduced in the multivariate Tweedie case is the method of moments (see [Alai et al., 2016]). There has also been much interest in using a method of moments-style estimator for Stable models, but with the characteristic function as a moment condition (see e.g. [Bee and Trapin, 2018] or [Koutrouvelis, 1980]). However, this is inappropriate for loss reserving due to a lack of large enough sample sizes. That said, there is a generalization we may use.

First, let us review the classical Generalized Method of Moments (GMM) introduced by [Hansen, 1982]. The GMM has become overwhelmingly popular in econometrics. This success is in part due to the fairly arbitrary moment conditions required and the lack of distributional assumptions. The GMM can for example handle complex nonlinear regressions involving tricky economic concepts such as endogenous variables.

Unfortunately, the wide applicability of the GMM comes at the cost of statistical efficiency. Because they "throw away" the excess information of perfect specification and introduce ad hoc moment conditions, GMM estimators are widely viewed as less efficient than their more onerous MLE equivalents[2]. This is one reason why we should be skeptical about applying GMM methods to the small samples in loss reserving. That said, it is not hard to show that if the moment conditions are the score of the correctly specified distribution, we actually recover the MLE. To see this, we consider a sample of $\{\mathbf{x}_1, .., \mathbf{x}_l\}$ i.i.d. realizations from some random variable $\mathbf{X}$ and for $r \in \{1, ..., l\}$ specify moment conditions of the form

$$\mathbb{E}[\mathbf{g}(\theta : \mathbf{x}_r)] = 0, 1 \leq r \leq l \tag{2.19}$$

where $\mathbf{g}(\theta : \mathbf{x}) = (g_1(\theta : \mathbf{x}), g_2(\theta : \mathbf{x}), ..., g_{d_g}(\theta : \mathbf{x}))$ is a vector-valued function of some model parameters θ, $dim(\mathbf{x}) = d_x$, $dim(\mathbf{g}) = d_g$, $dim(\theta) = d_\theta$ and $d_g > d_\theta$; thus Eq. (2.19)

[2]There are recent optimal efficiency results (see [Ackerberg et al., 2014] for two-step semi-parametric models) but this is unsurprising as we will shortly see.

is *overdetermined*[3]. We can take a sample average to approximate (2.19):

$$\mathbf{g}_l(\theta) = \frac{1}{l}\sum_{r=1}^{l}\mathbf{g}(\theta : \mathbf{x}_r). \tag{2.20}$$

For some positive definite matrix $\mathbf{W}_{d_g \times d_g}$, we can define an inner product and corresponding norm and minimize the norm of Eq. (2.20) to arrive at the GMM objective function. The GMM estimates for θ are then[4]

$$\theta^*_{GMM} = \underset{\theta}{\operatorname{argmin}} \|\mathbf{W}^{1/2}_{d_g \times d_g}\mathbf{g}_l(\theta)\|. \tag{2.21}$$

The most statistically efficient matrix is $\mathbf{W}_{d_g \times d_g} = \mathbb{E}[\mathbf{g}_l(\theta_o) \circ \mathbf{g}_l(\theta_o)]^{-1}$, which is the inverse of the covariance matrix of our moment conditions. We can prove that such estimators exhibit asymptotic normality such that $\sqrt{l}(\theta^*_{GMM} - \theta_o) \sim \mathcal{N}(0, (\mathbf{G}^\top_{\mathbf{d_g}\times\mathbf{d}_\theta}\mathbf{W}_{\mathbf{d_g}\times\mathbf{d_g}}\mathbf{G}_{\mathbf{d_g}\times\mathbf{d}_\theta})^{-1})$, where θ_o is the true parameter value and

$$\mathbf{G}_{d_g \times d_\theta} = \mathbb{E}\left[\frac{\partial \mathbf{g}(\theta)}{\partial \theta}\Big|_{\theta=\theta_o}\right] \tag{2.22}$$

Where the expectation is w.r.t. $\mathbf{X}$ and the derivative $\frac{d}{d\theta}$ is understood as the gradient in the case $d_\theta > 1$. Interestingly, this suggests that the correct moment conditions can recover the same information that was "thrown away" from the MLE. Indeed, we will see that the result of [Carrasco and Florens, 2000] formalizes just this idea to construct CGMM estimators. For now, let us consider the moment condition of form

$$\mathbf{g}(\theta : \mathbf{x}_r) = \frac{\partial \ln(f_{\mathbf{X}}(\mathbf{x}_r|\theta))}{\partial \theta}. \tag{2.23}$$

That is, Eq. (2.23) is the score function. Obviously, $\hat{\theta}_{MLE}$ will solve Eq. (2.19), while $\mathbf{G}_{d_\theta \times d_\theta} = -\mathbf{I}(\theta)$ and $\mathbf{W}_{d_\theta \times d_\theta} = \mathbf{I}(\theta)$ where $\mathbf{I}(\theta)$ is the Fisher information matrix. Therefore

[3]Note abuse of notation: technically we should denote θ by $\boldsymbol{\theta}$, but as a set of parameters we want to set it aside for succinctness and to avoid confusion with data and other vectors.

[4]Obviously this is equivalent to $\mathbf{W}^{1/2}_{d_g \times d_g}\mathbf{g}_l(\theta)^\top \mathbf{W}^{1/2}_{d_g \times d_g}\mathbf{g}_l(\theta) = \mathbf{g}_l(\theta)^\top \mathbf{W}_{d_g \times d_g}\mathbf{g}_l(\theta)$, $\mathbf{g}^\top$ being the transpose.

$\sqrt{l}(\theta^*_{GMM} - \theta_o) \sim \mathcal{N}(0, \mathbf{I}^{-1})$. That is, the GMM estimators achieve the Cramér–Rao (CR) bound and are thereby as efficient as the MLEs. The intuition here is unsurprising as we are making use of the same information, namely specification, as in the MLE. In some sense, then, the GMM can be seen as more general than likelihood estimation. The CGMM in turn is a further generalization to a *continuum* of moment conditions.

2.4.2 The CGMM

Consider the specific problem of estimating models efficiently from their integral transforms, namely the aforementioned Tweedie and Stable models. In [Feuerverger and McDunnough, 1981], the authors suggested the following moment condition. For $\{\boldsymbol{\tau}_k\}_{k=1}^{d_g}$ we choose a moment condition involving a CF of the form

$$g_k(\theta : \mathbf{x}) = e^{i\boldsymbol{\tau}_k{}^\top \mathbf{x}} - \phi_{\mathbf{x}|\theta}(\boldsymbol{\tau}_k). \tag{2.24}$$

They claimed this could be thought of as a potentially infinitely over-specified moment condition. The idea is that by sampling an increasing number of $\boldsymbol{\tau}$ values and proceeding with the typical GMM, one could achieve the CR bound. However, they did not prove this result, nor did they take into account how to construct the proper covariance matrix $\mathbf{W}$ in such a way as to continuously match the infinite-dimensional moment condition of Eq. (2.24).

To see why a continuous moment matching is necessary consider the resulting vector-valued function $\mathbf{g}_l(\theta)$ (Eq. (2.20) and covariance matrix:

$$\mathbf{K}_{d_g \times d_g} = \mathbf{W}^{-1}_{d_g \times d_g} = \mathbb{E}[\mathbf{g}_l(\theta_o)\mathbf{g}_l(\theta_o)^\top] = \{k_{ij}\}_{i,j=1}^{l}$$

where

$$k_{ij} = \phi_{X|\theta_o}(\tau_i - \tau_j) - \phi_{X|\theta_o}(\tau_i)\phi_{X|\theta_o}(-\tau_j).$$

Note that $\mathbf{K}_{d_g \times d_g}$ is Hermitian and positive definite, and satisfies properties that we would normally find desirable. However, one can show that the smallest eigenvalue of such a matrix will approach zero as d_g goes to infinity. This will make operations involving $\mathbf{W}_{d_g \times d_g}$ and the GMM very unreliable numerically. In fact, if we were to take the limit as d_g approaches infinity, $\mathbf{W}_{d_g \times d_g}$ would become unbounded and not invertible. This leaves us in need of a procedure to *continuously* match our $\mathbf{W}$-norm in (2.21) with the moment conditions.

In [Carrasco and Florens, 2000] the authors tackled these problems by inventing the CGMM. Rather than simply sampling from a continuous object as in Eq. (2.24), they reformulated the GMM in *continuous* terms: vectors and matrices become elements of a Hilbert space and operators, respectively. Intuitively, if we consider $d_g \longrightarrow \infty$ instead of a *vector* moment condition we get a *function*, so that $\mathbf{g}_l(\theta) \longrightarrow \frac{1}{l}\sum_{r=1}^{l} e^{i\boldsymbol{\tau}^\top \mathbf{x}_r} - \phi_{X|\theta}(\boldsymbol{\tau})$, specifically a function of $\boldsymbol{\tau}$.

In the interest of brevity, given a function of several variables g we will write integrals like so:

$$\int g(\mathbf{x})d\mathbf{x} = \int_{R^n} ... \int g(x_1, ..x_n)dx_1...dx_n. \tag{2.25}$$

To avoid confusion with the discrete inner product above, the L^2 inner product for functions f and g is defined as

$$< g, f >_{L^2} = \int g(x)\overline{f(x)}dx. \tag{2.26}$$

In order to give the typical notion of L^2, we set

$$L^2 = \left\{ f \,\middle|\, < f, f >_{L^2} < \infty \right\}, \tag{2.27}$$

the space over which we will primarily be working in the CGMM. Finally, given $f \in L^2$ and a kernel function $k : \mathbb{R}^{d_\theta} \times \mathbb{R}^{d_\theta} \to \mathbb{R}$ we will denote the notion of a Fredholm-style operator $\mathcal{K}$ as

$$(\mathcal{K}f)(\mathbf{t}) = \int k(\mathbf{t}, \mathbf{s})f(\mathbf{s})d\mathbf{s}. \tag{2.28}$$

By way of example to illustrate the CGMM, we again consider an i.i.d. sample $\{x_1, ..., x_l\}$, but this time define the moment condition by a *function* of τ:

$$h(\boldsymbol{\tau}, \theta : \mathbf{x}) = e^{i\boldsymbol{\tau}^\top \mathbf{x}} - \phi_{\mathbf{x}|\theta}(\boldsymbol{\tau}) \tag{2.29}$$

so that

$$h_l(\boldsymbol{\tau}, \theta) = \frac{1}{l} \sum_{r=1}^{l} h(\boldsymbol{\tau}, \theta : \mathbf{x}_r). \tag{2.30}$$

And rather than a matrix we define a positive-definite kernel function:

$$k(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \mathbb{E}[h_l(\boldsymbol{\tau}_1, \theta)\overline{h_l(\boldsymbol{\tau}_2, \theta)}]. \tag{2.31}$$

We can generalize the GMM estimators to their continuous counterparts:

$$\theta^*_{CGMM} = \underset{\theta}{\operatorname{argmin}} \|\mathcal{K}^{-1/2} h_l\|_{L^2} = \underset{\theta}{\operatorname{argmin}} < \mathcal{K}^{-1/2} h_l, \mathcal{K}^{-1/2} h_l >_{L^2}. \tag{2.32}$$

While the operator $\mathcal{K}$ and its inverse are self-adjoint, the expression $< h_l, \mathcal{K}^{-1} h_l >_{L^2}$ is often undefined. The inverse square root of the operator has a larger domain than the simple inverse and that is why we write Eq. (2.32) as here.

Estimators of the kind defined in Eq. (2.32) can be shown to be *as efficient* as the MLE, achieving the CR bound. The intuition behind this is discussed in [Carrasco et al., 2014]. The GMM can recover the MLE if the moment conditions involve the score; by generalizing the GMM to Hilbert spaces we can relax this. We now only require the new moment conditions to contain the score within their linear *closure*. Naturally, this is true for moment conditions of form (2.30). Figure 2.3 summarizes the situation.

$\mathbf{g}(\theta : \mathbf{x}) = (g_1(\theta : \mathbf{x}), ..., g_{d_g}(\theta : \mathbf{x}))$	$h(\tau, \theta : \mathbf{x}) = e^{i\boldsymbol{\tau}^\top \mathbf{x}} - \phi_{\mathbf{x}\|\theta}(\boldsymbol{\tau})$
$\mathbf{g}_l(\theta) = \frac{1}{l}\sum_{r=1}^{l} \mathbf{g}(\theta : \mathbf{x}_r)$	$h_l(\boldsymbol{\tau}, \theta) = \frac{1}{l}\sum_{r=1}^{l} h(\boldsymbol{\tau}, \theta : \mathbf{x}_r)$
$\mathbf{g}_l(\theta) : \begin{array}{ccc} \mathbb{R}^{d_\theta} & \longrightarrow & \mathbb{R}^{d_g} \\ \theta & \longrightarrow & \mathbf{g}_l(\theta) \end{array}$	$h_l(\boldsymbol{\tau}, \theta) : \begin{array}{ccc} \mathbb{R}^{d_\theta} & \longrightarrow & L^2 \\ \theta & \longrightarrow & h(\boldsymbol{\tau}, \theta) \end{array}$
$\mathbf{W}_{d_g \times d_g} : \begin{array}{ccc} \mathbb{R}^{d_g} & \longrightarrow & \mathbb{R}^{d_g} \\ \mathbf{g}_l & \longrightarrow & \mathbf{W}\mathbf{g}_l \end{array}$	$\mathcal{K} : \begin{array}{ccc} L^2 & \longrightarrow & L^2 \\ h(\boldsymbol{\tau}, \theta) & \longrightarrow & \int k(\mathbf{s}, \boldsymbol{\tau}) h(\boldsymbol{\tau}, \theta) d\boldsymbol{\tau} \end{array}$
$\theta^*_{GMM} = \underset{\theta}{\operatorname{argmin}} \|\mathbf{W}^{1/2}_{d_g \times d_g} \mathbf{g}_l(\theta)\|$	$\theta^*_{CGMM} = \underset{\theta}{\operatorname{argmin}} \|\mathcal{K}^{1/2} h_l\|_{L^2}$

Figure 2.3: Summary of GMM and CGMM counterparts

Evaluating the objective is equivalent to solving a Fredholm integral equation of the first kind. That is, the function $u(\tau) = (\mathcal{K}^{-1/2} h_l)(\tau)$ can be seen as a solution to the equation

$$\mathcal{K}^{1/2} u = h_l, \tag{2.33}$$

where $\mathcal{K} : D(\mathcal{K}) \to R(\mathcal{K})$ is the integral operator with kernel given by Eq. (2.31).

In what follows, we provide a quick introduction to solving integral equations in a practical way. In general, the true inverse of an operator, namely $\mathcal{K}^{-1} : R(\mathcal{K}) \to D(\mathcal{K})$ such that $(\mathcal{K}\mathcal{K}^{-1} f)(\cdot) = id(\cdot)$, is unbounded. Solutions of Eq. (2.33) will only be defined for a dense subset of $D(\mathcal{K})$. To ensure existence everywhere, we need to weaken our notion of a solution to one that solves the least squares problem. Therefore, rather than solving $\mathcal{K}u = f$, we solve

$$\mathcal{K}^\dagger u = \underset{u \in D(\mathcal{K})}{\operatorname{argmin}} \left\{ \|\mathcal{K}u - f\|^2 \right\}. \tag{2.34}$$

The operator $\mathcal{K}^\dagger : R(K) \to D(K)$ is called the *pseudoinverse* of K. This satisfies the existence problem. Unfortunately, problems of the kind (2.34) are still not well posed.

Given especially that our moment conditions (2.30) are estimated from data, we need to introduce the regularized problem that *is* well posed:

$$\mathcal{K}^{\dagger}u_{\lambda} = \underset{u \in D(\mathcal{K})}{\operatorname{argmin}} \left\{ \|\mathcal{K}u - f\|^2 + \lambda \|u\|^2 \right\}. \tag{2.35}$$

This regularized minimization problem has a unique solution $u_{\lambda} = (\mathcal{K}'\mathcal{K} + \lambda I)^{-1}\mathcal{K}'f$, where $\mathcal{K}'$ is the adjoint of $\mathcal{K}$. As $\lambda \to 0$, we recover (2.34). We can thus think of the regularized problem as approximating an ill-posed problem with a "nearby" well-posed one.

2.4.3 Loss Reserve Estimates

Let us revisit the example from Section 2.2 involving the Gamma losses in Figure 2.1. We will compute estimates for η, ν and γ using the CGMM instead of the MLE. We define the following:

- $X_{i,j} \sim \text{Gamma}\left(\frac{1}{\gamma_j}, \eta_i \nu_j \gamma_j\right)$
- $\theta = (\eta_i, \nu_j, \gamma_j)$ so that $\phi_{X_{i,j}|\theta}(\tau) = \mathbb{E}[e^{i\tau X_{i,j}}] = (1 - i\tau\eta_i\nu_j\gamma_j)^{-\frac{1}{\gamma_j}}$
- $h_{i,j}(\tau) = e^{i\tau X_{i,j}} - \phi_{X_{i,j}|\theta}(\tau)$
- $\mathcal{K}_{i,j}f = \int k_{i,j}(\tau, s) f(s) d\pi(s)$ where $k_{i,j}(\tau, s) = \phi_{X_{i,j}|\theta}(\tau - s) - \phi_{X_{i,j}|\theta}(\tau)\phi_{X_{i,j}|\theta}(-s)$

The objective function to be used is

$$(\hat{\eta}_i, \hat{\nu}_j, \hat{\gamma}_{i,j}) = \underset{\eta_i, \nu_j, \gamma_j}{\operatorname{argmin}} \sum_{i,j} < \mathcal{K}_{i,j}^{-1/2} \hat{h}_{i,j}, \mathcal{K}_{i,j}^{-1/2} \hat{h}_{i,j} > . \tag{2.36}$$

Evaluating the $\mathcal{K}_{i,j}^{-1/2}\hat{h}_{i,j}$ using the Nyström method and numerically optimizing (detailed in the next section), we obtain the mean expected claims shown in Table 2.1.

Given that for Tweedie losses with power parameter close to 2, we expect the CL and MLE estimates to be close, it is unsurprising that the CGMM parametric estimate for the development factors is also similar, as shown in Figure 2.4.

Accident Year	Realized Losses	Estimated Claims (CGMM)	Estimated Claims (MLE)
2	2.92	1.93	2.66
3	8.10	9.70	7.88
4	12.37	20.08	17.04
5	27.07	37.22	27.92
6	41.73	71.03	45.49
7	63.84	85.42	66.37
8	94.94	100.68	98.76
9	127.07	110.27	113.36
10	167.27	134.52	138.03

Table 2.1: Realized losses from the scenario used in Figure 2.1 vs the MLE- and CGMM-estimated mean losses

2.4.4 Numerical Considerations

Estimating the Kernel Function In both the GMM and the CGMM, we have thus far ignored the estimation of the matrix $\mathbf{W}_{d\times d}$ and the kernel operator $\mathcal{K}$ respectively. Recall they are defined as the expectation of the Kronecker product of our moment functions, which require the very quantity we are trying to estimate, namely θ_o. In the i.i.d. case, the natural choice of estimator is simply

$$\hat{k}(\tau_1,\tau_2) = \frac{1}{l}\sum_{t=1}^{l}(e^{i\tau_1 x_t} - \hat{\phi}(\tau_1))\overline{(e^{i\tau_2 x_t} - \hat{\phi}(\tau_2))}, \tag{2.37}$$

where $\hat{\phi}(\tau) = \frac{1}{l}\sum_{t=1}^{l} e^{i\tau x_t}$.

Unfortunately, in loss reserving we do not have the luxury of sizable i.i.d. samples. To address this issue, we continuously update the kernel function (defined in (2.31)) after a reasonable initial estimate (e.g. the CL estimates). This strategy is appropriate for the same reason that continuously updating in the discrete GMM case works: as [Carrasco and Florens, 2000] shows, any appropriate kernel function that produces a norm can produce a consistent estimate for θ_o. Unfortunately, this does add some computational cost.

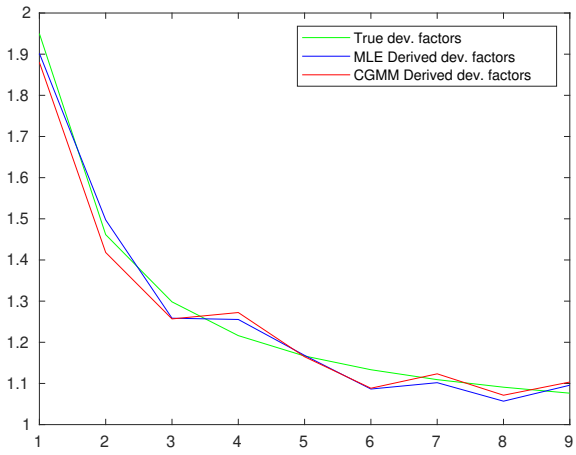


Figure 2.4: Comparison of true estimated development factors from the Gamma model (green) vs MLE model (blue) and CGMM estimates from Eq. (2.36) (red)

Estimation of λ The choice of regularization parameter λ obviously affects the estimators. If it is too large (resp. too small), the problem we are solving is too far from the original regularized problem (resp. we lose numerical stability). We direct the reader to the discussion in [Kotchoni, 2012] for possible simulation-based approaches to optimizing this parameter. In the loss-reserving setting, we find that a good "rule of thumb" is to set $\lambda \approx 10^{-7}$ when the number of development years is around 10.

Numerical Integral Equation Solutions 1: The Nyström method The most obvious way of evaluating (2.33) is via numerical quadrature. We introduce quadrature points[5] $\{s_q\}_{q=1}^{Q}$ and corresponding weights $\{w_q\}_{q=1}^{Q}$ such that

$$(\mathcal{K}h)(\tau) = \int k(\tau, s)h(s)d\pi(s) \approx \sum_{q=1}^{Q} w_q k(\tau, s_q)h(s_q).$$

[5]Capital Q here is taken to be deterministic.

We can then represent the continuous problem of the kind (2.33) as

$$\begin{pmatrix} w_1k(s_1,s_1) & \dots & w_qk(s_1,s_q) & \dots & w_Qk(s_1,s_Q) \\ \vdots & & \vdots & & \vdots \\ w_1k(s_q,s_1) & \dots & w_qk(s_q,s_q) & \dots & w_Qk(s_q,s_Q) \\ \vdots & & \vdots & & \vdots \\ w_1k(s_Q,s_1) & \dots & w_qk(s_Q,s_q) & \dots & w_Qk(s_Q,s_Q) \end{pmatrix}^{1/2} \begin{pmatrix} u(s_1) \\ \vdots \\ u(s_q) \\ \vdots \\ u(s_Q) \end{pmatrix} = \begin{pmatrix} \hat{h}_T(s_1) \\ \vdots \\ \hat{h}_T(s_q) \\ \vdots \\ \hat{h}_T(s_Q) \end{pmatrix}$$

or more succinctly as $\mathbf{K}^{1/2}_{Q\times Q}\mathbf{u}_{Q\times 1} = \mathbf{h}_{Q\times 1}$. We then evaluate the CGMM objective (2.32) as ${\mathbf{u}_\lambda}^\top\mathbf{u}_\lambda$, where $\mathbf{u}_\lambda = (\mathbf{K}_{Q\times Q} + \lambda\mathbf{I}_{Q\times Q})^{-1}\mathbf{K}^{1/2}_{Q\times Q}\mathbf{h}_{Q\times 1}$.

Numerical Optimization Consider a more general moment function of the form

$$h(\tau,\theta : X) = m(X,\tau) - E[m(X,\tau)]. \tag{2.38}$$

Thus far, we have only discussed expressions of the kind $m(X,\tau) = e^{i\tau X}$ leading to $E[m(X,\tau)] = \phi_{X|\theta}(\tau)$, the CF. The CGMM is not limited to such expressions. For example, given a CDF $F_X(x)$ it is just as natural to choose $m(X,\tau) = \mathbf{1}_{\{X>\tau\}}$, leading to an equivalent expression of form

$$h(\tau,\theta : X) = \mathbf{1}_{\{X>\tau\}} - F_X(\tau). \tag{2.39}$$

A closed-form CDF is, of course, lacking in the models we are interested in, so this point may seem moot. However, we emphasize that if the linear closure of expressions of the form $h(\tau,\theta : X)$ contains the score of our distribution, we can employ the CGMM confidently, as demonstrated by the following theorem.

Theorem 2.4.1 ([Carrasco et al., 2014])**.** *Consider the subspace $L^2(h)$ of L^2 formed by the linear hull of $\{h(\tau,\theta : X), \tau \in \mathbb{R}^{d_x}\}$. That is, for $w_j \in \mathbb{R}$ consider the space of random*

variables of the form

$$j_n = \sum_{j=1}^{n} w_j h(\tau_j, \theta : X),$$

and their mean square limits G such that

$$\mathbb{E}[\|j_n - j\|_{L^2}] \xrightarrow[n\to\infty]{} 0.$$

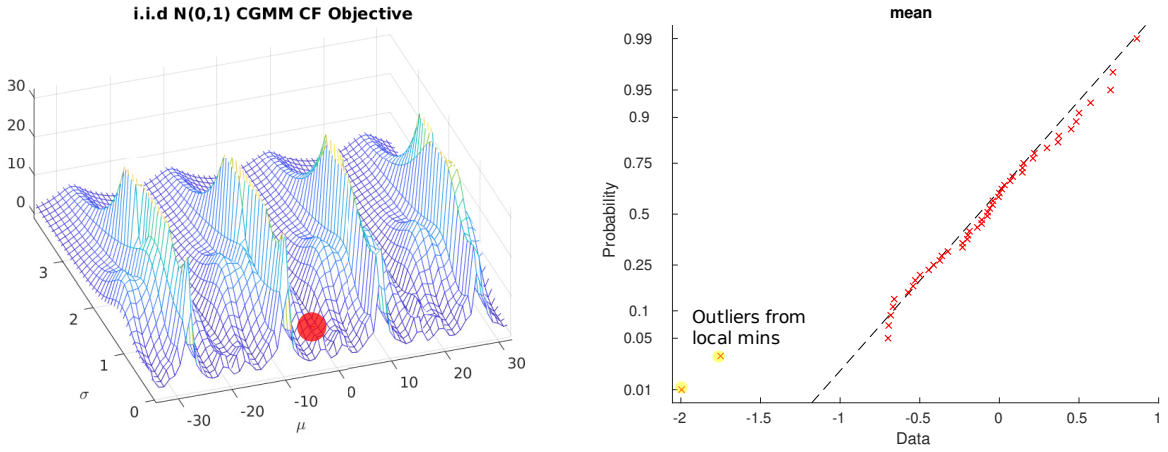
If the score $s_\theta(x) = \dfrac{\partial \ln(f_X(x|\theta))}{\partial\theta} \in L^2(h)$, then the resulting CGMM estimator for θ is asymptotically as efficient as the MLE.

Thus, for example in one dimension we can use the characteristic function and approximate a limit point in $L^2(h)$ by an integral as follows:

$$\begin{aligned}
\lim_{n\to\infty} \sum_{j=1}^{n} w_j h(\tau_j, \theta : x) &\approx \int h(\tau, \theta : X) w(\tau) d\tau \\
&= \int e^{i\tau x} w(\tau) d\tau - \int \phi_{X|\theta}(\tau) w(\tau) d\tau \\
&\qquad \left(\text{letting } w(\tau) = \frac{1}{2\pi} \int e^{-i\tau x} s_\theta(x) dx \right) \\
&= s_\theta(x) - \int s_\theta(x) \left[\frac{1}{2\pi} \int e^{-i\tau x} \phi_{X|\theta}(\tau) \right] dx \\
&= s_\theta(x) - E[s_\theta(x)] \\
&\qquad \text{(where the expectation is w.r.t. } \theta\text{, i.e. } E[s_\theta(x)] = 0) \\
&= s_\theta(x).
\end{aligned}$$

Theoretically, there are few issues with choosing the CF. Practically speaking, however, the CF may lead to non-convex objectives that are difficult to optimize. For instance, consider the following simple situation. Given an i.i.d. sample of $\{x_1, ..., x_l\}$, where $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = 0$ and $\sigma = 1$, and $l = 10$ and $m(x, \tau) = e^{i\tau x}$, we can plot the surface of (2.32) as shown in Figure 2.5a. Keep in mind, we wish to find the minimum.

The true minimum (shown by the red dot in Figure 2.5a) is still close to $(0, 1)$, as



(a) (b)

Figure 2.5: Complex-valued functions lead to periodicity and outliers

expected. We can see however that the presence of complex-valued functions creates oscillations and periodicity. Often, numerical errors and the sheer non-convexity of the objective can lead to standard gradient descent algorithms getting "stuck" out in neighbourhoods of adjacent local minimums. This is especially true in the heavier-tailed case, where it is harder to guess the initial point to start the optimization. To see this, look for instance at the Prob-Prob plot of the sampling distribution for μ (generated using $(l = 10)$-sized Stable samples with $\alpha = 1.5$) in Figure 2.5b. While the distribution appears mostly Normal (which, in theory, would be the case asymptotically) we can see that these numerical issues create a few notable outliers. In [Kotchoni, 2012] the application of the CGMM to heavy-tailed Stable variates required rather large sample sizes to overcome this problem. We suspect that this issue is partly why the CGMM has not yet seen wide application.

In the loss-reserving case, the picture is more challenging. The increased dimensions of the objective and possible local minimums mean that it is a near certainty that standard algorithms will get stuck. Two similar alternative approaches to finding a global minimum that we explore are MATLAB's built-in scatter-search and genetic algorithms. The former is used to generate the results in our Gamma example of Section 2.4.3. Both of these algorithms search a larger portion of the feasible space and often produce favourable results.

Another useful trick is to add a penalty for solutions found too far away from a good first estimate (say, the chain-ladder parameters). These approaches often produce the true optimum. However, the reality is that the fundamental non-convexity and periodicity of the objective are too much to overcome reliably.

Ideally, we would like to find moment conditions that are not just statistically valid but also produce well-behaved objective functions. It is preferable to make use of integral transforms that produce real and (log-)convex functions such as the moment-generating function. That is,

$$h_t(\tau, \theta : X) = e^{\tau X} - M_{X|\theta}(\tau) \tag{2.40}$$

with corresponding kernel function

$$k(s, \tau) = M_{X|\theta}(s + \tau) - M_{X|\theta}(s) M_{X|\theta}(\tau) \tag{2.41}$$

where $M_X(\tau)$ is the Moment Generating Function (MGF) of X. However, we run into issues showing this satisfies the conditions of Theorem 2.4.1. To begin with , if left unrestricted, h is no longer in L^2. Additionally, we would require $w(\tau) = -\frac{i}{2\pi} \int e^{-tx} s(x) dx$, which has no real coefficients.

In light of this, we can instead perform a "Wick rotation" and recover the MGF that way. Setting $\tau \to -i\tau$, we can make use of (2.40) as long as we keep the CF form for $\mathcal{K}$:

$$k(s, \tau) = M_{X|\theta}(s - \tau) - M_{X|\theta}(s) M_{X|\theta}(-\tau). \tag{2.42}$$

Using the MGF and this new kernel, we can generate a new objective function as in Figure 2.6 from the same scenario and data used in Figure 2.5a. As we can see, the objective function is now convex and much more amenable to numerical optimization.

One aspect of note that will reoccur in Section 2.5.2 is that the objective function has little curvature around the optimum. This "flatness" becomes even more pronounced as we add dimensions to the data (or lines of business in the loss-reserving problem).

Before continuing to numerical examples, we have one additional issue to consider. To

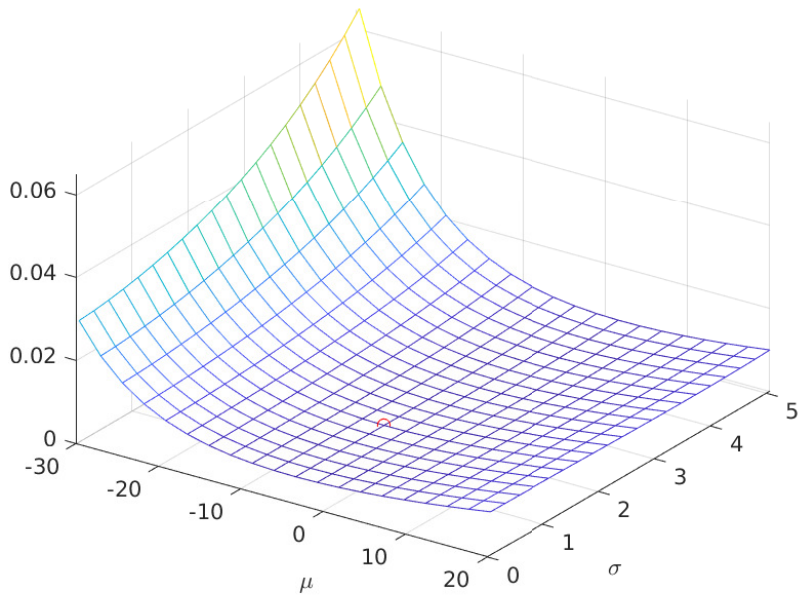


Figure 2.6: CGMM objective using the MGF (2.40) and the kernel from (2.42)

guarantee our conditions are properly bounded, we introduce a dampening function. In methods where the empirical characteristic function is used (e.g. kernel density estimation), it is well-known that some values of τ are unreliable. If we use the MGF instead, the same issue arises but now the actual MGF may not even be defined at some τ. As a remedy to both issues, dampening functions are employed to emphasize more reliable regions of τ over others. In our setting, this takes the form

$$h_t(\tau, \theta : X) = \Psi(\tau)(m(X, \tau) - E[m(X, \tau)]) \tag{2.43}$$

where $\Psi(\tau)$ is a dampening function. It is easy to show that this fits well into the CGMM framework. For instance, in [Carrasco and Florens, 2000] this takes the form of a new measure $\pi(\mathbf{x})$ and space $L^2(\pi) = \{f | f(\mathbf{x})\overline{f(\mathbf{x})}d\pi(\mathbf{x}) < \infty\}$ where the CGMM results still hold true.

2.5 Simulation Results

2.5.1 Single LoB Illustration

For each of the examples discussed in this section, we simulate 50 single LoB (line of business) loss triangles with 10 accident years and 10 development years. For simplicity, we use the same parameter values as in our Gamma examples. We use (2.40) and (2.42) to construct a CGMM objective of the form (2.36). We then find the estimators via MATLAB's `fmincon` function, which employs sequential quadratic programming.

Tweedie Losses: Compound Gamma and Gamma In the univariate case, the (reproductive) MGF of the Tweedie distribution is

$$M_{X|\mu,\sigma}(\tau) = \exp\left(\frac{1}{\sigma^2}[\kappa_p(g(\mu)+\tau\sigma^2) - \kappa_p(g(\mu))]\right). \tag{2.44}$$

For some values of p, the MGF cannot be defined past some positive τ. For example, in the $p = 2$ (Gamma) case, we require $\tau < \frac{1}{\mu\sigma^2}$. This may create some issues with the kernel function (recall the $s - \tau$ argument in the first term). To avoid this we choose a dampening function of the form $\Psi(\tau) = \mathbf{1}_{\{t<0\}}$ and consider the MGF only over $(-\infty, 0]$:

$$h(\tau, \theta : x) = \mathbf{1}_{\{\tau<0\}}\left(\exp\{\tau x\} - M_{X|\theta}(\tau)\right). \tag{2.45}$$

Using this formulation, we construct an objective of the form (2.36), but rather than the CF we use the MGF and a kernel of the form (2.41). Integration for $\|\mathcal{K}^{-1/2}\hat{h}_T\|$ is done via a simple trapezoidal scheme [6]

Table 2.2 summarizes the statistics for the CGMM estimates for Tweedie power parameters $p = 1.2$ and $p = 2$. For comparison, we include the chain ladder and MLE estimators,

[6] Interestingly, Newton–Coates schemes require fewer quadrature points than Gaussian quadrature. Despite having the freedom to select any π, it appears the function is well-behaved enough that linear approximations are sufficiently precise.

respectively. While the mean parameters of η and ν are slightly biased, the advantage goes to the CGMM in the estimation of the scale. The bias is present in the Stable example as well and we will discuss why in that case. For now, it is likely safe to conclude that the CGMM and CL/MLE estimators are at least comparable.

Stable Losses Given that Stable variables do not admit even second moments for all but $\alpha < 2$, using the MGF over the CF to generate a CGMM objective in the Stable case seems misguided at first glance. Fortunately, [Eaton et al., 1971] and [Samorodnitsky, 2017] have studied the Laplace transformation of an extreme Stable distribution ($\beta = -1$) and applied the Paley–Wiener theorems to conclude that for $X \sim S_\alpha(\mu, \sigma, -1,)$ the "moment generating function" is given by

$$M_{X|,\mu,\sigma}(\tau) = \exp\left(\mu\tau - \frac{\sigma^\alpha \tau^\alpha}{\cos(\alpha\pi/2)}\right), \text{ for } \tau > 0. \tag{2.46}$$

Note the fact that the resulting cumulant function is similar to the Tweedie, which formalizes the connection we hinted at earlier.

Again, for negative values of τ this obviously produces complex-valued arguments. Hence, once again we consider only the half-line that works for us. We also multiply by -1 to ensure positive losses:

$$h(\tau, \theta : x) = \mathbf{1}_{\{\tau<0\}} \left(\exp\{\tau(-x)\} - M_{-X|\theta}(\tau)\right). \tag{2.47}$$

In Figure 2.7, we compare a slice of the CGMM objective using the Stable MGF with another using the CF. As one can see, the difference is quite stark. Again, while a global search may give the correct minimum, it is not difficult to see why the MGF-based moment functions give much better and more consistent performance.

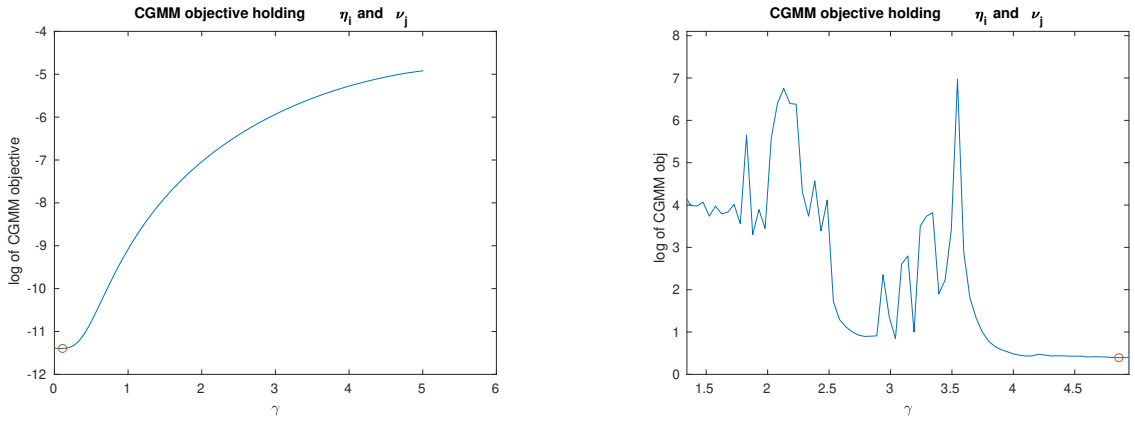
In Table 2.3, the estimates for η are slightly biased. This is in turn reflected in a small underestimation of the outstanding losses in Figure 2.8. We suspect this is due to the fact that compared to the MLEs, there is an extra parameter to optimize: the previously

discussed regularization parameter λ. We conduct a crude search by checking which values of λ give the best performance in a small series of simulations. A more systematic treatment could potentially offer better results and eliminate any bias. As previously mentioned, see discussion in [Kotchoni, 2012] for more on treatment of λ.

		CGMM		Chain Ladder	
	True	Median	SD	Median	SD
ν_1	5.00	4.71	0.77	5.12	1.14
ν_2	5.00	4.73	0.59	4.99	1.24
ν_3	5.00	4.84	0.76	5.05	1.17
ν_4	5.00	4.64	0.80	4.95	1.29
ν_5	5.00	4.60	0.66	4.77	1.39
ν_6	5.00	4.79	0.92	5.23	1.44
ν_7	5.00	4.49	0.80	4.77	1.44
ν_8	5.00	4.28	0.75	4.67	1.14
ν_9	5.00	4.29	0.91	4.69	1.20
ν_{10}	5.00	4.84	0.90	5.06	1.10
η_1	1.00	1.00	0.00	1.00	0.00
η_2	0.95	0.97	0.08	0.92	0.12
η_3	0.90	0.88	0.07	0.87	0.09
η_4	0.85	0.84	0.09	0.79	0.11
η_5	0.80	0.82	0.09	0.77	0.11
η_6	0.75	0.76	0.09	0.74	0.10
η_7	0.70	0.72	0.10	0.70	0.12
η_8	0.65	0.68	0.11	0.66	0.11
η_9	0.60	0.66	0.12	0.58	0.13
η_{10}	0.55	0.58	0.13	0.53	0.17
γ	0.20	0.22	0.10	0.08	0.07

		CGMM		MLE	
	True	Median	SD	Median	SD
ν_1	5.00	4.97	2.40	4.75	1.08
ν_2	5.00	4.59	2.00	5.24	1.04
ν_3	5.00	4.35	1.67	5.12	1.10
ν_4	5.00	4.41	2.25	5.16	1.05
ν_5	5.00	4.39	2.38	4.90	1.23
ν_6	5.00	4.82	2.42	4.72	1.20
ν_7	5.00	4.88	2.27	5.18	1.23
ν_8	5.00	4.29	2.05	4.57	1.40
ν_9	5.00	4.52	1.75	4.67	1.42
ν_{10}	5.00	4.64	1.70	4.78	1.86
η_1	1.00	1.00	0.00	1.00	0.00
η_2	0.95	0.90	0.12	0.94	0.15
η_3	0.90	0.82	0.14	0.87	0.18
η_4	0.85	0.76	0.14	0.79	0.18
η_5	0.80	0.74	0.14	0.77	0.18
η_6	0.75	0.72	0.12	0.71	0.15
η_7	0.70	0.67	0.14	0.66	0.16
η_8	0.65	0.68	0.17	0.63	0.22
η_9	0.60	0.63	0.18	0.59	0.23
η_{10}	0.55	0.61	0.21	0.54	0.26
γ	0.20	0.19	0.11	0.13	0.03

Table 2.2: Statistics of the CGMM estimators in the Tweedie case with $p = 1.2$ (overdispersed Poisson, left) and $p = 2$ (Gamma, right) after simulating 50 triangles. Included are the chain ladder and MLE estimators for comparison.



(a) MGF-based Objective

(b) CF-based Objective

Figure 2.7: Samples of the CGMM objective functions around the minima for MGF and CF moment functions

	True	Mean	SD
η_1	5	4.82	0.30
η_2	5	4.78	0.28
η_3	5	4.80	0.26
η_4	5	4.84	0.25
η_5	5	4.83	0.30
η_6	5	4.79	0.23
η_7	5	4.80	0.28
η_8	5	4.80	0.27
η_9	5	4.84	0.25
η_10	5	4.82	0.27

	True	Mean	SD
ν_1	1	1.00	0.00
ν_2	0.95	0.96	0.04
ν_3	0.9	0.92	0.04
ν_4	0.85	0.86	0.05
ν_5	0.8	0.81	0.05
ν_6	0.75	0.75	0.04
ν_7	0.7	0.70	0.05
ν_8	0.65	0.66	0.05
ν_9	0.6	0.59	0.06
ν_{10}	0.55	0.54	0.07

γ	0.2	0.18	0.02

Table 2.3: Statistics of the CGMM estimators in the stable case after simulating 50 triangles

2.5.2 Multiple LoB Illustration

In this section, we extend all the same methods from the single LoB to two LoBs. We simulate 50 joint triangles from a two-dimensional model of the kind specified by (2.5). To

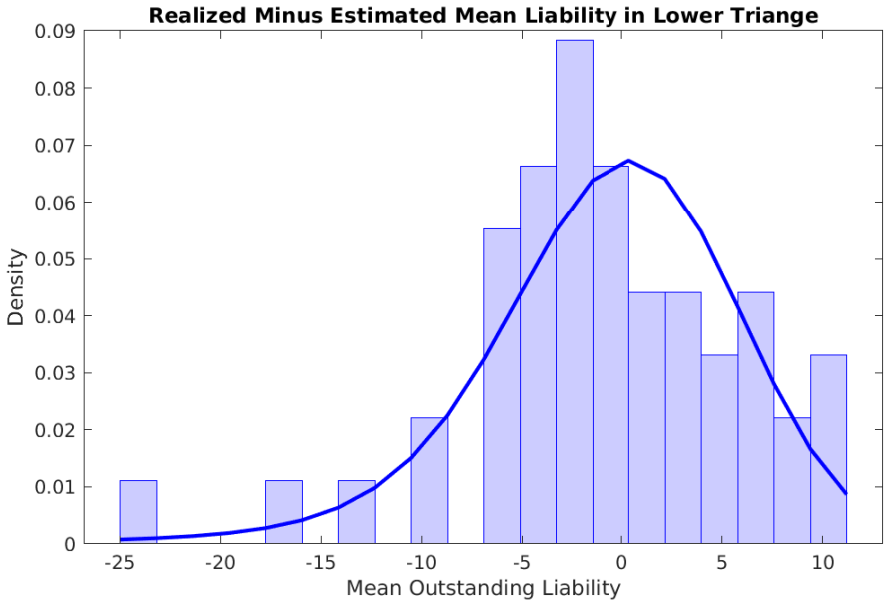


Figure 2.8: Mean outstanding losses calculated from Table 2.3 minus the realized losses in the 50 simulated triangles, fitted to a normal distribution (blue line)

create our objective we use a moment condition of the form

$$h((\tau_1,\tau_2),(\theta_1,\theta_2,\theta_Z):\mathbf{X}) = e^{\tau_1 a^{(1)} Y^{(1)} + \tau_2 a^{(2)} Y^{(2)} + (\tau_1+\tau_2) bZ} - M_{Y^{(1)}|\theta_1}(\tau_1 a^{(1)}) M_{Y^{(2)}|\theta_2}(\tau_2 a^{(2)}) M_{Z|\theta_Z}((\tau_1+\tau_2)b).$$

For the Stable case, we once again consider evaluating using $-X$ and the half-plane containing $\tau_i > 0$, but a practical problem immediately presents itself. As alluded to in Section 2.4.4, $\mathcal{K}^{-1/2}h$ produces an objective function with little curvature around the optimum, being basically flat for much of its support. Our current workaround to produce the results in Table 2.4 is as follows. First, we make use of a change of variable to concentrate as many of the quadrature points as possible where the function has the most substantial value. Second, we multiply the corresponding objective by a large number. This guarantees that any optimizing algorithm unable to distinguish between a flat region and a very lightly curved one would not terminate prematurely. That being said, formally understanding the relationship between moment conditions and the resulting objective, along with a more systematic approach to numerical optimization, remains an avenue for future work.

The results here are not as precise as in the previous section. However, virtually all parameters are within a standard deviation. Importantly, the CGMM is able to capture the systematic parameters motivating the exercise in the first place.

It goes without saying that the results for the single LoB cases seem more accurate. One aspect of the CGMM we lament is the degree of subjectivity of choices in the methodology. For example, to guarantee that the regularization term and MSE in Eq. (2.35) are balanced (guaranteeing that we converge to the "correct" solution as $\lambda \to 0$), we need to standardize the columns of $\mathcal{K}$. In the multiple LoB case, $\mathcal{K}$ is rather sparse and hence very sensitive to the choice of standardization, dramatically affecting the results. This is an example of the kind of problem we encounter in the multiple LoB case that is either not present or at least not as important as in the single LoB case. Furthermore, practical results (especially vs the MLE) would still take on the order of hours or days to produce, making thorough experimentation slow. That said, we are confident that finding the proper regularization is simply a matter of trial and error.

In the next section, we apply the same method to a well-studied data set, and succeed in producing seemingly favourable results. This suggests even as a preliminary conclusion that the CGMM can compete with existing methods.

Line 1				Line 2			
Parameter	True	Median	SD	Parameter	True	Median	SD
η_1^1	5.00	4.42	0.35	η_1^2	4.00	3.78	0.52
η_2^1	5.00	4.48	0.46	η_2^2	4.00	3.68	0.44
η_3^1	5.00	4.44	0.35	η_3^2	4.00	3.80	0.52
η_4^1	5.00	4.40	0.39	η_4^2	4.00	3.94	0.47
η_5^1	5.00	4.48	0.60	η_5^2	4.00	3.89	0.52
η_6^1	5.00	4.41	0.47	η_6^2	4.00	3.89	0.40
η_7^1	5.00	4.43	0.40	η_7^2	4.00	3.85	0.39
η_8^1	5.00	4.44	0.39	η_8^2	4.00	3.98	0.40
η_9^1	5.00	4.55	0.63	η_9^2	4.00	3.92	0.85
η_{10}^1	5.00	4.66	0.31	η_{10}^2	4.00	4.02	0.43
ν_1^1	1.00	1.00	0.00	ν_1^2	1.00	1.00	0.00
ν_2^1	0.93	1.01	0.08	ν_2^2	0.95	0.96	0.09
ν_3^1	0.87	0.95	0.07	ν_3^2	0.90	0.92	0.10
ν_4^1	0.80	0.88	0.07	ν_4^2	0.85	0.85	0.11
ν_5^1	0.73	0.81	0.08	ν_5^2	0.80	0.81	0.08
ν_6^1	0.67	0.73	0.06	ν_6^2	0.75	0.76	0.09
ν_7^1	0.60	0.65	0.05	ν_7^2	0.70	0.69	0.09
ν_8^1	0.53	0.60	0.08	ν_8^2	0.65	0.65	0.11
ν_9^1	0.47	0.51	0.06	ν_9^2	0.60	0.56	0.10
ν_{10}^1	0.40	0.43	0.17	ν_{10}^2	0.55	0.54	0.22
γ^1	0.20	0.16	0.10	γ^2	0.30	0.15	0.09

Systematic Parameters			
Parameter	1.00	Median	SD
μ	0.10	0.17	0.12
σ	0.10	0.12	0.07

Table 2.4: The results of CGMM estimation from 50 joint pairs of triangles

2.6 Real-world Data Analysis

In this section, we use the data first used in [Zhang and Dukic, 2013] and [Avanzi et al., 2016] from the Pennsylvania National Insurance Group (Schedule P, see Table E.1 and E.2) to study multivariate and copula Tweedie-based loss models from a Bayesian perspective. We estimate every model parameter using our multivariate CGMM methodology except for the Tweedie power parameter p, which we set *a priori* at $p = 1.32$. This is the value found in [Avanzi et al., 2016], derived from an analysis of the likelihood.

One advantage of the Bayesian methods employed in the aforementioned works is that they come equipped with a built-in uncertainty estimate for a single loss-reserve triangle. Calculating the variance of model parameters and confidence intervals is quite laborious in the CGMM. We instead opt to use the parametric bootstrap described in [Wüthrich and Merz, 2008]. To briefly summarize, we estimate the model parameters from the Schedule P data and generate new loss triangles from the results, re-estimating the parameters in this bootstrapped sample and outstanding claims. In Table 2.6, we can see the summary statistics from the bootstrapped samples, with the resulting outstanding claims reserve statistics in 2.7. We find that the results of Table 2.6 are similar to the celebrated chain ladder estimates for η and ν. Given that the Tweedie is in some sense a parametric equivalent to these classic estimators, this is both unsurprising and also a validation of the CGMM estimates.

Table 2.5 summarizes our outstanding estimates alongside previous results using the same data set. There appear to be smaller variances in the CGMM results. A possible reason is the fact that in other methods relatively harsh uniform priors were used in a Bayesian framework; besides, the [Zhang and Dukic, 2013] results use heavier-tailed log-Normal marginals. In light of the simulation results in the previous section, we must also consider the possibility that the CGMM may be systematically underestimating the variance of outstanding claims.

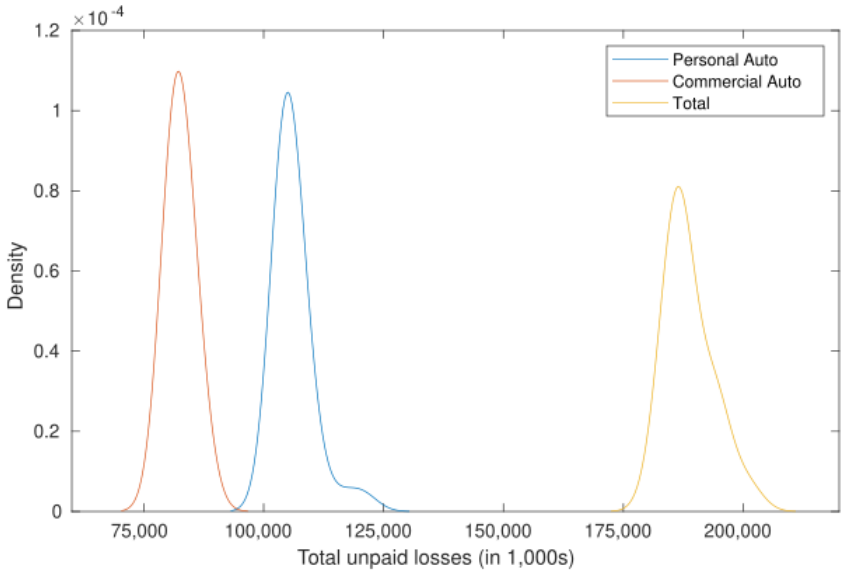


Figure 2.9: Kernel density estimates for outstanding loss reserves, calculated via parametric bootstrap

	Personal			Commercial			Total		
Model	Median	SD	Q(0.99)	Median	SD	Q(0.99)	Median	SD	Q(0.99)
Multivariate Tweedie (CGMM)	104,935	3,750	121,377	82,038	2,052	87,715	187,542	4,786	201,928
Multivariate Tweedie (Bayes)	103,374	9,373	127,075	88,385	9,029	112,258	192,148	13,780	226,110
Clayton Copula Tweedie (Bayes)	103,674	18,742	166,187	91,067	15,820	135,924	194,741	28,376	283,931
Gaussian Copula Tweedie (Bayes)	107,930	21,502	172,161	92,773	17,902	147,734	200,703	31,333	295,900

Table 2.5: Comparison of estimated outstanding reserves from our work (CGMM), [Avanzi et al., 2016] (Bayesian estimation applied to the multivariate Tweedie), and [Zhang and Dukic, 2013] (Bayesian estimation of a model with Tweedie marginals and copula dependence)

2.7 Conclusion

In this work, we have motivated the class of ABRMs with no closed-form PDF and proposed a novel application of the CGMM to estimate model parameters. Our methods are efficient and use moment generating functions alone. Though we primarily focus on Tweedie and Stable marginals in theory we are not limited to these distributions. We are also not bound by the form of ABRMs considered here. In the future, a more realistic model may

	Personal Auto					Commercial Auto			
	Median	SD	Q(0.05)	Q(0.95)		Median	SD	Q(0.05)	Q(0.95)
$\eta_1^{(1)}$	0.2635	0.0071	0.2506	0.2706	$\eta_1^{(2)}$	0.1478	0.0142	0.1264	0.1675
$\eta_2^{(1)}$	0.2429	0.0067	0.2328	0.2512	$\eta_2^{(2)}$	0.1528	0.0107	0.1335	0.1653
$\eta_3^{(1)}$	0.2295	0.0058	0.2250	0.2430	$\eta_3^{(2)}$	0.1517	0.0148	0.1291	0.1687
$\eta_4^{(1)}$	0.2299	0.0082	0.2233	0.2516	$\eta_4^{(2)}$	0.1432	0.0180	0.1203	0.1731
$\eta_5^{(1)}$	0.2324	0.0062	0.2271	0.2457	$\eta_5^{(2)}$	0.1611	0.0200	0.1286	0.1830
$\eta_6^{(1)}$	0.2268	0.0108	0.2191	0.2424	$\eta_6^{(2)}$	0.1457	0.0138	0.1327	0.1797
$\eta_7^{(1)}$	0.2453	0.0042	0.2385	0.2527	$\eta_7^{(2)}$	0.1804	0.0105	0.1649	0.1919
$\eta_8^{(1)}$	0.2495	0.0042	0.2400	0.2526	$\eta_8^{(2)}$	0.1844	0.0084	0.1655	0.1930
$\eta_9^{(1)}$	0.2946	0.0068	0.2887	0.3008	$\eta_9^{(2)}$	0.1982	0.0054	0.1900	0.2033
$\eta_{10}^{(1)}$	0.2792	0.0018	0.2773	0.2815	$\eta_{10}^{(2)}$	0.2107	0.0025	0.2061	0.2140
$\nu_1^{(1)}$	1.0000	0.0000	1.0000	1.0000	$\nu_1^{(2)}$	1.0000	0.0000	1.0000	1.0000
$\nu_2^{(1)}$	0.9951	0.0044	0.9896	0.9999	$\nu_2^{(2)}$	0.9999	0.0023	0.9944	1.0000
$\nu_3^{(1)}$	0.5828	0.0045	0.5787	0.5883	$\nu_3^{(2)}$	0.8753	0.0036	0.8685	0.8770
$\nu_4^{(1)}$	0.3518	0.0100	0.3393	0.3691	$\nu_4^{(2)}$	0.7074	0.0044	0.7036	0.7220
$\nu_5^{(1)}$	0.1977	0.0139	0.1802	0.2194	$\nu_5^{(2)}$	0.4635	0.0075	0.4578	0.4744
$\nu_6^{(1)}$	0.0932	0.0100	0.0839	0.1089	$\nu_6^{(2)}$	0.1915	0.0111	0.1875	0.2135
$\nu_7^{(1)}$	0.0310	0.0038	0.0223	0.0333	$\nu_7^{(2)}$	0.1428	0.0045	0.1375	0.1481
$\nu_8^{(1)}$	0.0153	0.0045	0.0082	0.0224	$\nu_8^{(2)}$	0.0517	0.0048	0.0446	0.0567
$\nu_9^{(1)}$	0.0142	0.0043	0.0067	0.0197	$\nu_9^{(2)}$	0.0222	0.0026	0.0179	0.0250
$\nu_{10}^{(1)}$	0.0008	0.0193	0.0002	0.0715	$\nu_{10}^{(2)}$	0.0003	0.0018	0.0001	0.0010
$\gamma^{(1)}$	0.0010	0.0001	0.0008	0.0011	$\gamma^{(2)}$	0.0011	0.0001	0.0009	0.0013

Systematic Parameters

α	0.0073	0.0013	0.0054	0.0100
β	0.0027	0.0003	0.0023	0.0030

Table 2.6: Parameter estimates based on parametric bootstrap

incorporate separate scale parameters for each cell or multiple systematic components for a richer dependence structure.

We believe the results we have obtained in simulations and in an application to Schedule

	Personal Auto		Commercial Auto		Total	
Accident Year	Mean	SD	Mean	SD	Mean	SD
2	149.89	302.47	31.79	22.73	181.68	318.62
3	654.18	599.86	314.57	75.74	968.75	636.11
4	1,572.97	937.92	999.35	114.05	2,572.32	985.10
5	3,410.71	1,364.64	3,023.16	167.58	6,433.87	1,389.11
6	7,946.91	2,154.56	7,377.44	364.11	15,324.35	2,314.57
7	20,212.86	3,290.60	18,393.45	876.04	38,606.32	3,738.48
8	39,242.16	3,837.55	35,019.65	1,392.14	74,261.82	4,661.08
9	62,104.63	3,415.46	55,170.37	1,646.02	117,274.99	4,223.66
10	106,151.94	3,750.46	82,571.57	2,052.32	188,723.51	4,786.30

Table 2.7: Cumulative Outstanding Claims Reserves by Accident Year, in $1000s

P data show promise for the CGMM in insurance applications. By constructing moment conditions that produce convex objective functions, we remove any need for overspecialized optimization packages or knowledge. Estimation remains generic and independent of the chosen model. That said, the art of solving ill-posed integral equations is challenging, and in the more complicated multi-LoB models further research is likely necessary.

Chapter 3

Risk Aggregation and Allocation

With a proper model of losses and parametric estimates for said losses in hand, we now turn to the question of modelling the risk capital. In the specific case of our Stable ABRM, confident in our CGMM estimation, we can ask (i) what is a prudent level of total risk capital and (ii) how can we meaningfully allocate it among business lines?

3.1 Introduction

Let χ be the set of random variables that represent the random liabilities of insurance contracts. Call the elements X of χ *risks*. Mathematically, $X \in \chi$ is a function on some probability space, measurable with respect to the sigma measure. *Risk measures* are then functionals used to assign finite real values or infinite values (corresponding to the *risk capital*) to elements of χ:

$$H[X] : \chi \to \mathbb{R} \cup \{\infty\}.$$

Closely related to risk functionals are Premium Calculation Principles (PCPs) denoted by $\pi[X]$, meaning the actual price the insurer charges for coverage of a risk. Often a PCP is explicitly derived from a risk measure. For example, indifference premiums are often

derived by solving $H[\pi[X]-X]=0$, which (assuming translation invariance) simply yields $\pi=-H[-X]$. Aside from pricing, the calculation of risk capital is important for shareholder and management purposes such as solvency requirements ([McNeil et al., 2005]). Infrequent but large losses in insurance and finance have led to the widespread adoption of tail-based measures of risk. Most prominent among these are Value at Risk (VaR), Expected Shortfall (ES) and the closely related Tail Conditional Expectation (TCE).

Let $X \in \chi$ have a Cumulative Distribution Function (CDF) given by $F_X(x)$. Given some prudence level $q \in (0,1)$, the VaR is simply defined as the q-th quantile of the distribution of X:

$$VaR_q[X] = x_q = \inf\{x | F_X(x) > q\},$$

or simply $x_q = F_X^{-1}(q)$ for continous distributions.

The ES attempts to capture the mean loss over a threshold by averaging the VaR over all prudence levels greater than or equal to q:

$$ES_q[X] = \frac{1}{1-q}\int_q^\infty VaR_\gamma[X]d\gamma.$$

Finally, when F_X is continuous the ES coincides with the more intuitive TCE, given by

$$TCE_q[X] = \mathbb{E}[X \mid X > x_q].$$

In other words, like ES, TCE measures the mean loss over some threshold for a given prudence level. The ES/TCE is often touted as an alternative to VaR because it is *coherent* in the sense of Artzner [Artzner et al., 1999]. The VaR is however coherent for elliptical loss variables[1].

Given a risk measure H and n random variables $X^{(1)},...,X^{(n)}$ representing the total losses from n individual business lines, the aggregate risk capital is $H[S]$ where $S = X^{(1)} + ...+X^{(n)}$. The calculation of S and $H[S]$ is the first step in any risk management framework

[1]This can easily be shown using properties of elliptical distributions and the triangle inequality.

and is mandatory under insurance and banking regulations (e.g. SolvencyII/Swiss Solvency Test, Basel III).

Corresponding to each choice of risk measure are capital allocation rules. Once $H[S]$ has been computed, it is natural to ask how the individual $X^{(i)}$'s contribute to $H[S]$. Consider again a financial institution with n business lines and corresponding aggregate loss $S = \sum_{i=1}^{n} X^{(i)}$. The capital allocated to each line is denoted $A[X^{(i)}, S]$. Given S and risk capital $H(S)$, the question is: how to appropriately calculate the capital allocations $A[X^{(i)}, S]$ such that $H[S] = \sum_{i=1}^{n} A[X^{(i)}, S]$? For profitability testing and other internal analyses (e.g. cost sharing, pricing [Venter, 2004]) it is important to know which business lines contribute the most to aggregate risk.

For the TCE, there is an extremely natural choice of allocation rule. Taking advantage of the additivity of expectation, we have

$$\mathbb{E}[S|S > s_q] = \sum_{i=1}^{n} \mathbb{E}[X^{(i)}|S > s_q].$$

Thus, given an aggregate loss that exceeds a prudence level, the k^{th} allocation is its expected contribution to this excess.

Generally, calculating allocations is very involved even when given a specific risk measure like the TCE. First there is the non-trivial task of determining the stochastic properties of S ([Miles et al., 2019] and references therein). Second, in addition to the potentially complicated relationship between the $X^{(i)}$'s and S, there is often a dependence structure between the different $X^{(i)}$'s. That being said, in certain cases this problem can be reduced in complexity to determining the aggregate risk. A good example of this was first put forward by Panjer [Panjer, 2002] for the TCE risk measure when bivariate Normal losses were assumed. In such a case allocations were shown to be linear in the aggregate risk capital:

$$A[X^{(i)}, S] = \mathbb{E}[X^{(i)}] + \gamma_{X^{(i)},S} \left(H(S) - \mathbb{E}[S] \right)$$

where $\gamma_{X^{(i)},S}$ is the typical regression coefficient, namely $\mathbb{C}\mathbb{o}\mathbb{v}[X^{(i)}, S]/\mathbb{V}\mathbb{a}\mathbb{r}[S]$ where $\mathbb{C}\mathbb{o}\mathbb{v}[X, Y]$ and $\mathbb{V}\mathbb{a}\mathbb{r}[X]$ are the covariance and variance respectively. This is obviously reminiscent of

exposure to systematic risks *à la* CAPM in finance. The assumption of normality was heuristically justified by regarding business lines as the sum of enough individual policies to invoke the Central Limit Theorem (CLT). Later results showed this linear risk decomposition could be extended to more general elliptical models [Landsman and Valdez, 2003]. More recently, [Furman and Zitikis, 2010] and then [Furman et al., 2018] have shown that such a decomposition is possible under very general considerations. This has led to the development of an insurance analogue to the CAPM known as the Weighted Insurance Pricing Model (WIPM) ([Furman and Zitikis, 2017]). Crucially, [Furman et al., 2018] shows that the linearity of elliptical conditional expectation is key in decomposing the allocations into linear functions of aggregate risk capital. There are however limitations to using elliptically distributed losses. For more general criticisms on the topic of elliptical linear dependence in the risk management literature, see [Embrechts et al., 2002] and [Bilodeau, 2004].

The instinct to use the CLT in risk aggregation is broadly correct. But in order to approximate S via the classical CLT the following must be true: i) n must be large, ii) the $X^{(i)}$'s must be sufficiently similar and iii) the moments of $X^{(i)}$'s must be "well behaved" (i.e. $\mathbb{V}\mathrm{ar}[X^{(i)}] < \infty$, small skewness, etc.). In practice, at least one of these usually fails. Insurance losses for example are often totally skewed and highly heterogeneous, and can have infinite second moments. In finance, there is a perennial thread in the literature disputing the use of Normal models (starting with [Mandelbrot, 1967]; for a more modern reference see [Rachev et al., 2011]). The inapplicability of Normal models may seem discouraging in applying the CLT but the Normal distribution is *not* the only possible limit to schemes of sums of normalized random variables. The class of Stable distributions and the GCLT can address the abovementioned shortcoming.

The organization of Chapter 3 is as follows. In Sections 3.2 and 3.3 we will outline our proposed model along with the application of the GCLT. We will provide a more detailed background on the linear decomposition of allocations in Section 3.4 vis-à-vis Stable distributions. Section 3.5 will include a formula for the TCE in the Stable case and the corresponding allocations. We end the chapter with some notes on how to compute the quantities related to stable distributions in Section 3.6 followed by concluding remarks

in Section 3.7. All technical proofs are included in Appendix D; for more background on Stable random variables see Appendix A.3.

3.2 Multivariate Stable Random Variables

Recall that we can model the losses of the n business lines of an insurance company by the vector $\mathbf{X} = (X^{(1)}, ..., X^{(n)})$ and the total losses by $S = X^{(1)} + ... + X^{(n)}$. The quantity S, being the sum of Stable marginals in a Stable vector, will also be Stable[2]. Likewise, $(X^{(i)}, S)$ will be a bivariate Stable vector with results that will ultimately make the calculations of risk capital and allocations relatively straightforward (Section 3.4.3). Before moving on to the questions of risk capital and allocation, in this Section we describe some of the unique properties of Stable random vectors.

Stable random vectors can be defined in a similar way to definition A.3.2. However, more involved results will require the introduction of some extra machinery. General Stable d-dimensional random vectors that preserve the essential properties of Stable distributions are determined by a standard shift vector $\boldsymbol{\mu}$ and a finite measure Λ on the Borel sets of the d-dimensional unit sphere. The measure Λ is often termed a "spectral measure" in the literature. This spectral measure determines the dependence structure of the random vector's components. As in the univariate case, in only a few cases will a closed-form pdf of the distribution actually exist.

Definition 3.2.1 (Stable Random Vector)**.** $\mathbf{X} \in \mathbb{R}^d$ *is a Stable vector if it has characteristic function*

$$\phi_{\mathbf{X}|\mu,\Lambda}(\boldsymbol{\tau}) = \mathbb{E}\left[\exp\{i\mathbf{X}^\top\boldsymbol{\tau}\}\right] = \exp\left\{-\int_{S_d} \Upsilon_\alpha(\boldsymbol{\tau}^\top\mathbf{s})\Lambda(d\mathbf{s}) + i\mu^\top\boldsymbol{\tau}\right\}, \tag{3.1}$$

denoted $\mathbf{X} \backsim S_\alpha(\Lambda, \mu)$*, where*

[2]The Stables share this property with the Normal class. However, in the same way that a vector with Normal marginals may not be distributed as a multivariate Normal (for a classic counterexample see [Melnick and Tenenbein, 1982]), Stable vectors have more properties than those of a vector with straightforward Stable marginals.

1. Λ *is a spectral measure on the unit sphere* S_d *determining the dependence structure and distribution of* $\mathbf{X}$

2. $\boldsymbol{\mu}$ *is a standard location vector*

3. Υ_α *is the CF of a totally skewed univariate stable variable (see A.21):*

$$\Upsilon_\alpha(u) = \begin{cases} |u|^\alpha(1 - ia\,\text{sign}\,(u)) & \alpha \neq 1 \\ |u|(1 + i\frac{2}{\pi}\,\text{sign}\,(u)\ln(u)) & \alpha = 1 \end{cases}$$

Stable random vectors defined by (3.1) do share a nice property with Normals and Normal-scale mixtures: linear combinations of their marginals are univariate Stable. Let $\boldsymbol{\mu} = 0$ and define the following:

$$\sigma^\alpha(\boldsymbol{\tau}) = \int_{S_d} |\boldsymbol{\tau}^\top \mathbf{s}|^\alpha \Lambda(d\mathbf{s}) \tag{3.2}$$

$$\beta(\boldsymbol{\tau}) = \frac{1}{\sigma^\alpha(\boldsymbol{\tau})} \int_{S_d} \text{sign}\,(\boldsymbol{\tau}^\top \mathbf{s}) |\boldsymbol{\tau}^\top \mathbf{s}|^\alpha \Lambda(d\mathbf{s}) \tag{3.3}$$

$$I_{\mathbf{X}}(\boldsymbol{\tau}) = \begin{cases} \sigma^\alpha(\boldsymbol{\tau})\,(1 - i\beta(\boldsymbol{\tau})\tan(\frac{\pi\alpha}{2})) & \alpha \neq 1 \\ \sigma(\boldsymbol{\tau})\left(1 - i\int_{S_d} \boldsymbol{\tau}^\top \mathbf{s} \ln(\boldsymbol{\tau}^\top \mathbf{s})\Lambda(d\mathbf{s})\right) & \alpha = 1. \end{cases} \tag{3.4}$$

Then $\boldsymbol{\tau}^\top \mathbf{X}$ is a one-dimensional random variable with characteristic function

$$\mathbf{E}\exp\{iu\mathbf{X}^\top \boldsymbol{\tau}\} = \exp\{-I_{\mathbf{X}}(u\boldsymbol{\tau})\}.$$

Breaking from the multivariate Normal, the converse is not always true:

Theorem 3.2.1 ([Samorodnitsky, 2017])**.** *Let* $\mathbf{X}$ *be a random vector in* $\mathbb{R}^N$*. If* $\forall \mathbf{t}, \mathbf{t}^\top \mathbf{X}$ *is stable with* $\alpha > 1$ *then* $\mathbf{X}$ *is a Stable vector.*

Obviously, a potential issue for the *practical* usage of Stable random vectors is the specification of the dependence structure and therefore the spectral measure $\Lambda(\cdot)$. Fortunately, the spectral measure can be naturally approximated by a much simpler discrete object to arbitrary precision:

Definition 3.2.2 (Discrete Spectral Measure)**.** *Given a set of points* $\mathbf{s}_i \in S_d$ *and a corresponding set of weights* $\gamma_i > 0$*, we can define a spectral measure* $\Lambda : S_d \to \mathbb{R}^+$ *as follows:*

$$\Lambda(\cdot) = \sum_{i=1}^{m} \gamma_i \delta_{\mathbf{s}_i}(\cdot). \tag{3.5}$$

For a proof and means of construction of this measure, see [Byczkowski et al., 1993]. Given that discrete measures appear naturally in many contexts ([Nolan, 2003]) and are much easier to manipulate, we will simply consider them on their own. Additionally, discrete measures give rise to a powerful stochastic representation. A nice feature of elliptical distributions is the ease with which they can be manipulated under linear transformations. Fortunately, this is still true when a Stable vector is described by a discrete spectral measure.

Assume $\mu = 0$ and define $Z^{(j)} \sim S_\alpha(1, 1, 0)$ i.i.d. and

$$\mathbf{X} := \begin{cases} \sum_{j=1}^{n} \mathbf{s}_j \gamma_j^{1/\alpha} Z^{(j)} & \alpha \neq 1 \\ \sum_{j=1}^{n} \mathbf{s}_j \gamma_j (Z^{(j)} + \frac{2}{\pi} \ln \gamma_j) & \alpha = 1. \end{cases} \tag{3.6}$$

It is easy to show that the characteristic function of $\mathbf{X}$ is

$$\phi_{\mathbf{X}}(\boldsymbol{\tau}) = \exp\left(- \sum_{i=1}^{n} \gamma_i \Upsilon_\alpha(\boldsymbol{\tau}^\top \mathbf{s_i}) \right). \tag{3.7}$$

That is, X has a characteristic function of the form (3.7) if the Lévy measure in (3.1) is a discrete measure. Once again we will exclusively look at the $\alpha \neq 1$ case. Without loss of generality we will include in the set of $(\gamma, \mathbf{s})$ the twin pairs $(\gamma_i, \mathbf{s_i})$ and $(\gamma_{-i}, \mathbf{s_{-i}}) = (\gamma_{-i}, -\mathbf{s_i})$, so that

$$\begin{aligned} \mathbf{X} &= \sum_{j=1}^{2n} \gamma_j^{1/\alpha} \mathbf{s}_j Z^{(j)} \\ &= [\mathbf{s}_1 \dots \mathbf{s}_n, -\mathbf{s}_1 \cdots - \mathbf{s}_n] \mathrm{diag}[\gamma_{+1}, \dots, \gamma_{+n}, \gamma_{-1} \dots, \gamma_{-n}]^{\frac{1}{\alpha}} \mathbf{Z} \\ &= \mathbf{S}_{d \times 2n} \mathbf{D}^{\frac{1}{\alpha}}_{2n \times 2n} \mathbf{Z}. \end{aligned}$$

If $\gamma_i = \gamma_{-i}$ then we will say that Λ is *symmetric*. It is tempting to look at the expression $\mathbf{X} = \mathbf{S}_{d\times 2n}\mathbf{D}^{\frac{1}{\alpha}}_{2n\times 2n}\mathbf{Z}$ and conclude that $\mathbf{X}$ is elliptical if Λ is symmetric. In general, however, this will not be true:

$$\begin{aligned}\mathbf{X} &= \sum_{j=1}^{2n} \mathbf{s}_j(\gamma_{+j}^{1/\alpha} Z^{+j} - \gamma_{-j}^{1/\alpha} Z^{-j}) \\ &= \frac{1}{2^{1/\alpha}} \sum_{i=1}^{n} \mathbf{s}_i Z_i^{sym} \\ &= \frac{1}{2^{1/\alpha}} \ [\mathbf{s}_1 \dots \mathbf{s}_n]\mathbf{Z}^{sym}.\end{aligned}$$

Each marginal in $\mathbf{Z}^{sym}$ has $\beta = 0$. Only if $[\mathbf{s}_1 \dots \mathbf{s}_n]$ is a Cholesky factorization of a positive definite matrix do we have an elliptical Stable vector.

We conclude this section on Stable vectors by stating a theorem that will be highly relevant in the following sections.

Theorem 3.2.2. *([Samorodnitsky, 2017]) Let $(X_2, X_1) \backsim S_\alpha(\Lambda, 0)$ be a jointly distributed Stable random vector. Then*

$$\mathbb{E}[X_2|X_1 = x] = \kappa_{2,1}x + a\sigma_1^\alpha(\lambda_{2,1} - \beta_1\kappa_{2,1})\frac{h(x)}{\pi f_{X_1}(x)}$$

where

$$\begin{aligned}\kappa_{2,1} &= \frac{[X_2, X_1]}{\sigma_1^\alpha} = \frac{\int_{S_d} s_2|s_1|^{\alpha-1}\text{sign}(s_1)\Lambda(d\mathbf{s})}{\sigma_1^\alpha} \\ \lambda_{2,1} &= \frac{\int_{S_d} s_2|s_1|^{\alpha-1}\Lambda(d\mathbf{s})}{\sigma_1^\alpha} \\ h(x) &= \int_0^\infty e^{-\sigma_1^\alpha t^\alpha} t^{\alpha-1}\cos(xt - a\beta_1\sigma_1^\alpha t^\alpha)dt\end{aligned}$$

and $f_{X_1}(x)$ is the pdf of X_1.

3.3 Multivariate Stable Insurance Losses

In Section 2.3.2 we discussed briefly how and why a multivariate Stable model may be a natural choice for heavy-tailed insurance data. Given that it is a central object of study in this work we will take some time to examine this model and its possible drawbacks more fully. Ultimately we want to be able to approximate N business lines with a multivariate Stable vector. Consider an insurance company and assume for simplicity that it has two business lines. Further suppose each line sells identical policies whose losses can be (at least asymptotically) described as Pareto-tailed random variables $L^{(1)}$ and $L^{(2)}$ with tail parameters α_1 and α_2 respectively[3].

Assuming a roughly equal number of policies, n, are sold for each line, we want to model the total aggregate loss random vector:

$$\mathbf{X}_n = \sum_{i=1}^{n} \begin{pmatrix} L_i^{(1)} \\ L_i^{(2)} \end{pmatrix}$$

Obviously we would like to show that we can model $\mathbf{X}_n$ as a Stable random *vector*. As we will see, however, the only case in which we can do so without introducing degenerate marginals is when $\alpha_1 = \alpha_2$. Since $\alpha > 1$ we can set $q_n = (n^{1-\frac{1}{n\alpha}})(\mu_1, \mu_2)^T$ where μ_i are the respective means. Consider the linear combinations of the centred normalized losses:

$$\begin{aligned} \frac{1}{n^{1/\alpha}} \boldsymbol{\tau}^\top \mathbf{X_n} - q_n &= \frac{\sum_{i=1}^{n} t_1(L_i^{(1)} - \mu_1) + t_2(L_i^{(2)} - \mu_1)}{n^{1/\alpha}} \\ &= \frac{\sum_{i=1}^{n} (t_1 L_i^{(1)} + t_2 L_i^{(2)}) - (t_1\mu_1 + t_2\mu_2)}{n^{1/\alpha}}. \end{aligned}$$

To make sense of the above expression we will make use of a result of [Tucker et al., 1968].

Theorem 3.3.1 ([Tucker et al., 1968], Lemma 3)**.** *Consider two random variables Z_1 and*

[3]Specifically, $1 - F_X(x) \sim \frac{1}{x^\alpha}$ as $x \to \infty$ and $F_X(x) \sim \frac{1}{|x|^\alpha}$ as $x \to -\infty$.

Z_2 with Pareto tails of index α_1 and α_2. We have that

$$\frac{1-F_{Z_1+Z_2}(kx)}{1-F_{Z_1+Z_2}(x)} \sim x^{-\min\{\alpha_1,\alpha_2\}} \text{ as } k \to \infty.$$

This theorem implies that $t_1 L_i^{(1)} + t_2 L_i^{(2)}$ will have tail index α, and as $n \to \infty$, by the GCLT we have

$$\frac{1}{n^{1/\alpha}} \boldsymbol{\tau}^\top \mathbf{X}_n - q_n \sim S_\alpha(\sigma(\tau), \beta(\tau), \sigma(\tau)). \tag{3.8}$$

This is true for any linear combination and so $\frac{1}{n^{1/\alpha}} \mathbf{X}_n - \mathbf{q}_n$ weakly converges to a Stable random vector by Theorem 3.2.1 and the Cramer–Wold theorem.

Assume now that $\alpha_1 \neq \alpha_2$. In this case the tail index of $\boldsymbol{\tau}^\top \mathbf{X_n}$ will depend on $\boldsymbol{\tau}$. For $\boldsymbol{\tau}$ where $t_2, t_1 \neq 0$ the previous picture does not change significantly; by Theorem 3.3.1 the tail of $t_1 \ell_i^{(1)} + t_2 \ell_i^{(2)}$ will be $\alpha^* = \min\{\alpha_1, \alpha_2\}$ and we will still have the situation in (3.8).

However, say $\boldsymbol{\tau} = (0,1)^T$ then

$$\frac{1}{n^{1/\alpha^*}} \boldsymbol{\tau}^\top \mathbf{X_n} - q_n = \frac{\sum_{i=1}^n L_i^{(2)} - \mu_2}{n^{1/\alpha^*}} \tag{3.9}$$

$$= \frac{1}{n^{\frac{1}{\alpha^*} - \frac{1}{\alpha_2}}} \frac{\sum_{i=1}^n \ell_i^{(2)} - \mu_2}{n^{1/\alpha_2}} \tag{3.10}$$

$$\sim S_{\alpha_2}(n^{\frac{1}{\alpha_2} - \frac{1}{\alpha^*}} \sigma, \beta, 0) \tag{3.11}$$

$$\to \delta(x) \tag{3.12}$$

and indeed, the proof of Theorem 3.2.1 in [Samorodnitsky, 2017] begins by showing that if all linear combinations of a random vector are Stably distributed then those that are non-degenerate have the same tail index!

In light of this we could envision (for N business lines) a model of the following kind:

$$\mathbf{X}_n = \sum_{j=1}^n \begin{pmatrix} a^{(1)} I_j^{(1)} + b^{(1)} M_j \\ \vdots \\ a^{(N)} I_j^{(N)} + b^{(N)} M_j \end{pmatrix} \tag{3.13}$$

where we have not yet specified our "idiosyncratic" and "market" factors, the $I_j^{(i)}$'s and M_j's, beyond the fact they are Pareto-tailed. As previously mentioned, according to [Embrechts et al., 2013] and others many insurance losses exhibit Pareto-tailed behaviour over a certain threshold. We believe it is prudent from a risk management perspective to choose the tail index of losses to be the minimum such index observed across all business lines. This also allows us to approximate the losses the insurance company faces as a Stable vector *without* the degeneracies encountered in (3.9)–(3.12).

Take all the tail indexes of the losses to be α. Abusing notation, the GCLT implies that

$$\frac{1}{n^{1/\alpha}}\boldsymbol{\tau}^\top \mathbf{X_n} \longrightarrow \boldsymbol{\tau}^\top \mathbf{X} \sim S_\alpha(\sigma(\boldsymbol{\tau}), \beta(\boldsymbol{\tau}), \mu(\boldsymbol{\tau})).$$

If this is true for all $\boldsymbol{\tau}$ and $\alpha > 1$ then $\mathbf{X}$ is a Stable vector by Theorem 3.2.1, and by the Cramer-Wold theorem we have

$$\frac{1}{n^{1/\alpha}}\mathbf{X}_n \longrightarrow \mathbf{X} = \boldsymbol{a} \circ \mathbf{Y} + \boldsymbol{b}Z$$

where we can recover $\mathbf{Y}$ and Z from the CGLT convergence of the I_j and M_j sums. This is the model we introduced in Section 2.5.

3.4 Weighted Insurance Pricing

3.4.1 Weighted Risk Measures

Consider a loss random variable X. We can compute its expectation using the inverse CDF and integrating over probabilities:

$$\mathbb{E}[X] = \int_0^1 F^{-1}(p)dp$$

In order to avoid ruin with probability one, insurers require net premiums to be at least $\mathbb{E}[X]$. The easiest way to do this is to distort probabilities of events in such a way

as to guarantee that the net premiums satisfy this lower bound. That is, we calculate the expectation or net premium under a distorted distribution. This is commonly achieved through a *distortion function* $g : [0,1] \to [0,1]$, an increasing function such that $g(0) = 0$ and $g(1) = 1$. Define for the net premium the class of risk measures

$$H[X, g] = \int_0^1 F^{-1}(p) g'(1-p) dp$$

called distortion risk measures (see [Balbás et al., 2009]). Note that $g'(1-p)$ is non-negative and non-increasing: large losses are emphasized and lossless scenarios are de-emphasized. These distortion risk measures encompass a large class of well-studied risk functionals and corresponding premium calculations.

A similar procedure for achieving the same goal of reweighing the loss probabilities is to directly re-weight the distribution function. Given a random variable S and a weight function w such that $0 < \mathbb{E}[w(S)] < \infty$, we can define the CDF of the weighted distribution as

$$F_{w;S}(s) = \frac{\mathbb{E}[\mathbf{1}\{S \leq s\} w(S)]}{\mathbb{E}[w(S)]}.$$

We can define the weighted risk measures similarly to our definition of the distortion class, as an expectation with respect to the new distribution:

$$H_w[S] = \frac{\mathbb{E}[S w(S)]}{\mathbb{E}[w(S)]}.$$

Note that the class of distortion risk measures are a special case[4]. Additionally, while we will primarily be interested in the case of the TCE ($w(s) = \mathbf{1}\{s > s_q\}$), this class easily recovers other standard risk measures (see [Furman and Zitikis, 2008b, Furman and Zitikis, 2008a]).

[4] $w(s) = g'(\bar{F}_S(s))$

3.4.2 Weighted Allocations

Perhaps the most important and useful property of the class of weighted risk measures is the ease with which corresponding allocation rules can be derived and interpreted. To simplify the analysis, we assume continuous risks, so that

$$dF_{w;S}(s) = \frac{w(s)}{\mathbb{E}[w(S)]} f_S(s)dx.$$

Recall the notation that assigns, for a financial institution with n business lines, losses $X^{(1)}, X^{(2)}, ..., X^{(n)}$ and aggregate loss $S = \sum_{i=1}^{N} X^{(i)}$. Using additivity of expectation,

$$\begin{aligned}
\frac{\mathbb{E}[Sw(S)]}{\mathbb{E}[w(S)]} &= \sum_{i=1}^{n} \frac{\mathbb{E}[X^{(i)}w(S)]}{\mathbb{E}[w(S)]} \\
&= \sum_{i=1}^{n} \iint X^{(i)} \frac{w(S)}{\mathbb{E}[w(S)]} f_{(X,S)}(x,s)dxds \\
&= \sum_{i=1}^{n} \int \left[\int X^{(i)} f_{X|S}(x|s)dx \right] \frac{w(S)}{\mathbb{E}[w(S)]} f_S(s)ds \\
&= \sum_{i=1}^{n} \int \mathbb{E}[X^{(i)}|\, S = s]\, dF_{w;S}(s).
\end{aligned}$$

One can show that the quantity $\int \mathbb{E}[X^{(i)}|\, S = s]\, dF_{w;S}(s)$ satisfies many properties desired in an allocation rule: no unjustified loading, consistency and of course full additivity. Furthermore [Furman and Zitikis, 2008b] shows that it is non-negative and no undercut holds in the TCE case. To that end, we define for a given weight function w the allocation

$$A_w[X,S] = \frac{\mathbb{E}[X^{(i)}w(S)]}{\mathbb{E}[w(S)]} = \int \mathbb{E}[X^{(i)}|\, S = s]\, dF_{w;S}(s).$$

This allocation is easily interpretable in the case that $f_S, f_X \in L^2$. Assume an insurer's preferences or utility for profit and losses is quadratic for each business line. This is a standard assumption in many basic versions of various financial models such as the CAPM

[Panjer, 1998]. We have

$$\min_{a_i} \mathbb{E}\left[(X^{(i)} - a_i)^2 \frac{w(S)}{\mathbb{E}[w(S)]}\right] = \min_{a_i} \mathbb{E}\left[\mathbb{E}\left[(X^{(i)} - a_i)^2 \middle| S\right] \frac{w(S)}{\mathbb{E}[w(S)]}\right]$$
$$= \min_{a_i} \int \mathbb{E}\left[(X^i - a_i)^2 \middle| S = s\right] dF_w(s).$$

We can easily prove that:

$$a_i = \int \mathbb{E}[X^{(i)} |\, S = s]\, dF_w(s).$$

3.4.3 Weighted Allocations given Stable Losses

Having elucidated the necessary properties of both weighted allocations and Stable vectors, we are now ready to derive results for weighted allocations in the Stable case. We begin by referring back to Section 3.4.2 and specifically the derivation of $A_w[X^{(i)}, S]$. Making use of Theorem 3.2.2 we can find a very general form for allocations when the joint losses $X^{(i)}$ are described by a Stable random vector:

$$\begin{aligned}
\mathbb{E}[w(s)]A_w[X^{(i)}, S] &= \int \mathbb{E}[X^{(i)}|S = s]w(s)f_S(s)ds \\
&= \int \left(\kappa_i s + a\sigma_S^\alpha(\lambda_i - \beta_S\kappa_i)\frac{h(s)}{f_S(s)}\right) w(s)f_S(s)ds \\
&= \kappa_i \int s w(s) f_S(s)ds + a\sigma_S^\alpha(\lambda_i - \beta_S\kappa_i) \int \frac{h(s)}{f_S(s)} w(s) f_S(s)ds.
\end{aligned}$$

Where $\kappa_i = \frac{[X_i,S]}{\sigma_S^\alpha}$ and λ_i is similarly defined with respect to X_i and S. This gives us that for *any* appropriate weight function,

$$\boxed{A_w[X^{(i)}, S] = \kappa_i H_w(S) + a\sigma_S^\alpha(\lambda_i - \beta_S\kappa_i)\frac{\int h(s)w(s)ds}{\mathbb{E}[w(S)]}.} \tag{3.14}$$

Before continuing, it is worth making a couple of observations about Eq. (3.14). While we can simplify the task of calculating allocations to the evaluation of the aggregate risk capital $H_w(S)$ and some skewness term (the integral involving $h(s)$), this in itself is not a trivial task. Surprisingly, as we will see, the skewness term disappears, leaving just the risk capital term. That being said, there are in general not many well-known results for risk measures involving Stable Losses. Only *numerical* results are known on $H_w(S)$ for Stable S in the TCE case [Stoyanov et al., 2006]. Specifying Eq. (3.14) for the TCE ($w(x) = \mathbf{1}_{(x>x_q)}$) case will be our focus for the rest of this paper.

As Applied to the Model The quantity $a\sigma_S^\alpha \frac{\int h(s)w(s)ds}{\mathbb{E}[w(s)]}$ is shared for all $X^{(i)}$: the effects of skewness enter the allocation only in the $\lambda_i - \beta_S\kappa_i$ term. Interestingly, our model allows for a non-elliptical dependence while preserving the results of [Landsman and Valdez, 2003]. If the losses $L_i^{(j)}$ are Pareto-tailed and totally skewed then $1 - F_S(s) \sim Cs^{-\alpha}$ and $F_S(s) \sim 0$. In the limit when approximated by a Stable, the marginals will be totally skewed to the right (see Theorem A.3.2). Assume a discrete spectral measure for the vector $(X_1, ..., X_N, S)^T$ where for $\mathbf{s}_j \in S_{N+1}$ we have $\mathbf{s}_j = (s_j^{(1)}, ..., s_j^{(i)}, ...s_j)$. Recalling (3.2) and (3.3), this yields

$$\sigma_S^\alpha = \sum_{j=1}^m |s_j|^\alpha \gamma_j$$
$$\beta_S = \frac{1}{\sigma_S^\alpha} \sum_{j=1}^m |s_j|^\alpha \text{sign}(s_s^{(j)})\gamma_j.$$

Clearly, there is no nonzero support in the spectral measure where $\text{sign}(s_s^{(j)}) = -1$ and

$$\kappa_i = \frac{1}{\sigma_S^\alpha} \sum_{j=1}^m s_j^{(i)} |s_j|^{\alpha-1} \text{sign}(s_s^{(j)})\gamma_j = \frac{1}{\sigma_S^\alpha} \sum_{j=1}^m s_j^{(i)} |s_j|^{\alpha-1}\gamma_j = \lambda_i.$$

So then $(\lambda_i - \beta_S \kappa_i) = 0$, and in the model,

$$A_w[X^{(i)}, S] = \kappa_i H_w[S]. \tag{3.15}$$

Recall that Theorem 3.2.2 considered the case with no location parameters (or means where they exist). One can easily add them in:

$$A_w[X^{(i)}, S] = \mathbb{E}[X^{(i)}] + \kappa_i (H_w[S] - \mathbb{E}[S]). \tag{3.16}$$

Calculating the allocations in this case will only require us to calculate κ_i and $H_w(S)$. In the next section we do just that, providing a result for the TCE in the Stable case and a simple example involving κ.

3.5 Stable TCE and Portfolio Risk Allocation

In this section we will provide a representation of the TCE in the Stable case using the Fox H-functions described in Appendix B to represent the Stable pdf. We will use two lemmas proved in Appendix D. It is worth noting that in [Stoyanov et al., 2006] a formula for the Stable TCE is developed through direct numerical integration, whereas the Fox H-function representation used in this work allows us to leverage the numerical convenience therein (see Section 3.6).

Following the derivation of this representation, we will work with our simple example of an insurance company, using 3.5.3 and (3.15) to compute the allocations. Given that $H_w[S]$ is the TCE when $w(s) = \mathbf{1}_{s>s_q}$, then

$$H_w[S] = \frac{1}{1-q} \int_{s_q}^{\infty} s f_S(s) ds$$

where $\mathbb{E}[w(s)] = \int_{s_q}^{\infty} f_S(s) ds = 1 - u$ and, given the cdf $F_S(s) = P(S \leq s)$, we have $s_q = F_S^{-1}(q)$. To compute such an integral we will need to represent $f_S(s)$ as the inverse Fourier transform of the characteristic function (A.21), entailing a double integral. In order

to do this, will need to state a few results, beginning with expressing the following Laplace transform in terms of a simple H-function.

Lemma 3.5.1 (The Laplace transform of $t^j e^{-bt^\alpha}$)**.**

$$\mathcal{L}\left\{t^j e^{-bt^\alpha}\right\}(x) = \frac{1}{\alpha b^{\frac{j+1}{\alpha}}} H_{1,1}^{1,1}\left[\frac{x}{b^{\frac{1}{\alpha}}}\middle|\begin{matrix}\left(1-\frac{j+1}{\alpha},\frac{1}{\alpha}\right)\\(0,1)\end{matrix}\right].$$

When applying Lemma 3.5.1 we shall often need to take the real part of an H-function with complex arguments. To that end we shall need the following:

Lemma 3.5.2. *For $z_1, z_2 \in \mathbb{C}$, $\nu_1 \in \mathbb{R}$ and $\nu_2 \in \mathbb{R}^+$,*

$$\begin{aligned}&z_1^{\nu_1} H_{1,1}^{1,1}\left[z_2^{\nu_2} x\middle|\begin{matrix}(a_1,A_1)\\(b_1,B_1)\end{matrix}\right]\\&= \pi r_1^{\nu_1}\left(H_{2,2}^{1,1}\left[r_2^{\nu_2}x\middle|\begin{matrix}(a_1,A_1) & \left(\frac{1}{2}-\frac{\theta_1\nu_1}{\pi},\frac{\theta_2\nu_2}{\pi}\right)\\(b_1,B_1) & \left(\frac{1}{2}-\frac{\theta_1\nu_1}{\pi},\frac{\theta_2\nu_2}{\pi}\right)\end{matrix}\right] + iH_{2,2}^{1,1}\left[r_2^{\nu_2}x\middle|\begin{matrix}(a_1,A_1) & \left(1-\frac{\theta_1\nu_1}{\pi},\frac{\theta_2\nu_2}{\pi}\right)\\(b_1,B_1) & \left(1-\frac{\theta_1\nu_1}{\pi},\frac{\theta_2\nu_2}{\pi}\right)\end{matrix}\right]\right)\end{aligned}$$

where $r_i = |z_i|$ and $\theta_i = \arg(z_i)$.

Theorem 3.5.3. *Let $S \sim S_\alpha(\sigma,\beta,0)$, $s \geq 0$, $r = \sqrt{1+(a\beta)^2}$, $\phi = \tan^{-1}(a\beta)$ and $\gamma = \frac{1}{2} - \frac{\phi}{\alpha\pi}$. Then the TCE is given by*

$$TCE_q[S] = \frac{\sigma r^{\frac{1}{\alpha}} H_{2,2}^{1,1}\left[\frac{s_q}{\sigma r^{\frac{1}{\alpha}}}\middle|\begin{matrix}\left(1-\frac{\alpha-1}{\alpha},\frac{1}{\alpha}\right) & (\gamma,\gamma)\\(0,1) & (\gamma,\gamma)\end{matrix}\right]}{1-q}$$

Proof. We need to evaluate the aforementioned double integral:

$$
\begin{aligned}
\int_{s_q}^{\infty} s f_S(s) ds &= \int_{s_q}^{\infty} s \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-its} \phi_X(t) dt \right] ds \\
&= \frac{1}{\pi} \int_{s_q}^{\infty} s \operatorname{Re} \left[\int_{0}^{\infty} e^{-its} e^{-t^\alpha \xi} dt \right] ds \\
&= \frac{1}{\pi} \int_{s_q}^{\infty} s \operatorname{Re} \left[\mathcal{L} \left\{ e^{-t^\alpha \xi} \right\} (is) \right] ds \\
&= \operatorname{Re} \left[\frac{1}{\pi} \int_{s_q}^{\infty} s \mathcal{L} \left\{ e^{-t^\alpha \xi} \right\} (is) \, ds \right] \\
&\qquad\qquad \text{(given the exterior integral is real)} \\
&= \operatorname{Re} \left[\frac{\alpha \xi}{\pi} \mathcal{L} \left\{ t^{\alpha-2} e^{-t^\alpha \xi} \right\} (is) \Big|_{s=\infty}^{s=s_q} \right].
\end{aligned}
$$

Recall Lemma 3.5.1:

$$
\mathcal{L} \left\{ t^{\alpha-2} e^{-t^\alpha \xi} \right\} ((is)) = \frac{1}{\alpha \xi^{\frac{\alpha-1}{\alpha}}} H_{1,1}^{1,1} \left[\frac{is}{\xi^{\frac{1}{\alpha}}} \middle| \begin{matrix} \left(1 - \frac{\alpha-1}{\alpha}, \frac{1}{\alpha}\right) \\ (0,1) \end{matrix} \right].
$$

Fortunately, $H_{1,1}^{1,1} \xrightarrow{z \to \infty} 0$[5], so

$$
\operatorname{Re} \left[\frac{\alpha \xi}{\pi} \mathcal{L} \left\{ t^{\alpha-2} e^{-t^\alpha \xi} \right\} (is) \Big|_{s=\infty}^{s=s_q} \right] = \sigma r^{\frac{1}{\alpha}} H_{2,2}^{1,1} \left[\frac{s_q}{\sigma r^{\frac{1}{\alpha}}} \middle| \begin{matrix} \left(1 - \frac{\alpha-1}{\alpha}, \frac{1}{\alpha}\right) & (\gamma, \gamma) \\ (0,1) & (\gamma, \gamma) \end{matrix} \right].
$$

■

While this characterization may seem uninformative, the Fox H-functions are a class of special functions with many known results. Furthermore, we will present a few natural ways of computing H-functions in Section 3.6.

[5]Refer to the definition of the Fox H-function in Appendix B.1. As the Bromwich path will have all positive real parts, then we can show that the integrand goes to zero and use the Dominated Convergence Theorem.

3.6 Evaluating Fox H-Functions

At this point it is natural to ask how to make use of the H-function representations in Section 3.5 and Appendix B.2. In this section we will detail how one can compute the density function of the univariate Stable distribution via its H-function representation. The same methods can be repurposed for the evaluation of the TCE and other quantities found in Section 3.5 without any major changes.

There are two approaches to consider:

1. Numerically invert the integral transform defining the H-functions.
2. Find an equivalent series representation.

Integral Transform Inversion We will start with the more straightforward approach. Consider the Stable pdf as an H-function:

$$f_X(x) = \frac{1}{\alpha\pi\sigma r^{\frac{1}{\alpha}}} H_{2,2}^{1,1}\left[\frac{x}{\sigma r^{\frac{1}{\alpha}}}\middle|\begin{matrix}(1-\frac{1}{\alpha},\frac{1}{\alpha}) & (1-\gamma,\gamma)\\ (0,1) & (1-\gamma,\gamma)\end{matrix}\right].$$

Interpreting this expression as an inverse Mellin transform and simplifying the integrand using the Euler reflection formula gives

$$f_X(x) = \frac{1}{2\pi i}\frac{1}{\alpha\pi\sigma r^{\frac{1}{\alpha}}}\int_{c-i\infty}^{c+i\infty}\Gamma(s)\,\Gamma\left(\frac{1-s}{\alpha}\right)\sin(\pi[\gamma-\gamma s])(\frac{x}{\sigma r^{\frac{1}{\alpha}}})^{-s}ds. \tag{3.17}$$

In order for the integral in (3.17) to converge, the path of integration must separate the poles of the two Gamma functions in the integrand (it being a Mellin–Barnes integral). The poles of the Gamma functions in (3.17) are $s=-k_1$ and $s=1+\alpha k_2$ for $k_1,k_2\in\mathbb{Z}^+$. The path of integration will be the line in the complex plane running from $c-i\infty$ to $c+i\infty$ with $c\in(0,1)$. In this case, (3.17) corresponds to the usual definition of an inverse Mellin transform.

The transformation $s' = \frac{s-c}{i}$ yields

$$f_X(x) = \frac{1}{2\pi}\frac{1}{\alpha\pi}\int_{-\infty}^{\infty}\Gamma\left(c+is'\right)\Gamma\left(\frac{1-c-is'}{\alpha}\right)\sin(\pi[\gamma-\gamma(c-is')])e^{-(c+is')\ln(x)}ds. \quad (3.18)$$

So (3.17) can also be evaluated as an inverse Fourier transform (or inverse two-sided Laplace transform):

$$f_X(x) = \frac{e^{-c\ln(x/\sigma r^{\frac{1}{\alpha}})}}{\alpha\pi\sigma r^{\frac{1}{\alpha}}}\mathcal{F}^{-1}_{\ln(x/\sigma r^{\frac{1}{\alpha}})}\left[\Gamma\left(c+is'\right)\Gamma\left(\frac{1-c-is'}{\alpha}\right)\sin(\pi[\gamma-\gamma(c-is')])\right].$$

There are several well-known ways of inverting Fourier and Laplace transforms numerically (see e.g. [Kuznetsov, 2013] and references therein). The naive approach is simply to truncate the contour at the points $\pm b$ (or $c \pm ib$ in the original coordinates). This truncated approximation is indeed practical here, yielding

$$|\Gamma(a+ib)|^2 = |\Gamma(a)|^2\prod_{k=0}^{\infty}\frac{1}{1+\frac{b^2}{(a+k)^2}} \quad (3.19)$$

That is, as the imaginary part of the contour increases in magnitude, the complex Gamma function decays rapidly.

The reader can implement these inversions using their preferred numerical methods. Additionally, there are commercial numerical Mellin/Fourier/Laplace inversion packages available for example in computing systems such as Mathematica and Matlab

Series Representations Deriving power series and asymptotic expansions directly from (3.17) is a relatively simple application of complex analysis. Consider the two half-circles in the complex plane formed from the diameter running between $c \pm ib$ where $c \in (0,1)$ and $b \in \mathbb{R}$ (Fig. 3.1).

As we allow $b \to \infty$, the segment of the contour along the diameter will become what we need to evaluate (3.17). Provided the contribution from the arc of the half-circle disappears

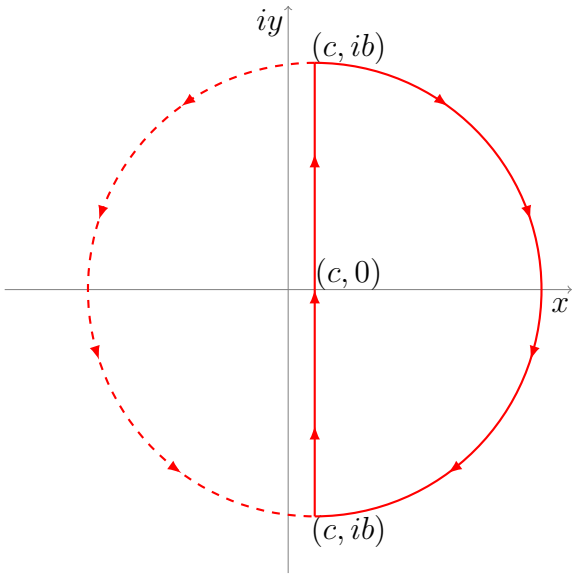


Figure 3.1: Possible contours

in the usual way, can use Cauchy's Residue Theorem to evaluate the contour integral. From (3.19) we know that the modulus of the integrand of (3.17) on the half circle will achieve its maximum on the point intersecting the real line. The remaining issue is to decide which half-circle and therefore what residues to use. Assume $\alpha > 1$ and consider two distinct limits of the Gamma function product:

$$\begin{aligned} &\Gamma(s)\,\Gamma\left(\frac{1-s}{\alpha}\right) \\ &= \frac{\Gamma\left(\frac{1-s}{\alpha}\right)}{\Gamma(1-s)}\frac{\pi}{\sin(\pi[1-s])} \\ &\xrightarrow[s\to-\infty]{\alpha>1} 0 \end{aligned} \qquad \begin{aligned} &\Gamma(s)\,\Gamma\left(\frac{1-s}{\alpha}\right) \\ &= \frac{\Gamma(s)}{\Gamma(1-1/\alpha+s/\alpha)}\frac{\pi}{\sin(\pi[\frac{1-s}{\alpha}])} \\ &\xrightarrow[s\to\infty]{\alpha>1} \infty \end{aligned}$$

We can see that for the integrand in (3.17), if we integrate along the left semi-circle, in the limit of an infinitely large radius the contribution from the left arc disappears by the estimation lemma and we recover (3.17). This allows us via residues to easily derive series

representations:

$$\begin{aligned}
&\frac{1}{\alpha\pi\sigma r^{\frac{1}{\alpha}}}\frac{1}{2\pi i}\left(\int_{\text{diameter}}\tilde{f}(s)ds+\int_{\text{arc}}\tilde{f}(s)ds\right)\\
&\to\frac{1}{\alpha\pi}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\tilde{f}(s)ds\ \ +\ \ 0\\
&=\sum_{k=0}^{\infty}\text{Res}(\tilde{f}(s),-k)
\end{aligned}$$

where

$$\tilde{f}(s)=\Gamma\left(s\right)\Gamma\left(\frac{1-s}{\alpha}\right)\sin(\pi[\gamma-\gamma s])\left(\frac{x}{\sigma r^{\frac{1}{\alpha}}}\right)^{-s}$$

and given $\text{Res}(f,-k)=\lim_{z\to c}(z+k)f(z)$, we have

$$\text{Res}(\tilde{f}(s),-k)=\text{Res}(\Gamma\left(s\right),-k)\Gamma\left(\frac{1+k}{\alpha}\right)\sin(\pi[\gamma+\gamma k])\left(\frac{x}{\sigma r^{\frac{1}{\alpha}}}\right)^{k}.$$

For any $k\in\mathbb{N}$ we can use the recurrence formula:

$$(z+k)\Gamma(z)=\frac{\Gamma(z+k+1)}{z(z+1)\cdots(z+k-1)}.$$

The numerator at $z=-k$ is $\Gamma(-k+k+1)=\Gamma(1)=1$ and the denominator is $(-1)^k k!$. So the residues of the gamma function at those points are

$$\text{Res}(\Gamma,-k)=\frac{(-1)^k}{k!}$$

which finally gives

$$\begin{aligned}
f_X(x)&=\frac{1}{\alpha\pi\sigma r^{\frac{1}{\alpha}}}H_{2,2}^{1,1}\left[\frac{x}{\sigma r^{\frac{1}{\alpha}}}\middle|\begin{matrix}(1-\frac{1}{\alpha},\frac{1}{\alpha}) & (1-\gamma,\gamma)\\ (0,1) & (1-\gamma,\gamma)\end{matrix}\right]\\
&=\frac{1}{\alpha\pi\sigma r^{\frac{1}{\alpha}}}\sum_{k=0}^{\infty}\frac{\Gamma\left(\frac{1+k}{\alpha}\right)\sin(\pi[\gamma+\gamma k])}{k!}\left(-\frac{x}{\sigma r^{\frac{1}{\alpha}}}\right)^{k},\qquad\alpha>1.
\end{aligned}$$

For $|x| \leq 1$ this series will converge rather slowly as it will take many terms for the gamma function to overpower the x^k term. This would make some kind of asymptotic series desirable. Interestingly the series for the $\alpha < 1$ case is such a series.

If we instead examine the opposite half circle (with residues $s = 1 + \alpha k$) we would derive the series for the case of $\alpha < 1$:

$$\operatorname{Res}(\tilde{f}(s), 1+\alpha k) = \Gamma(s) \operatorname{Res}\left(\Gamma\left(\frac{1+k}{\alpha}\right), 1+\alpha k\right) \sin(\pi[\gamma - \gamma(1+\alpha k)]) \left(\frac{x}{\sigma r^{\frac{1}{\alpha}}}\right)^{-1-\alpha k}.$$

Using the Recurrence Formula in a similar fashion yields

$$f_X(x) = \frac{1}{\alpha \pi \sigma r^{\frac{1}{\alpha}}} \sum_{k=0}^{\infty} \frac{\Gamma(1+\alpha k) \sin(\pi[\gamma + \gamma(1+\alpha k)])}{k!} (-1)^k \left(\frac{x}{\sigma r^{\frac{1}{\alpha}}}\right)^{-1-\alpha k}, \qquad \alpha < 1. \tag{3.20}$$

To see how this will give an asymptotic series in the $\alpha > 1$ case requires the use of a modification of the standard Jordan's Lemma (for proof see Appendix (C.3)).

Lemma 3.6.1 (Jordan's Lemma)**.** *Given the right-hand semicircle in Fig. 3.1 with the part-contour arc* $= \{c + Re^{-i\theta} \mid \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]\}$,

$$\left| \int_{arc} e^{-az} g(z)\, dz \right| \leq e^{-a} \frac{\pi}{a} M_R \quad \textit{where} \quad M_R := \max_{\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]} \left| g\left(c + Re^{i\theta}\right) \right|.$$

We have from (3.17) that $a = \ln(x/\sigma r^{\frac{1}{\alpha}})$ and since $g(\cdot)$ acheives it's max on $\theta = 0$ $M_R = \Gamma(c+R)\,\Gamma\left(\frac{1-c-R}{\alpha}\right) \sin(\pi[\gamma - \gamma(c+R)])$. So we have $\frac{e^{-a}\pi M_R}{a} \to 0$ as $x \to \infty$. Since the contribution from the arc vanishes for large x regardless of R, that leaves just the contribution from the diameter, and thus as $x \to \infty$,

$$f_X(x) \sim \frac{1}{\alpha \pi \sigma r^{\frac{1}{\alpha}}} \sum_{k=0}^{\infty} \frac{\Gamma(1+\alpha k) \sin(\pi[\gamma + \gamma(1+\alpha k)])}{k!} (-1)^k \left(\frac{x}{\sigma r^{\frac{1}{\alpha}}}\right)^{-1-\alpha k}, \ \alpha > 1. \tag{3.21}$$

The TCE Case Following the same steps as above, we can show for example that

$$\frac{\sigma r^{\frac{1}{\alpha}} H_{2,2}^{1,1}\left[\frac{s_q}{\sigma r^{\frac{1}{\alpha}}}\middle|\begin{matrix}(1-\frac{\alpha-1}{\alpha},\frac{1}{\alpha}) & (\gamma,\gamma)\\ (0,1) & (\gamma,\gamma)\end{matrix}\right]}{1-q}$$
$$= \frac{\sigma r^{\frac{1}{\alpha}}}{1-F_S(s)} \sum_{k=0}^{\infty} \frac{\Gamma\left(1-\frac{1}{\alpha}+\frac{k}{\alpha}\right)\sin(\pi[1-\gamma+\gamma k])}{k!}\left(-\frac{s_q}{\sigma r^{\frac{1}{\alpha}}}\right)^k.$$

An asymptotic series is again available in the $\alpha < 1$ case; however, Fig. 3.1 no longer applies. Instead, the Bromwich path in the $\alpha < 1$ case must separate poles at $s = -k$ and $\alpha - 1 - \alpha k$.

As a general comment, we would recommend the integral transform inversion approach whenever possible. The main issues are that the transition from the truncated power series to the asymptotic regime can create inaccuracies at points of interest (e.g. at high q values in the TCE case).

3.7 Conclusion

In this chapter we have presented a new way of computing allocations among business units with shared systemic shocks, via the GCLT and some useful properties of Stable distributions. Notably, we make use of the fact that as long as our business units share a tail index we can make a Stable approximation.

The methods presented here allow for a very general, prudent and simple way of handling the allocation problem. By assuming the worst-case tail index for each loss, the Stable approximation is viable. This in turn leads to very simple allocations as linear functions of total risk capital. Additionally, the total risk capital has been computed in the Stable approximation for the first time in the TCE case using the Fox H-functions.

The possibility of computing TCE allocation estimates should be valuable for insurance companies and banks trying to define their minimum capital requirements under the Basel

framework. This approach may also be generalized for other useful risk functionals.

Chapter 4

A Tale of Two Correlations: Comparing FIA Products at Scale

4.1 Introduction

In the United States, fixed annuities hold an important place in retirement portfolios. In addition to guaranteed fixed returns that are usually above and beyond other fixed-income instruments, they also come with a full array of useful benefits that would neither be easy nor practical to replicate otherwise, such as tax deferral, income, and death benefit provisions (see [NAFA, 2022]). In an environment of low interest rates (like the recent past) and prolonged market volatility, one particular subcategory of fixed annuities, the Fixed Index Annuity (FIA), provides two additional benefits, in equity market participation and principal protection respectively. The first, albeit sometimes very limited, is enabled by the insurance carriers through interest-crediting option strategies that are tied to a particular index, while the second is the guarantee offered by the insurance company via their investments of premiums in their general account. Selection and allocation to different interest-crediting options available within FIAs have become important questions to which many near-retirees need answers. The National Association of Insurance Commissioners (NAIC) in the United States prescribes (at least for FIAs) that illustrations and therefore

comparisons be performed via "back testing" [NAIC, 2022], that is using historical financial data to illustrate annuity performance. Looking into the most recent, worst and best index performance scenarios, one can definitely understand the "movements" inside the annuity. However, this kind of inspection can by no means determine which annuity construct (comprising crediting term + strategy + index) will provide the desired results. Back tests can be extended to cover a greater number of historical scenarios, but all of this comes down to what is usually referred to as "driving forward while looking in the rear-view mirror."

An alternative to back testing would be generating annuity return forecasts based on individual indices and then comparing their point estimates (means, medians, standard deviations etc.). One issue that we identify is that, while one can believe that the compared indices may have similar distributions, the annuity crediting strategies transform the returns so significantly that distributions are no longer comparable. This is easily overcome with side-by-side simulations of different indices and therefore annuity structures. In this way we can still compare point estimates but can also go further, drawing entire return distributions and determining which annuity did better and how often. At this point we arrive at what we believe are the main contributions of this chapter: doing this comparison, consistently, for many annuities struck on many different indices and with realistic results. We propose ways to model esoteric indices (in our case volatility controlled ones) more precisely, without re-engineering index mechanisms in our simulations but instead simply relying on known machine learning (ML) techniques to recreate them. To enable large-scale forecasting, we utilize long-term capital market assumptions that are established in the industry and provided by many institutions. We do this by fitting each of our index models with features from a common "core" economy consisting of major asset classes and their benchmark indices. We also show that in many cases this broad exposure works better when compared to the correct set of index features. Finally, we explore what models are best suited for the proper simulation of this common economy.

In order to make a selection or decide on the allocation within an FIA, it is important to understand both the interest-crediting structure of the annuity and the underlying index features. In order to properly compare two or more annuities whose returns are linked to

different underlying indices, the payout and index have to be modelled and forecasted in a consistent manner. Given that the interest-crediting formula is essentially a deterministic function of index returns, we look into ways to replicate index performance, pass obtained returns through annuities and ultimately allow for comparison and informed selection. We discovered two main challenges to this: i) cross-sectional index returns are difficult to model and ii) intertemporal return dependence in financial markets interacts with index calculations in novel ways. The former problem is a known issue in financial modelling generally. The latter issue, intertemporal dependence, is the motivation for stochastic volatility models. That said, these considerations have not been studied in this context before. As we will demonstrate, however, they are of critical importance.

It is also important to emphasize the aspects we are not going to address, and where and how we chose to make certain assumptions. We will neither explore nor elaborate on the merits and appropriateness of annuity-crediting strategies. Despite ongoing criticism in the light of the famous example of [Babbel et al., 2010], we will not model annuity parameters (caps, participation rates, spreads etc.) alongside index simulations, instead keeping them constant for the duration of calculations.

This chapter is organized in two parts, as follows. In Part 4.2 we explore the cross-sectional correlation issue. Section 4.2.1 examines some toy models of annuities, illustrating the importance of cross-sectional return correlations among indices. In Section 4.2.2, we discuss the kinds of indices linked to FIAs. While there are many kinds of indices involving various calculations, we discuss their relative challenges and choose to focus on an extremely popular and easily understood category within risk-controlled (RC) or volatility-managed indices. Section 4.2.3 discusses our common-economy method of sidestepping the issue of direct cross-correlations by using machine learning on a smaller subset of benchmark indices. Finally, in Section 4.2.4 we simulate single and multiple annuity rates using the aforementioned methods.

In Part 4.3, section 4.3.1 draws on our analysis in the previous section to understand the ways in which our methods may fall short. As we are replicating the index methodology instead of modelling the index levels directly, we ask what features are important to account for in the underlying index components. In Section 4.3.2 we discuss the nature of

index return statistics generally and explore the intertemporal dependence of returns. We then explore stochastic volatility models as a means of accounting for these intertemporal features, along with the relative advantages and disadvantages of well-known models. In Section 4.3.3 we explore the specific features of risk control models and hypothesize on the behaviour we observed in Part 4.2. Finally, we bolster this hypothesis with results in Section 4.3.3. We find that more sophisticated fractional or "rough" stochastic volatility models are needed to properly account for the effects of the risk control mechanism. We conclude by discussing the implications of the analysis covered in this chapter, as well as possible future avenues of research.

At this point, it is also important to bring to attention the fact that this chapter is a product of joint work with another PhD student and (then) VP of research at Cannex Financial Exchanges Ltd, Branislav Nikolić. I was a MITACS intern at CANNEX on two occasions and Branislav has a decade-long career interest in annuity products. To contextualize our respective contributions, much of the insight for this work was framed within an expansive MITACS[1] project hosted by CANNEX and York University where the collaboration that led to this project was established.

4.2 Cross-Sectional Correlation: Annuity Forecasting at Scale

4.2.1 Motivation and Examples

We will discuss FIAs in more detail in Section 4.2.2, but to begin with we present the main FIA building blocks in our modelling. A growing number of FIAs are linked to volatility- or risk-controlled (RC) indices, denoted by I_t in our particular example. Basic RC indices shift their exposure between an equity index, say the S&P 500 (SPX), and a less volatile fixed-income index or even cash. We will refer to these "primary" indices as

[1] https://www.mitacs.ca/

P_t and B_t, respectively. The return on the RC index is then passed through an annuity-crediting formula; this return, denoted here X_i, is what the investor receives in their FIA accumulation account.

Consider for example a 10-year FIA using a 55% participation strategy and linked to the S&P 500 Average Daily Risk Control 10% USD Price Return Index (SPXAV10P). In this case, P_t is the SPX and B_t would be a cash account. The index provider, in this case Standard & Poor's, tracks index levels of P_t and B_t to construct index levels for I_t. Finally, the annuity carrier applies 55% of the annual returns of I_t for the duration of the FIA.

To summarize,

$$\begin{aligned} I_t &= f(\{P_s\}_{s\leq t}, \{B_s\}_{s\leq t}) \\ X_{t_i} &= \max\left(0.55\frac{I_{t_i} - I_{t_i-\delta t}}{I_{t_i-\delta t}}, 0\right) \\ i &= 1, ..., 10 \\ \delta t &= 1 \text{ year}. \end{aligned} \tag{4.1}$$

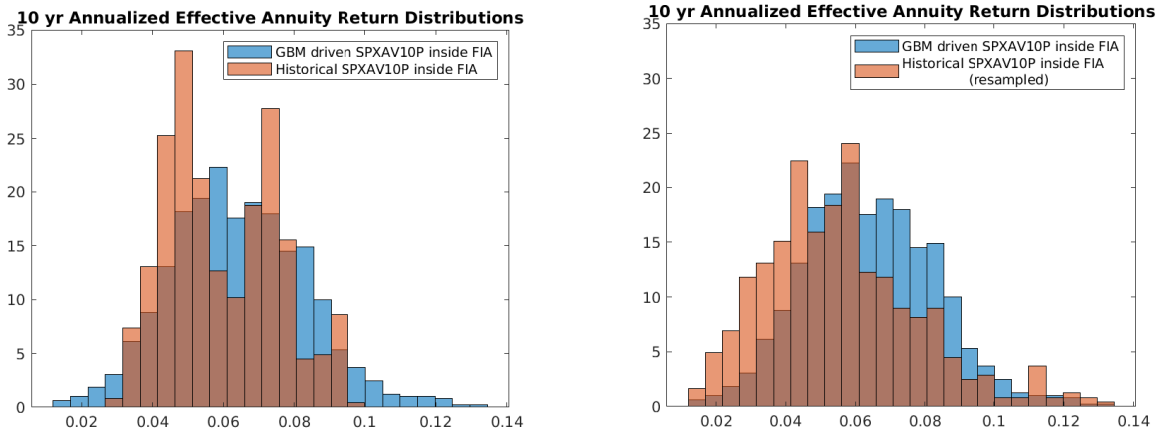
The exact nature of the mapping from the basic indices to the RC index $f(\cdot, \cdot)$ will be discussed in Section 4.2.2. For now, suffice to say that this constant reallocation from a risky asset class to a less risky one is designed to keep the volatility below a prescribed level.

The current state of the art[2] in forecasting FIA returns (i.e. the X_i's) is to model the underlying index returns (the I_t's) by a stand-alone Geometric Brownian Motion (GBM) and then pass them through the annuity crediting formula. We now model the mean effective 10-year return of the annuity in the previous example and consider some problems arising from the results. This is achieved by computing terminal accumulation value at the 10-year mark and calculating the annual rate of return that would yield it, namely

$$Y = \left(\prod_{i=1}^{10}(1 + X_i)\right)^{1/10} - 1.$$

[2]Evidenced by Branislav Nikolić's experience while working at CANNEX. For more details see white papers produced there during his tenure.

We present two sets of results (Figure 4.1a), one in which index I_t is modelled by a GBM and another for which we rely on the historical data for the index; note however that both represent returns after these are passed through the annuity crediting formula (i.e. X_{t_i} in Eq. 4.1). More precisely, for the historical data we took 10-year windows of monthly index returns, starting in 1986 (when the SPXAV10P data set begins) and moving forward in monthly increments to create our data set. The GBM was calibrated to the historical data set of monthly price levels for SPXAV10P obtained from CANNEX.



(a) (b)

Figure 4.1: Histograms of 10-year annualized annuity returns without and with resampling of returns (panels (a) and (b) resp.)

	GBM Index	Real Index Returns (Actual)	Real Index Returns (Resampled)
Mean	0.064	0.060	0.056
SD	0.019	0.016	0.023
Skewness	0.286	0.322	0.758

Table 4.1: Annuity return comparison of a direct GBM model for the index with the actual and resampled returns

On the other hand, compare our previous results to the case where we have taken the sample of historical index returns and resampled them with replacement before passing

them through the annuity formula (Figure 4.1b). That is, we have removed any dependence structure that exists through time on the index returns while still using their historical distribution. Keep in mind the caveat that this historical distribution may not be representative of a stationary distribution, if it exists. Even so, we can at least *visually* observe that the resampled results align much closer to the GBM results.

We may be tempted to think we can salvage the GBM model by simply altering the GBM parameters to give results more consistent with the real distribution. However, this ignores two issues:

Issue 1: Intertemporal Dependence of Returns While shifting the parameters of the GBM may yield some reasonable values, it ignores any underlying mechanisms accounting for why such a shift would be needed. While the underlying index or P_t (SPX in this case) itself appears to behave *roughly* like a GBM at certain time scales [Cont, 2001], it is not immediately clear whether the behaviour of the underlying index or the risk control mechanism is the source of any intertemporal dependence and deviations from the GBM model. Without understanding these issues, it would not be prudent to alter the GBM parameters manually to fit historical results, as there may be hidden risks in doing so. We will study this in much greater detail in Part 4.3 of the chapter.

Issue 2: Comparing at Scale Point estimates are often insufficient to determine how often one specific annuity product will outperform another. Assume for now there are no issues with the GBM methodology and consider a new experiment involving annuities being struck on two benchmark indices (there is no RC mechanism under consideration in this case). Here we have simulated two indices, the S&P 500 (annual average return (μ) of 0.08 and volatility (σ) of 0.16) and the Russell 2000 ($\mu = 0.11$ and $\sigma = 0.19$) and have considered a few annuity strategies in various commonly occurring comparisons. Clearly, the higher mean return of the Russell index translates into greater annuity returns most of the time. How often this is true is reliant on the dependence structure of the index returns (a historical estimate of the correlation between returns gives $\rho = 0.63$) and thus will not

be captured by point estimates alone. This is true in general: any consistent comparisons will rely heavily on the dependence structure between index returns (see Table 4.2).

	SPX with 7.75% cap		Russell with 25% par		
	Mean	SD	Mean	SD	$P(\text{SPX} > \text{Russ})$
$\rho = -1$	4.6%	1.1%	3.5%	1.1%	69.4%
$\rho = 0$					74.9%
$\rho = 0.63$					82.6%
$\rho = 1$					92.5%

	SPX with 7.75% cap		Russell with 7.25% cap		
	Mean	SD	Mean	SD	$P(\text{SPX} > \text{Russ})$
$\rho = -1$	4.6%	1.1%	4.7%	1.0%	48.3%
$\rho = 0$					48.2%
$\rho = 0.63$					47.9%
$\rho = 1$					44.1%

	SPX with 30% par		Russell with 25% par		
	Mean	SD	Mean	SD	$P(\text{SPX} > \text{Russ})$
$\rho = -1$	3.3%	1.1%	3.5%	1.1%	45.3%
$\rho = 0$					44.6%
$\rho = 0.63$					41.7%
$\rho = 1$					0.0%

Table 4.2: Relative performance of 10-year annuities written on benchmark indices for different strategies and dependence structures

Today there are between 100 and 150 indices linked to FIA products of relevance, many of which are only a few years old. What this means in practice is that many of these indices will have only a few years of "live" data compared to back-tested performance. It is also well known that estimating correlations among stock returns is problematic, especially in the case of multiple assets (see e.g. [Bun et al., 2017]); we believe this also applies to equity indices. Indeed, this may help explain the many theoretical irregularities in real portfolios ([Liu and Zeng, 2017] or [Huang et al., 2017]).

Compare the following two scenarios: (i) A sample of 100 indices with five years of daily returns and (ii) 15 indices with a 10-year daily return universe. Scenario (i) captures the

data we need to work with when studying FIA returns. Scenario (ii) motivates our choice of a common economy in Section 4.2.3. In both cases we model the indices with jointly correlated GBMs. In Figures 4.2 and 4.3 we look at the statistics of *estimated* correlation matrices of returns in the two scenarios[3]. Specifically we focus on the distribution of the highest eigenvalue (λ_{max}) in Figure 4.2 and the mean absolute error of the entries in Figure 4.3. In scenario (i), in blue, higher errors within the correlation matrix itself can clearly be seen and indeed the sampling distribution for the largest eigenvalue is not representative of the actual value. In scenario (ii), in red, we see an improvement on both counts. Indeed it is not obvious that we can reliably estimate the dependence of the 100 indices to such a precision that a joint GBM model will meaningfully deliver index returns and therefore annuity comparisons. Our main contribution here is the claim that we can sidestep the issue of fitting 100 indices by moving to scenario (ii). Many FIA indices are constructed from the same basic building blocks. For example the whole class of SPX RC indices is constructed from the SPX and fixed-income indices alone (see Figure 4.5).

There remains the issue of accurately replicating the (proprietary) risk control calculations. Our first attempts at doing so using the calculations provided by some carriers were comparatively unsuccessful. Without access to the exact computations we found we could not perfectly replicate the index even in sample. That is, given the correct components and the formulas provided by the index provider, our reconstructed index deviated from the true value. Starting in Section 4.2.3 we instead use statistical learning (a.k.a. machine learning) techniques, with much greater success.

4.2.2 Fixed Index Annuities

When introduced in the United States in the mid-90s, FIAs (then called equity-indexed annuities) were intended to provide owners with interest credit that would outpace banks' certificates of deposit (CDs). Fast forward a couple of decades and we have an array of index-linked products that still provide equity index-based interest credits but also offer

[3]We chose a random positive definite matrix for the "true" correlation matrix in each scenario, but the results are universal. See e.g. [Bun et al., 2017].

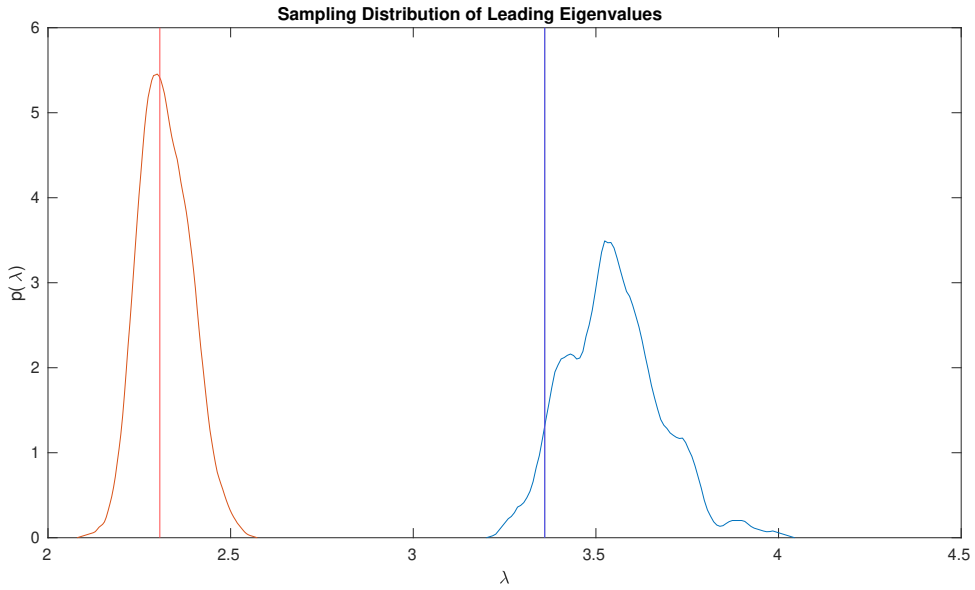


Figure 4.2: Sampling distribution of λ_{max}

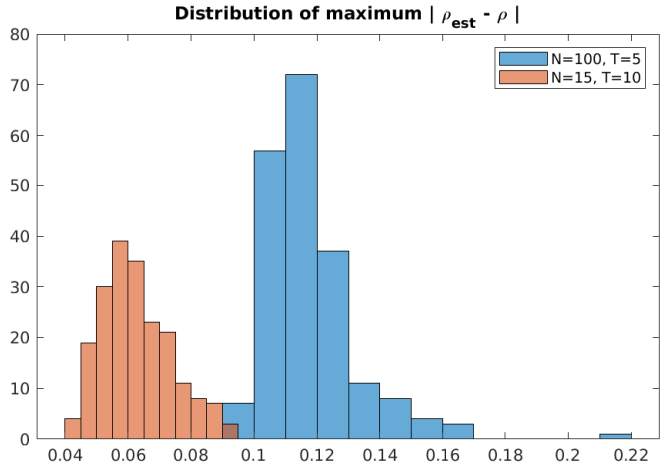


Figure 4.3: Sampling distribution of the MAE for $\rho_{i,j}$

principal protection for near- or new retirees. Through over a dozen distinct crediting methods in conjunction with the 100+ indices available, these products offer a unique value proposition that demands insights from several areas for the consumer to make an informed choice. Selecting one that is just right for a particular individual is an almost impossible task and there is now a sizable literature exploring product types, crediting strategies, index characteristics and so on (see [Alexandrova et al., 2017] and references therein). If annuities are linked to the same index, the choice of crediting strategy boils down to the highest available rate and which crediting method is likely to produce favourable returns given belief about how well (or not) the index will perform in the future. However, we have a number of options when it comes to understanding how different annuities struck on different indices compare.

In terms of index crediting strategies, a popular choice is point-to-point participation with and without a cap. The participation strategy without the cap is the best illustration of why the indices have to be modelled consistently, as it provides minimal return transformation, while the capped version is the most popular from the sales perspective. We focus on the former strategy in this chapter.

The point-to-point participation rate strategy measures the performance of the index from one point in time to another, usually running from the day the contract is purchased until a year later and then repeating the measurement every year. The strategy then uses a "participation rate," a percentage of index gains that determines the amount of credited interest so that it increases in proportion to the index returns. The participation rate may be above or below 100%, depending on the nature of the index itself. On the other hand, the point-to-point cap strategy similarly measures the performance of the index from one point in time to another; however the strategy then uses a cap that limits the amount of interest, no matter how well the index performs. Different levels of participation are technically available when the cap is used, but the level is usually set at 100%. As with all FIAs, the interest credited in any year cannot go below 0% no matter how much the index loses. Participation strategy returns mimic the underlying returns effectively but are usually dampened or enhanced, while the cap provides full exposure within the limit (see [Nikolić et al., 2018]).

While the ideal scenario would be to address a broad universe of indices available in the market, after initial experimentation we decided to focus on risk-controlled indices, as the machinery we built would need different tweaks depending on the index category. Another important reason for this particular choice was the availability of "live" index history. Once an index goes live, its returns are a true representation of how it works compared to what was intended by the index manufacturer. Absent that, what is available is the "back-tested" or "synthetic" history where the index provider illustrates the behaviour that would have been observed in the past based on the underlying index mechanics and movement of the constituents. Here we draw attention to a potential pitfall. The index behaviour observed in the back-test may be substantially different to the one after the index goes live, in which case our model would learn from a not-very-representative sample. In other words, if the index methodology was overfitted to past market conditions and if history does not repeat, the index will diverge from the expected behaviour and therefore the model's forecast may be misleading.

One example that illustrates some of the idiosyncrasies of more complex index structures is the Barclays Trailblazer Sectors 5 Index (BXIITBZ5). As a "smart beta" index its behaviour is motivated by Modern Portfolio Theory (MPT) and Arbitrage Pricing Theory (APT), and it aims to use momentum in the market to inform index constituent selection and to rebalance among selected constituents (see [BXI, 2022] and [Kim and Francis, 2013]). Its ability to recognize and act differently in a highly stressed period (e.g. during the opening months of COVID) as opposed to more typical market cycles causes problems for our machine learning approach. Due to the complexity of the algorithm, each instance of return computation is sufficiently novel that an effective training set would need to be prohibitively large, while making certain out-of-sample events not generalizable. Conversely, risk control calculations are consistent and simple enough that even a relatively small training set can be effective, even when we have events out of sample that are in hindsight seen as significant or even seismic.

A detailed description of machine learning techniques used is outlined in Section 4.2.3. At this point, for illustrative purposes only we provide the summary of this exercise in Table 4.3. We see that low R^2 values are mainly unaffected by different training/test set

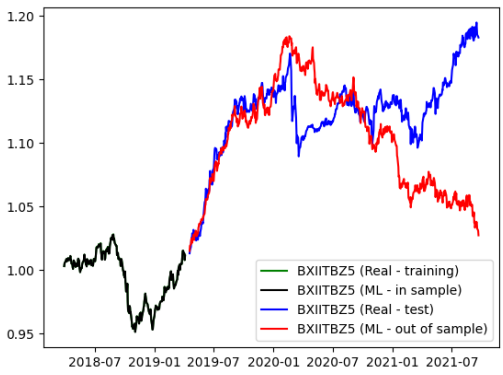
ratios in the neural network cases and that our estimates are not up to par. This conclusion is even possible by visual inspection of Figure 4.4.

		Train/Test Split		
		30% / 70%	50% / 50%	70% / 30%
NN Model R^2 (Out of Sample)	SPX10AV10P	0.455	0.664	0.730
	BXIITBZ5	−0.130	−0.109	−0.012

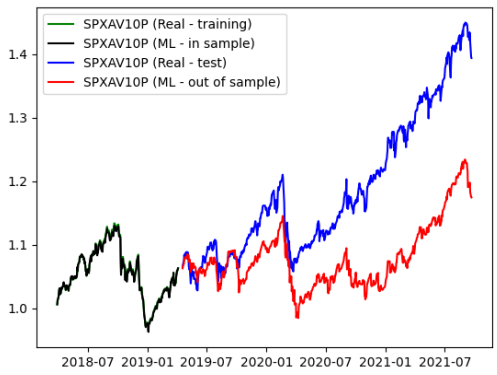
Table 4.3: Comparison of R^2 for out-of-sample predictors for different training/test data splits

Volatility-Controlled Indices

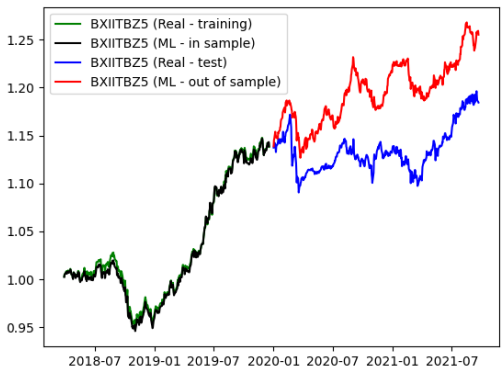
Managing volatility emerged as a noteworthy exercise and rose in prominence particularly within the insurance industry during and immediately after the 2008 financial crisis. Its intent was to reduce risk and severity of losses for securities investors, but it enabled variable annuity issuers to offer secondary guarantees on their annuities because of lowered hedging costs, even in the face of the volatile period of low interest rates and bond yields that followed. Managed volatility expanded far beyond this use with its application to index-linked products. Prior to volatility management, there was little or no need for such management for index-linked structured products such as structured notes and FIAs. Structured product writers, namely banks and insurance companies, were able to use benchmark indices in their products and still offer competitive crediting rates compared to other fixed-income vehicles. As most of these products use options struck on the underlying index to fashion the payouts, both the prevailing interest rate and the underlying volatility play significant roles not only in the option pricing but also in the variability of the price at any given time. As low interest rates persisted, the option budgets required to provide credits in the FIAs, being tied closely to bond yields, were simply not adequate to offer attractive payouts using relatively volatile broad market indices. Issuers were able to address this by selecting or creating a less volatile underlying index that would result in more predictable and attractive rates. Lowering and stabilizing the volatility of the underlying index would ensure lower and more stable option prices for annuity providers and



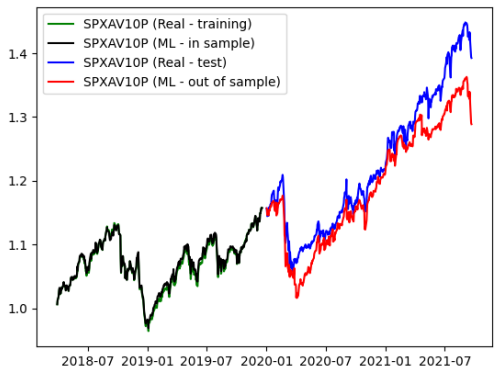
BXIITBZ5 30%/70%



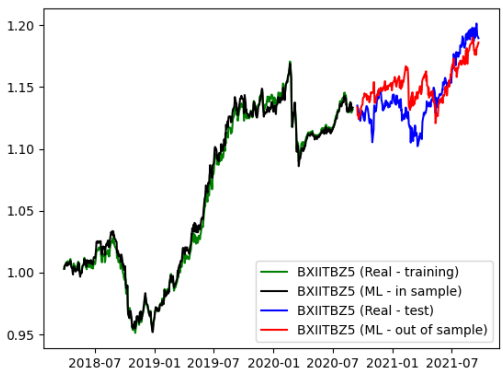
SPXAV10P 30%/70%



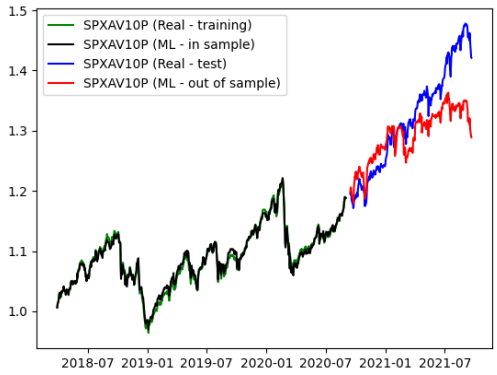
BXIITBZ5 50%/50%



SPXAV10P 50%/50%



BXIITBZ5 70%/30%



SPXAV10P 70%/30%

Figure 4.4: Comparison of neural network models in and out of sample for various train/test splits. SPXAV10P is more easily captured than BXIITBZ5 when there is less data to train on.

therefore better offerings for annuity purchasers. In general, volatility-controlled indices operate on the rudimentary principle of moving exposure between high- and low-volatility assets. In its simplest form, this transition takes place between equities and bonds (see Figure 4.5). Pursuing an explanation on a slightly deeper level immediately becomes more complicated, as different index providers use different methods to measure and achieve efficient and worthwhile "shifting" while preserving enough of the upside of credited interest as well. These methods include volatility targets or levels, predetermined rule sets and leverage conditions, or a combination thereof to provide an index that allows for attractive option pricing and therefore favourable annuity rates (higher participation, caps and lower spreads). Index providers tackled this problem in a number of ways over the years, with methodologies evolving and often being kept secret. Without extensive work put into reverse engineering it has been very difficult to forecast the performance of these indices, as well as any structured products that use them as the underlying index.

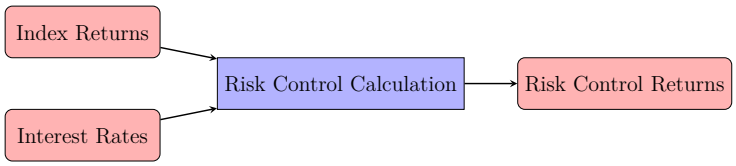


Figure 4.5: Schematic of a risk-controlled index

Consider one generic implementation of a RC index in the following[4]: we model the price level of some equity or commodity index at time t by a stochastic process P_t. We model the dynamics of P_t with the following SDE:

$$\frac{dP_t}{P_t} = \mu dt + \sigma_t dW_t$$

where W_t is a standard Wiener process. Note the time dependence imposed on volatility, σ_t. We can model this using a local or stochastic volatility model. RC or volatility-controlled indices take the real-world stochastic nature of volatility into account and seek to provide the same equity premium exposure of another index while maintaining a volatility

[4]Inspired by the discussion in [spg, 2022], reformulated here using stochastic calculus.

around a target value. They do this by shifting exposure between fixed-income and equity components. Consider a generic bond (or interest rate index) modelled by another process B_t. For now we assume $dB_t = rB_t dt$. We can construct an RC index I_t in the following generic way. Using time-dependent weights $w_t^{(i)}$, the RC index as sum of positions in equity and bonds can be modelled by the following SDE:

$$dI_t = w_t^{(1)} r B_t dt + w_t^{(2)} P_t(\mu dt + \sigma_t dW_t) \tag{4.2}$$

$$\frac{dI_t}{I_t} = \underbrace{\left(\frac{w_t^{(1)} B_t}{w_t^{(1)} B_t + w_t^{(2)} P_t}\right)}_{1-\pi_t} r dt + \underbrace{\left(\frac{w_t^{(2)} P_t}{w_t^{(1)} B_t + w_t^{(2)} P_t}\right)}_{\pi_t} (\mu dt + \sigma_t dW_t) \tag{4.3}$$

$$= (r + \pi_t(\mu - r))dt + (\pi_t \sigma_t) dW_t. \tag{4.4}$$

We can write this as $dI_t = \mu_t^* I_t dt + \sigma_t^* I_t dW$ where $\mu_t^* = r + \pi_t(\mu - r)$ and $\sigma_t^* = \pi_t \sigma_t$. Assuming the fixed-income exposure is fairly close to the theoretical risk-free rate, the Sharpe ratio of the new index I_t is the same as the original[5].

If we set $\overline{\sigma}$ as the "target" volatility, then the natural choice is to set $\pi_t = \frac{\overline{\sigma}}{\hat{\sigma}_t}$ where $\hat{\sigma}_t$ is an estimate of the current volatility. This ensures that the RC index (at least in theory) maintains a constant target volatility of $\sigma_t^* = \frac{\overline{\sigma}}{\hat{\sigma}_t}\sigma_t \approx \overline{\sigma}$. There are of course caveats to this. There are different methodologies in use for calculating $\hat{\sigma}_t$ that will not perfectly capture σ_t. Standard choices include a rolling window of the standard deviation of index returns or an Exponentially Weighted Moving Average (EWMA) model. Many RC indices also take a maximum or average of different estimates to smooth the estimates out (again, see [spg, 2022] for details). Most important for our purposes however is the reality that in order for carriers to provide FIA exposure to an RC index, they have to purchase a custom derivative over the counter. The writer of this derivative contract in turn needs to hedge their exposure with options on the original equity index. As a result, the RC index cannot be too dissimilar from the original, so that constructing the RC index entails

[5] Considering that $\frac{dI_t}{I_t}$ are the returns (see Part 4.3), $\frac{\mu_t^* - r}{\sigma^*} = \frac{\mu - r}{\sigma_t}$.

leverage restrictions (denoted K_u) and maximum bond portions (denoted K_d) such that

$$\pi_t = \max(\min(\hat{\sigma}_t, K_u), K_d). \tag{4.5}$$

Equation 4.5 will be very relevant in our discussions in Part 4.3. As we shall see, the RC index inherits enough essential features of the original index that any modelling at scale will require some sophisticated stochastic volatility modelling.

Before moving on to the next section, we would like to discuss some experiments and results that for brevity's sake we do not include here. For the rest of the chapter we detail how we used ML to reconstruct RC indices from other components in the market. Given the simplicity of (4.4) and (4.5) it would be natural to ask why we could not directly reconstruct the index, no ML required. When we attempted to model RC indices directly using (4.4) and (4.5) and the methods provided to the public in the index fact sheets, we found the results to be poor in comparison. In the ML case, as we shall see, we provide a set of trailing returns to allow the machine to make the necessary volatility calculations. When we provided $\hat{\sigma}_t$ as an input in lieu of trailing days of returns, the ML results underperformed in comparison. This leads us to believe there are idiosyncrasies in how $\hat{\sigma}_t$ is estimated (data cleaning, weekend/seasonal effects etc.) to which we do not have access, and for this reason ML models are more favourable. Furthermore, their generic nature simply makes for a more efficient methodology. Our technique does not require novel reproduction of the index methodology associated with each specific RC index.

4.2.3 Index Methodology

The CORE Hypothesis

A key motivating question for our work was "What percentage of time does one annuity struck on a particular index outperform another?" To answer this we need to be able to produce the relevant index return distributions (back-tested or forecasted) for different annuities and determine the overlap, the proportion of times one annuity return was higher

than the other. In any event we need two pieces of machinery: an index simulator and an annuity return calculator. Given the straightforward and deterministic nature of the latter we can forgo a detailed explanation. In this section we focus on index simulations.

In Section 4.2.1 we saw how we could perform direct index modelling via stochastic processes with limited success. We also encounter scalability problems when we try to expand this approach to capture a larger set of annuities. As said in Section 4.2.1, the difficulty comes down to the impossibility of producing a good correlation matrix for the large number of underlying indices. In this section we detail our solution via ML techniques.

A first approach to consider is to "back out" the index methodology function using known index components. This quickly presents a problem, however: even if the set of index components may be smaller than the set of FIA indices we are interested in (e.g. the SPX is a known component of all SPX RC indices), we are still relying on the ability to simulate *all* known annuity index components with high-dimensional correlation matrices. We use this first approach as a benchmark for our "core economy" hypothesis. We spent some time looking into the overarching set of indices that would encompass as much of the economy as possible (from now on referred to simply as CORE) while staying within a reasonable number of indices due to the correlation issues identified in Section 4.2.1. We hypothesize that our machine learning algorithm should be able to pick up the index function equally well from the "core economy" set of indices as compared to calibrating it on the set of known index components. With this separation, we open the door for future research on how to model and simulate the economy optimally while using ML-derived functions for indices.

One question that remains is the proper choice of the CORE economy. We define the CORE as our set of indices representing asset classes that capture the broader market well (See Table 4.4 for individual indices). From now on, for simplicity we will refer to the set of equity indices listed in Table 4.4 as CORE equities and the set of fixed-income indices as CORE bonds. The CORE data was provided by CANNEX and consists of daily price-level index time series from 2010/01/29 to 2021/09/21.

Traditionally, long-term capital market assumptions available from investment banks,

Asset Class	Description	Name	Ticker
Equities	US Small Cap Equities	Russell 2000	RTY
	US Small–Mid Cap Equities	Russell 2500 Index	R2500
	US Mid Cap Balance Equities	MSCI US Mid Cap 450 Index	MZUSM
	US Large Cap Equities	Russell 1000	RIY
	US All Cap Equities	Russell 3000	RAY
	World Equities	MSCI World Index	MXWO
	International Equities	MSCI ACWI ex US Index	MXWDU
	Emerging Markets Equities	MSCI Emerging Markets IMI	MXEFIM
Fixed Income	High-Quality US Fixed Income	Bloomberg Barclays US Aggregate Bond Index	LBUSTRUU
	Low-Quality US Fixed Income	Bloomberg Barclays High-Yield Very Liquid Bond Index	I33743US
	World Fixed Income	Bloomberg Barclays Global Aggregate Bond Index	LEGATRUU
	International Fixed Income	Morningstar Global ex-US Core Bond	MGXUSN
	Money Market	Bloomberg Barclays 1–3 Month US Treasury Bill Index	I00078US

Table 4.4: The CORE set of indices

asset managers and insurance companies (JPM, PGIM, Callan etc.), are provided on an asset-class basis in terms of their returns, volatility and correlations. These are forward-looking assumptions and can be used for forecasting in the models. Ideally, these would coincide with the asset classes available in the training data set. For that reason, we chose to represent each asset class with the index benchmark. The chosen index benchmark is preferably a good representation of the asset class but not necessarily an explicit component of the risk control index that we are trying to train on, to avoid overfitting in the model. This would allow for calibration to the history via the index benchmark and the use of long-term capital market assumptions for more consistent forecasting. Ultimately, this is just one possibility and the technology we developed for index calculations should in principle be able to use any other variations of the CORE economy representation (i.e. alternative sets of indices), subject to size constraints that would make forecasting and projections feasible in regards to the correlation issues that we outlined earlier.

Machine Learning Techniques

In examples presented in this section we use the CORE dataset from Section 4.2.3 in addition to the S&P itself and the yield on U.S. Treasury Securities at 10-Year Constant Maturity (DSG10). These later additions will serve as the true ingredients for some RC

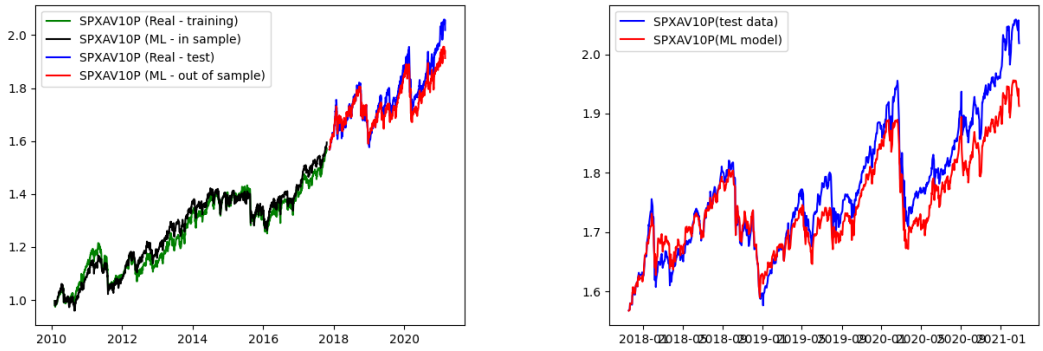
Equity Features	Fixed income Features	Training Set			GPR R^2	NN R^2
		# of Daily Returns	# of Indices	Total # of Returns		
S&P	DGS10	8875	2	17750	0.884	0.878
	I00078US	2807	2	5614	0.807	0.82
	LBUSTRUU	2807	2	5614	0.804	0.816
	Core Bonds	744	6	4464	0.756	0.617
Core Equity	DGS10	3055	9	27495	0.92	0.902
	I00078US	3055	9	27495	0.919	0.898
	LBUSTRUU	3055	9	27495	0.92	0.896
	Core Bonds	918	13	11934	0.907	0.809

Table 4.5: Summary of results; neural network with 10 layers, with 10 trailing days as inputs for SPXAV10P. R^2 is calculated *out of sample* (30% of the total data set allocated for training)

indices as a point of comparison. For our features or predictors, we exclusively used daily returns (as this is the true input during index construction) and included 10 trailing days of index returns to allow for any algorithm to compute current volatility levels. Our response variables were the returns for the index of interest.

We started our experimentation with a simple ordinary regression model over index returns and as anticipated found this not to be a worthwhile avenue. Looking at Eq. (4.5) in Section 4.2.2 we can see that index exposure will be dynamic and nonlinear. We did however discover a feature that we carried over into more complex ML techniques. Including a set of lagged or trailing returns for the underlying indices produced better results than incorporating traditional volatility calculations in the case of RC indices. Clearly there were idiosyncrasies in the real volatility calculation that were hard to capture with the products' stated volatility methodology (e.g. EWMA or rolling window). It is unlikely that the index documentation provided (albeit possibly vague) is incorrect; rather, we suspect our inability to reproduce the index precisely is based on our interpretation of the rules and our dataset.

The natural starting point beyond a regression model when dealing with a "black box" function like an index calculation is a neural network (NN) of the kind repopularized by [Hinton, 1990]. These have become increasingly popular in recent years. For a practical or introductory reference on these and other statistical learning topics, see [Wüthrich and Merz, 2022] and references therein. We used standard stochastic gradient descent (SGD)



(a) (b)

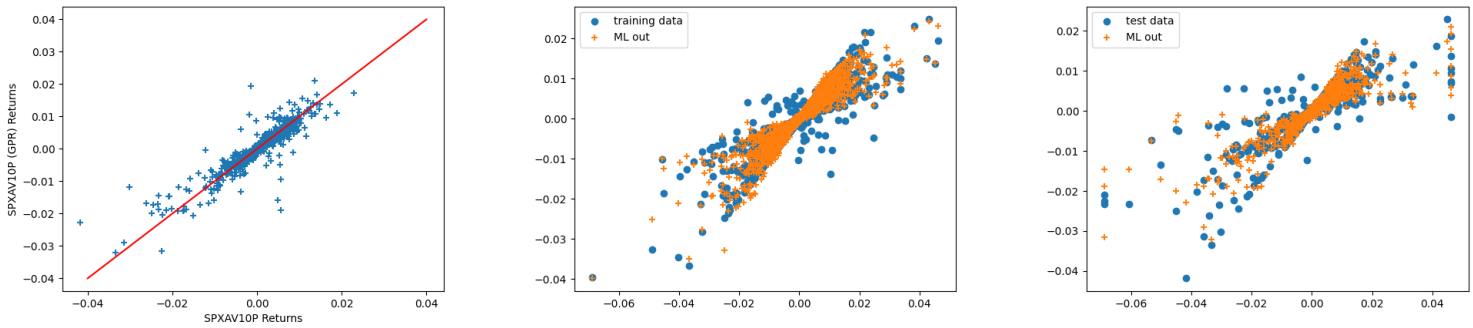
Figure 4.6: SPXAV10P results with 20 trailing days: (a) all data, (b) zooming in on out-of-sample performance

training with a tanh activation function from [scikit-learn Developers, b]. Knowing exact index components as features in an NN gave promising results (see the first row of Table 4.5). As discussed in the previous section, we were ultimately interested in the performance of the generic CORE economy which delivered comparable if not better performance out of sample.

As we set up our NN with 10 layers, one issue we quickly discovered was that including trailing days of returns meant very wide NNs. For instance, when trying to model the SPXAV10P, from Table 4.5 we see that the number of features is 10 trailing days times the number of indices, for example 130 in the CORE bonds and equity row. Relative to the 10 hidden layer results presented in Table 4.5, our NNs are very shallow. In order to remain confident that we were not simply "fitting the data" but rather accessing the full potential of deep learning[6] to correctly handle novel inputs, we needed to explore larger, deeper networks.

Our first attempt to improve the fit by increasing the depth of the NN from 10 hidden layers to 20 is shown in Figure 4.6. We observe that we can get reasonably good fit in

[6]see [Nielsen, 2015] for information about deep learning and [Safran and Shamir, 2017] for more on the depth vs. width trade-off.



(a) (b) (c)

Figure 4.7: Same NN model as in Figure 4.6 but now comparing returns of (a) the model and real data (b) the in-sample index vs SPX returns and (c) the out-of-sample index vs SPX returns

and out of the sample, but cumulative returns viewed through index price levels start to diverge out of sample. The divergence is mostly caused by a few outliers whose magnitude is not properly captured, even in sample. This nuance is best observed in Figure 4.7b, which depicts relatively good agreement of actual returns with the ones produced via ML (visually, most blue dots are covered with orange ones) in sample, while still having few observations that disagree.

Unfortunately, we observed that the NNs in this case started to suffer from vanishing gradients and a low convergence rate (see [Hochreiter, 1998] and references therein). We claim that a more appropriate technique is Gaussian process regression (GPR).

We will not reiterate the theory of GPR here, nor explain the implementation we used from [scikit-learn Developers, a] other than to mention that we made use of a standard RBF kernel. Briefly, GPR is a technique that can be thought of as a Bayesian stochastic kernel machine or a generalization to Bayesian linear regression accounting for more complex, nonlinear functional forms. Our motivation for using it here was twofold: computational cost and certain theoretical qualities. Heuristically, the time complexity of SGD is about $\mathcal{O}(dn^2)$[7] where d is the number of parameters and n is the training sample size. By

[7] To see this, $\mathcal{O}(d) \times \mathcal{O}(n) \times \mathcal{O}(1/\eta)$, the cost of gradient computation per epoch multiplied by $\mathcal{O}(1/\eta) \approx$

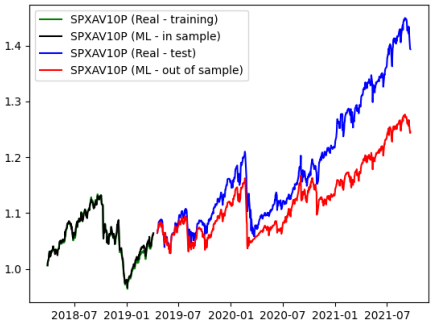
comparison, the time complexity of fitting a Gaussian process regressor is $\mathcal{O}(n^3)$. The large NNs in this case make GPR comparable or more practical. Theoretically, a Gaussian process regressor with an RBF kernel can be thought of as the limit of an infinitely deep NN (albeit an RBF network, not an MLP; see [Domingos, 2020]). So we can sidestep the vanishing gradient issue, while hopefully achieving better out-of-sample performance at the same cost. Table 4.5 seems to bear this intuition out. Additionally, many common financial models are themselves Gaussian processes with a particular kernel. So we can model the CORE indices with some stochastic process and GPR will produce the resulting index models as a jointly distributed stochastic process. For the mathematically inclined, GPR provides a common language between potential CORE models and our machine learning models: stochastic processes. We refer the interested reader to [Williams and Rasmussen, 2006], [Lilley and Frean, 2005], [Domingos, 2020] and references therein.

Discussion of Results

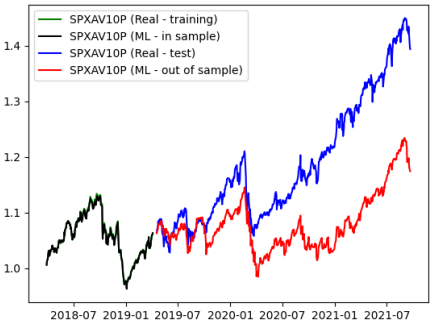
From Table 4.5 we see that the CORE data set performed fairly well as predictors. Even though more cross-validation is required, we see an improvement in out-of-sample performance when compared to the case of known index elements. Additionally, we achieved these results with fewer trailing days than the actual volatility estimation techniques for these indices.

It is important to note a difference in the results summarized in Table 4.5. As a starting point, we took S&P 500 returns and "added" interest rate inputs. In addition, we tested adding our entire set of CORE bonds and observed no material difference in the quality of the estimate, as measured through R^2 out of sample. This was true regardless of the proxy for interest rates used. Under this data setup, GPR performs much better than NN with 10 layers, and moreover the NN started to suffer from vanishing gradients when we tried to increase the network depth. Alternatively, if we substitute S&P 500 as a feature with the broader set of CORE equity indices, we see notable improvements across the board. Even though CORE bond selection yields worse results compared to more precise interest rate

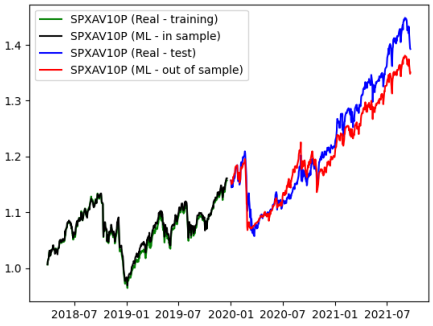
$\mathcal{O}(n)$ where η is the learning rate.



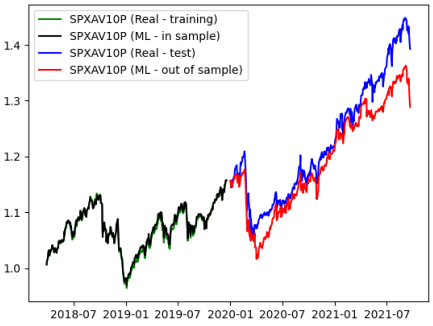
SPXAV10P (GPR) 30%/70%



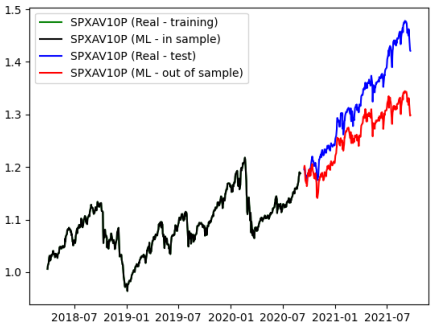
SPXAV10P (NN) 30%/70%



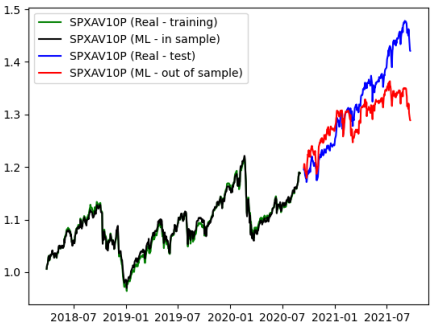
SPXAV10P (GPR) 50%/50%



SPXAV10P (NN) 50%/50%



SPXAV10P (GPR) 70%/30%

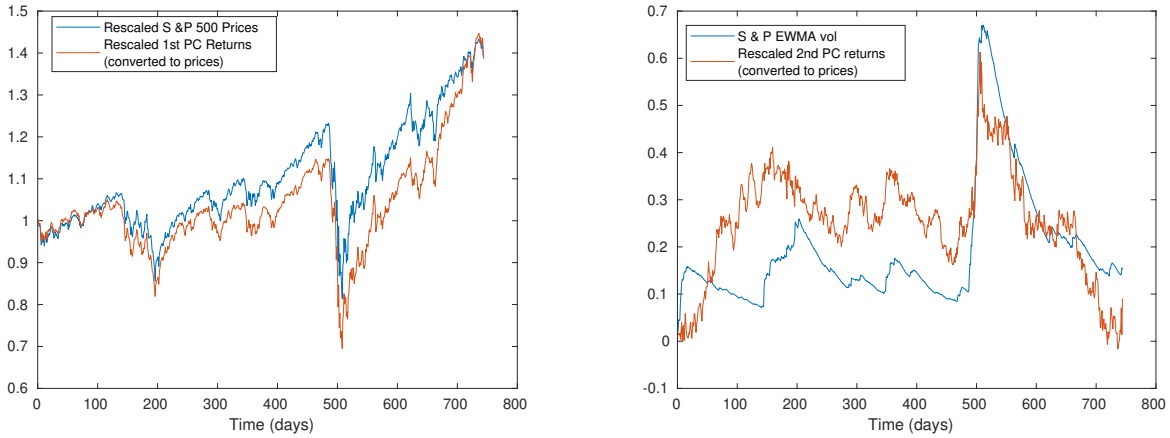


SPXAV10P (NN) 70%/30%

Figure 4.8: Comparison of NN and GPR models for SPXAV10P in and out of sample for various train/test splits

proxy choices, we are satisfied with the R^2 of over 0.907 in case of GPR and slightly less for the NN with 10 layers. In addition, inclusion of more indices to represent the broader equity market seems to solve the vanishing gradients issue in the impractically deep NN.

Looking at the striking difference in performance between NN and GPR, we conclude that we obtain the best R^2 out of sample with CORE data (generalized economy) compared to known custom index components used as features. This result is counterintuitive at first, so we seek an explanation via principal component analysis. Unsurprisingly, the principal components of the CORE data set daily returns corresponding to the two largest eigenvalues are related to "market movements" and "market volatility" (see Figure 4.9). This gave us confidence that the effects our ML models were picking up are real and explainable in our context. Despite the relatively short time window of the CORE data (approximately three years), the core data set contains a great deal of cross-sectional data across many indices, as evidenced by the total number of returns column in Table 4.5. As a result, we can extract the equivalent information on exposure to a wider set of market factors compared to the deeper historical data set of just a few indices.



(a) (b)

Figure 4.9: Principal components corresponding to the largest two eigenvalues and (a) rescaled S&P returns, (b) volatility

4.2.4 Results

A Single Annuity

Having established the practicality of using the CORE set of indices in conjunction with ML to reproduce risk-controlled indices within FIAs, we would like to apply this to accomplish our goal of annuity interest credit forecasting. With simplicity in mind, we looked only at participation strategies of the kind summarized in Eq. 4.1. Not only are these easy to reproduce, but the resulting return distribution is more closely related to the underlying index distribution. Any underlying issues with the index calculations are easier to spot in these annuity crediting rates.

As a base case we simulated the CORE economy (equity and bond indices from Table 4.4) using a multivariate GBM[8]. That is, we modelled P_t and B_t by a GBM with regards to the FIA summary in Eq. 4.1. These multivariate normal returns were then used as the input into our ML-obtained RC index function (f in Eq. 4.1). The distributions of FIA returns based on this scenario are shown in Figure 4.10 in red and orange for GPR and NN respectively. Alternatively, we could simulate the RC index directly as a GBM, which means simulating I_t from Eq. 4.1 with a GBM directly. In Figure 4.10 the distribution of FIA returns obtained by this direct index simulation via GBM is shown in blue.

Surprisingly, at first glance, this does not appear to translate into comparable cumulative annuity returns over a 10-year period. What is surprising is not that the index itself does not seem to evolve according to a GBM (Figure 4.1) but rather that using a GBM-based input to an RC index function produced such narrow distributions. Consider one example of the FIA with participation strategy, struck on SPXAV10P and a 55% participation rate over a 10-year accumulation period. Leveraging our ML models described in Section 4.2.3 with the CORE equity and I00078US as an interest rate input[9] we simulated a sample of annuity returns (see Figure 4.10, left panel).

We observe that distributions of FIA returns obtained via ML are wildly different than

[8]We estimate the parameter directly by fitting a multivariate normal to the daily returns in the dataset.

[9]Specifically, U.S. Treasury Bills 1–3 Months index. As discussed in 4.2.3, the choice of fixed-income exposure seems to make little difference to the final index result.

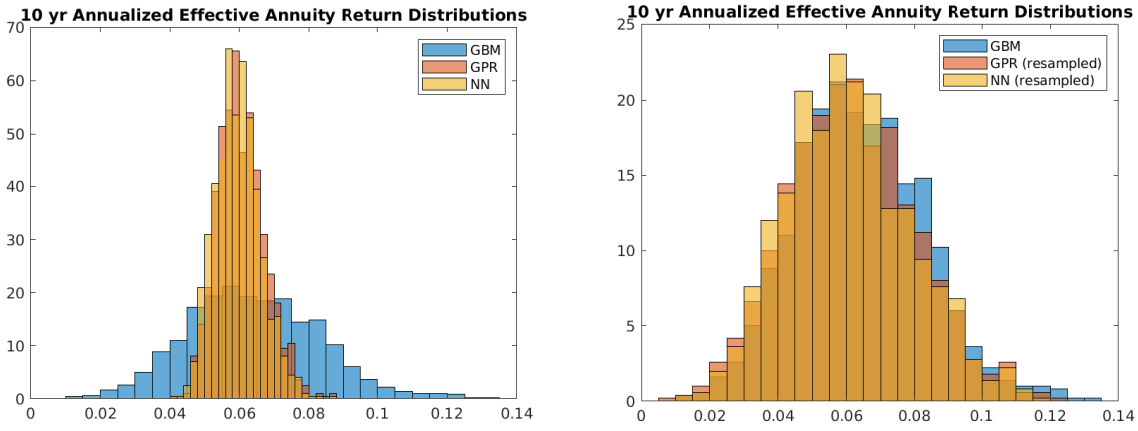


Figure 4.10: FIA return distributions for 1000 simulations for SPXAV10P with 55% participation, alongside the same simulation results with resampled monthly returns before crediting to the FIA. On the left, note that the relative performance between the two products is driven more by the mean as the ML models have narrower yield distributions.

those obtained through direct index modelling via the GBM. Despite reproducing the mean we can see the shape of the distributions of FIA returns is very different.

We initially suspected that our ML models were poorly standardized. Not only was this not the case, but we were happy to observe that the annual credited returns are *essentially identical*. In Table 4.6 we look at the yield Y computed from annual credited returns X_i such that $(1+Y)^{10} = \prod_{i=i}^{10}(1+X_i)$. We found that the X_i's are correctly distributed but the distributions diverge when looking at the total returns, $\prod_{i=i}^{10}(1+X_i)$. The only way

		GBM	GPR	NN
Actual total returns (i.e. not assuming independence between years)	Mean	6.43%	6.05%	5.95%
	SD	1.90%	0.68%	0.65%
	Skewness	28.56%	44.21%	39.66%
Total returns assuming independence between years	Mean	6.43%	6.25%	6.12%
	SD	1.90%	1.87%	1.81%
	Skewness	28.56%	42.75%	41.98%

Table 4.6: 10-year annualized FIA return statistics

to reconcile these facts is to realize that the X_i's are not independent. To support this we resampled the monthly returns of our ML-derived index returns and redid the experiment with the annuity returns (Figure 4.10, right panel).

The intuition is as follows: even though the features we used (underlying economy returns) were simulated via GBMs, the ML model is picking up on the volatility calculations in SPXAV10P. Even though there is no autocorrelation structure in the case of a pure GBM, in our ML case large swings are followed by a higher "estimate" for $\hat{\sigma}$ and π_t (see Eq. 4.5), dampening returns in the next period. This is the key benefit of the RC indices as marketed. Even though the volatility is constant in our GBM inputs, *sample* volatility is not, and so π_t will be non-constant. This will introduce a native autocorrelation to the RC index output of our ML model, effectively independent of the model for underlying components.

Compared to the results with the same products in Section 4.2.1, we would expect some change in the distribution (e.g. perhaps some reduced variance). The results shown by Figure 4.10, on the other hand, seem too extreme. Even ignoring the data, consider the following heuristic argument. The original RC index is constructed to aim at a value of $\sigma = 0.10$, so the participation of 55% should give $0.55\sigma = 0.055$ and the floor at zero percent will *roughly* halve this again to give $\approx 0.25\sigma$ at 0.025 or 2.5%, on the same order of magnitude of our observations. But a volatility of less than one percent, especially given the high returns of $\approx 6\%$, seems counterfactually low and defies financial sense. So, regardless of the choice of benchmark (direct GBM or historical data) for our results, we can see that the shape of our distribution is too narrow and the standard deviation of rates is an order of magnitude too small. These findings motivate the results studied in Part 4.3.

Multiple Annuities

Before moving on to Part 4.3, in which we will study the effects of dependent returns, we take a moment to explore their importance. Recall the original goal was a comparison between two annuities linked to different indices. For this, we need to perform simul-

	SPXAV10P with 55% par		SPMARC5P with 85% par		
	Mean	SD	Mean	SD	$P(\text{SPX} > \text{MARC5})$
GBM	0.059	0.018	0.063	0.016	42.8%
GPR	0.050	0.006	0.065	0.004	2.2%
NN	0.049	0.006	0.057	0.004	11.8%

Table 4.7: Summary statistics, comparison of annualized returns for two annuities struck on two different indices

taneous simulations of two indices, comprised of different components and with different participation rates. Here we add Standard and Poor's MARC 5% Excess Return Index (SPMARC5P), which is similar to the other two indices explored previously but with an added gold index component. For this index, the prevailing participation rate (in late 2021) is set at 85% for the strategy. We simulated the same CORE economy with the addition of a gold index, in hopes that this would improve our estimates' performance, and repeated the same procedure as before. Figure 4.11 shows similar behaviour. It is worth noting that in Figure 4.12 we see a larger difference between the NN and GPR results in the MARC5 case than for SPXAV10P. This is not too surprising as the MARC5 index is a more complex product and the two methods likely struggle to sufficiently capture the index dynamics to the point that they converge.

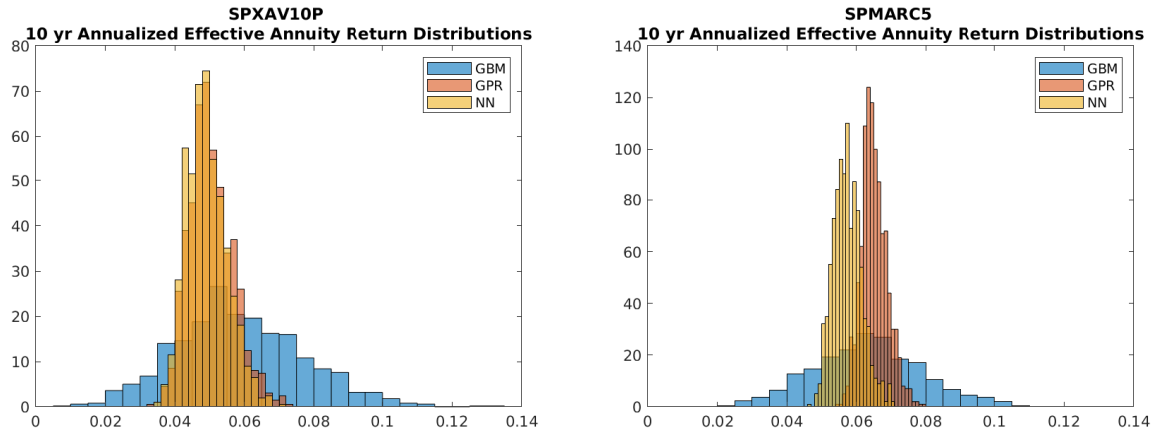


Figure 4.11: The MARC5 yield results have similar properties to the SPXAV10P product.

Not only do we see the same dramatic change in the summary statistics, we also see a striking change in relative performance. This is to be expected as the lower variance drives a narrower yield distribution and therefore less overlap and more dramatic relative performance, especially compared to the GBM base case (see Figure 4.12 and Table 4.7).

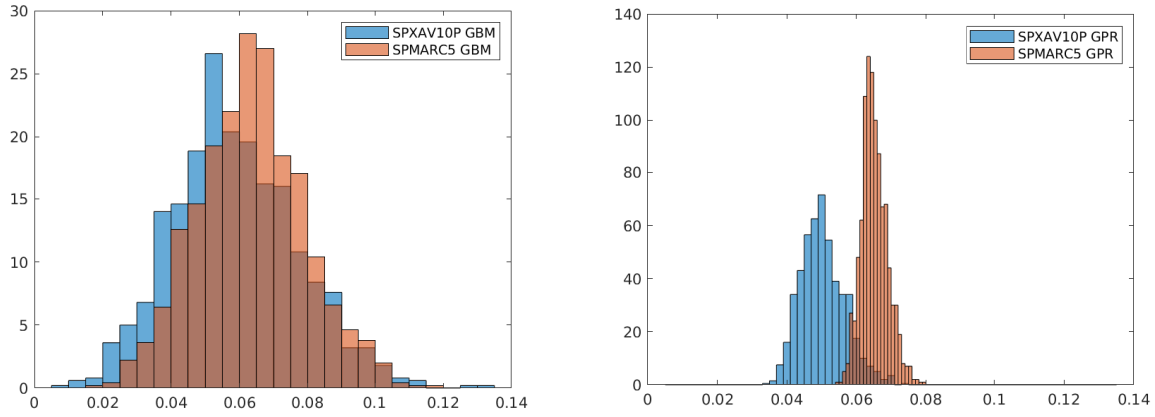


Figure 4.12: The relative performance between the two products is driven more by the mean, as the ML models have narrower yield distributions.

While these results were not what we had expected to uncover, they are nevertheless very encouraging. An appropriate and well-trained ML model was able to capture the very feature on which these indices are marketed, namely volatility control. However, we do see the overreaction of the ML-trained index functions due to the insufficiently sophisticated "economy" simulation model. This suggests (as will be seen in Part 4.3) that a more sophisticated CORE economy model is likely needed in place of correlated GBMs. To obtain more realistic results, the autocorrelation that affects variance through aggregate returns in the index function must interact with some autocorrelation native to the CORE set of indices. These issues, as well as the features needed in the CORE model and how to obtain proper annuity returns, are the topics we explore in Part 4.3.

4.3 Intertemporal Correlation: Risk Control Indices and Volatility Modelling

4.3.1 FIA Index Construction and Implications

In this second part of the chapter we aim to explain the results of Section 4.2.4. We first provide some context relevant to FIA construction. As mentioned in Section 4.2.2, the low interest rate environment that resulted from the Great Financial Crisis was the main motivator for the design of risk-controlled indices. With historically low yields for Treasuries and corporate bonds, insurance providers needed a new product to keep annuity returns competitive. Providing index exposure with a principal guarantee is one way to do this. However, reliance on a broad market index with high volatility (e.g. the SPX) is not tenable in an economic environment where the option budget available under low interest rates renders the crediting rate unattractively low to annuity buyers. The solution widely adopted by the industry was RC indices of the kind discussed in Section 4.2.2. Risk control mechanisms modulate exposure to the respective primary indices to make hedge positions cheaper and pricing of the associated options more stable. Ultimately, an entity in the supply chain has to have the underlying index or components on their books somehow. An FIA provider needs to take some of the premiums received for the FIA and purchase an option or portfolio of options giving them the requisite exposure to the index to credit the FIA returns. The option provider in turn needs to hedge this exposure to the index in the open market. A nice schematic of this process can be found in [Watson, 2022].

This process greatly constrains the kind of index behaviour we would expect to see. Any successfully marketed index with sufficiently idiosyncratic behaviour would very quickly run into capacity issues, with hedge providers trying to create exposure and hedging in the open market. Indices therefore must be constructed from components that are very liquid or similarly exposed to broad risk factors to make substitution possible (e.g. hedging the SPX with another "market proxy"). For instance, in Section 4.2.3 we explored (via PCA) the possibility that our CORE economy returns were mostly driven by a market and volatility risk factor. Risk-controlled indices must necessarily be driven by exposure

to these same factors; otherwise, the hedge provider would not be able to sell the necessary options to make a market for FIAs possible. Likewise, the risk control mechanisms cannot be too complex and sensitive, or the number of transactions needed to hedge would be cost-prohibitive. We argue as evidence of this the fact that RC indices still have sufficiently interesting intertemporal return dependence that, when passing a GBM through the RC mechanism and aggregating returns over an FIA crediting period, we get counterfactual results[10] of the kind in section 4.2.4. Put another way, while RC indices may have been motivated in part by FIA products, they have a life separate from FIAs and are subject to all the same construction and modelling challenges present with any index. This lies at the heart of our results in this chapter part.

In Part 4.2, we discussed how to recreate cross-sectional index returns by reproducing the index calculations themselves and passing a smaller CORE economy through them. We saw that the simulated economy needed certain essential features that interact with the index calculations. These are features that simple GBMs do not possess. In this chapter part, we explore what those essential features may be, and which are sufficient to reproduce reasonable FIA returns. We do this by first studying the stylized facts of indices *in general* and how these affect the aggregation of returns in Section 4.3.2. In Sections 4.3.2, 4.3.2 and 4.3.2 we apply this to the popular class of stochastic volatility models, the very class of model that inspired the RC mechanisms we are interested in. We then return to RC indices specifically in Section 4.3.3 before finally addressing the question of FIA returns in Section 4.3.3. A brief conclusion tying the results of Parts 4.2 and 4.3 together follows.

4.3.2 Stochastic Volatility: Statistics of Indices

As a general rule, asset prices are often less useful than relative asset growth or loss. This is why we employ the notion of a return. Given a time series of prices $\{P_i\}_{i=1}^N$ sampled at regular intervals of time Δt (e.g. daily), we consider two different notions of asset returns,

[10] To note briefly, dependence in the aggregation of returns is a known, perhaps under-investigated phenomenon in finance generally. See for instance [Schwartz and Whitcomb, 1977],[Perry, 1982],[Baltussen et al., 2019] or [Xiao et al., 2021].

the simple or gross return (R_i) and the logarithmic return (r_i):

$$R_i = \frac{P_i - P_{i-1}}{P_{i-1}}, \quad r_i = \log\left(\frac{P_i}{P_{i-1}}\right).$$

Comparing the two, the key point is that logarithmic returns are additive across periods of time while simple returns are additive across portfolio components. What is often lost is the understanding that these are *statistics* of the price data, with all the associated statistical questions and issues. For instance, a gross return is clearly most useful when describing a quantity that grows additively or linearly in time, whereas a logarithmic return captures the rate of an exponential growth. Depending on the data, the context and application of these two notions of return are not a priori interchangeable. The difference can be captured when considering geometric Brownian motion (GBM), a common null model of financial returns going back to the work of [Merton, 1969], [Samuelson, 1973] and the thousands of works derived since:

$$\frac{dP_t}{P_t} = \mu dt + \sigma dW_t \Longrightarrow P_t = P_{t-1}\exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma W_{\Delta t}\right\}. \tag{4.6}$$

GBM growth is inherently geometric or multiplicative: $r_t \sim \mathcal{N}((\mu - \frac{\sigma^2}{2})\Delta t, \sigma^2\Delta t)$, whereas R_t will be a shifted Log-Normal distribution with mean $\exp(\mu\Delta t) - 1$ and variance $[\exp(\sigma^2\Delta t) - 1]\exp(2\mu\Delta t)$. From a statistical perspective, logarithmic returns are favourable, as $\frac{r_t}{\Delta t}$ will be scale-free and the mean will report a typical portfolio trajectory (for recent discussions of the relevance of this, see [Peters, 2019] and [Carr and Cherubini, 2022]). Despite this, simple returns are often used in many financial calculations (e.g. annuity crediting formulas) of interest to us.

Fortunately, at small enough time scales these two notions of return are functionally interchangeable. To see this we need only consider the Taylor approximation to $\exp(x)$:

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \exp(r_t) - 1 \approx r_t.$$

Consider the justification for the GBM as a suitable null model of financial asset dynamics. Regardless of the kind of return considered, in any event we can express longer time returns using a sum of log price ratios:

$$1 + R_T = 1 + \frac{P_T - P_0}{P_0} = \prod_{t=1}^{T} \frac{P_t}{P_{t-1}} = \exp\left(\sum_{t=1}^{T} \log\left(\frac{P_t}{P_{t-1}}\right)\right) \longrightarrow \text{LogNormal by CLT}$$

$$r_T = \log\left(\frac{P_T}{P_0}\right) = \log\left(\prod_{t=1}^{T} \frac{P_t}{P_{t-1}}\right) = \sum_{t=1}^{T} \log\left(\frac{P_t}{P_{t-1}}\right) \longrightarrow \text{Normal by CLT}$$

As long as the $\log\left(\frac{P_t}{P_{t-1}}\right)$ terms are sufficiently i.i.d. and well behaved, we can justify the GBM as a model by invoking the CLT. Indeed, at long enough time scales this model does appear sufficient ([Cont, 2001]).

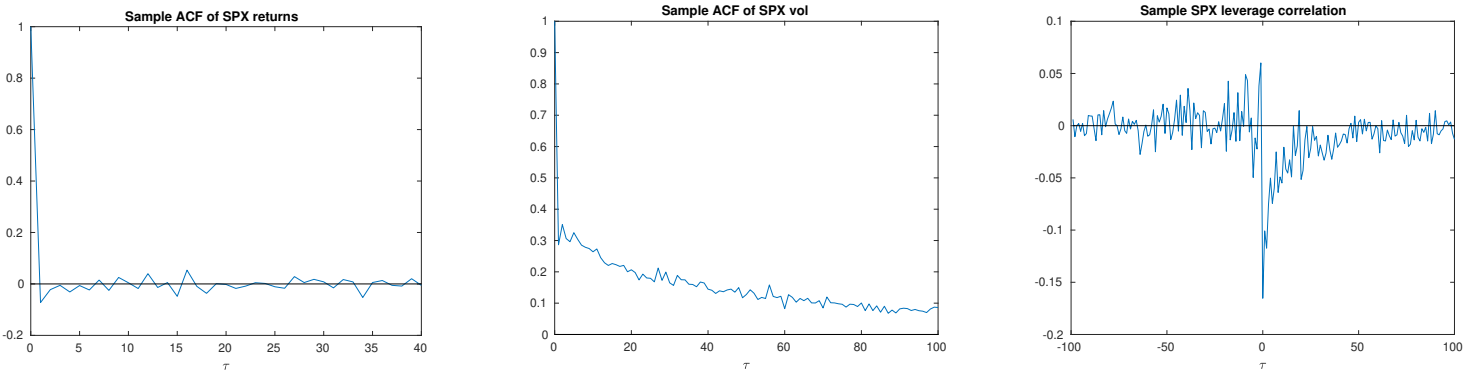
That said, the exact nature of the temporal dependence of returns is the subject of much debate and from some perspectives can be thought of as *the* open problem of mathematical finance. Investigators often turn to statistical autocorrelation functions (ACFs) to provide insight. Consider the autocorrelation of S&P returns (i.e. $\mathbb{Corr}[r_t, r_{t+\tau}]$). Figure 4.13a suggests that any dependence between returns is nonlinear, as returns are not linearly correlated through time. Indeed many nonlinear functions of returns express nontrivial correlations (again see [Cont, 2001] and references therein). Nonlinear dependence can be expressed as the sum of contributions of correlations between powers of returns via Taylor series. So it is useful to consider products of powers of returns alone. For instance, in finance we often consider the magnitude of squared returns such as the volatility squared (typically denoted by σ_t^2).

Let us define the instantaneous squared volatility at time i by $\hat{\sigma}_i^2 = \frac{r_i^2}{\Delta t}$ (or $\frac{R_i^2}{\Delta t}$ where $R_i \approx r_i$). Two significant nonlinear correlations appear to be persistent features of many traded assets across different kinds of markets (see previous references, for example [Cont, 2001]). We demonstrate them here with daily S&P 500 (SPX) returns:

1. Volatility autocorrelation: In Figure 4.13b we plot the autocorrelation of instantaneous volatility in terms of a fixed lag τ via $c(\tau) = \mathbb{Corr}[\hat{\sigma}_t, \hat{\sigma}_{t+\tau}]$. The results strongly

suggest that there is a persistent sizable dependence in volatility across time.

2. Leverage correlation: In Figure 4.13c we plot $L(\tau) = \mathbb{C}\mathbb{o}\mathbb{r}\mathbb{r}\left[\left(\frac{r_t - \mathbb{E}[r_t]}{\Delta t}\right), \hat{\sigma}^2_{t+\tau}\right]$. Large negative price movements seem to be correlated to increases in volatility that decay after the initial shock. The correlation is generally strongest following negative returns; that is, volatility clustering cannot meaningfully "anticipate" a sudden price drop.



(a) (b) (c)

Figure 4.13: Sample autocorrelation functions of (a) returns, (b) volatility and (c) the leverage correlation

Higher-order correlations may have some impact on prices, but for reasons of both statistical quality and modelling convenience we typically restrict ourselves to volatility and leverage. This motivates the common strategy of replacing the volatility σ in Eq. (4.6) by a stochastic process σ_t with a positive decaying autocorrelation. In order to induce the leverage correlation, a dependence between σ_t and W_t is typically imposed. Now reconsider the issues raised by Figure 4.10 and Table 4.6, as well as the analogous results for multiple annuities in Section 4.2.4. To understand how the variance of returns can depend on the aforementioned correlations, we look at the variance of the sum of log returns given by $\mathbb{V}\mathbb{a}\mathbb{r}[\sum_{i=0}^{N} r_i]$.

If we take $dP_t = \mu P_t dt + \sigma_t P_t dW_t$ then we simply have

$$r_T = \log\left(\frac{P_T}{P_0}\right) = \int_0^T d\log(P_s) = \int_0^T \left(\mu - \frac{\sigma_t^2}{2}\right) dt + \int_0^T \sigma_t dW_t$$

which is a GBM-like formula with a stochastic volatility.

To explain the aggregated variance of index returns, we aim to decompose $\mathbb{Var}[r_T]$ into a sum of the contributions from autocorrelation and leverage. To do so we will need some assumptions about the form of σ_t and possible dependence on r_t. In the rest of this section we will study models of the following form:

$$\begin{aligned} r_T &= \int_0^T \left(\mu - \frac{\sigma_t^2}{2}\right) dt + \int_0^T |\sigma_t| dW_t^{(1)} \\ \sigma_t &= m + \lambda \int_{t_0}^t k(t,t') dW_{t'}^{(2)} \\ \rho &= \mathbb{E}[W_t^{(1)} W_t^{(2)}]/t \end{aligned} \tag{4.7}$$

This general framework can be augmented further for realism. For example we could make the change $|\sigma_t| \to f(\sigma_t)$. The choice of $f(\cdot)$ would control the ultimate stationary distribution (if it exists) of volatility; for instance, $f(x) = e^x$ would produce Log-Normal distributions for volatility ([Tegnér and Poulsen, 2018]). In the remainder of this section we will consider the model presented above and focus on the choice of volatility kernel $k(t,t')$, which will determine the stochastic volatility model under consideration. This model for r_t and σ_t is flexible enough to capture many cases but parsimonious enough to compute the variance we are interested in:

Theorem 4.3.1. *If returns are generated by a model of the form (4.7), then*

$$\mathbb{Var}[r_T] = \underbrace{\mathbb{E}\left[\int_0^T \sigma_s^2 ds\right]}_{=\sigma^2 T \text{ in GBM}} + \int_0^T \int_0^T \left(\underbrace{\frac{1}{4} C(t,u)}_{\text{vol. autocorr.}} - \underbrace{2\rho\lambda\Theta k(t,u) H(t-u) \mathbb{E}\left[\sigma_t \sigma_u\right]}_{\text{leverage corr.}} \right) dt du \tag{4.8}$$

where $C(t,u) = \mathbb{E}[(\sigma_t^2 - \mathbb{E}[\sigma_t^2])(\sigma_u^2 - \mathbb{E}[\sigma_u^2])]$, $\Theta = P(\sigma_u > 0) - P(\sigma_u \leq 0)$ *and* $H(\cdot)$ *is the*

Heaviside step function.

We can simplify Eq. (4.8) further via Lemma D.1.1 by making the substitution $C(t,s) = 2(\mathbb{E}\left[\sigma_t\sigma_u\right]^2 - m^4)$. This conveniently allows us to compute only one model correlation. The leverage term in Eq. (4.8) is so called for the following reason: if we say $r_t \approx \frac{dP_t}{P_t}$ and $\Delta t \approx dt$ and use the standard Ito conventions and results from Appendix D.2, we can show that

$$\mathbb{E}\left[\left(\frac{r_t - \mathbb{E}[r_t]}{\Delta t}\right), \hat{\sigma}_u^2\right] \approx 2\rho k(u,t)\theta(t-u)\mathbb{E}[\sigma_t\sigma_u]; \tag{4.9}$$

that is, we can see that the leverage term in (4.8) is proportional to the leverage correlation $\mathcal{L}(\tau)$. So we have the expression we were looking for: the variance of r_T in terms of the two significant correlations we have discussed.

Many indices conform to the S&P pattern laid out in Figs. (4.13b) and (4.13c), namely a positive volatility autocorrelation and a negative leverage correlation implying $\rho < 0$. A negative ρ implies in turn that the sum of the correlation terms in Eq. (4.8) is positive, so that the variance of the T-period return distribution will be larger given intertemporal dependence than otherwise. RC indices are constructed specifically with such a dependence induced by stochastic volatility of the kind in Eq. (4.7) in mind. Indeed, as the RC mechanism attempts to dampen the shocks of stochastic volatility (shrinking the leverage and autocorrelation terms) we should see smaller T-period return variance in the RC index versus the underlying index.

We now apply Eq. (4.8) to some well-known stochastic volatility models in the form of (4.7) to gain some insight, keeping in mind that to fit the stylized features of index data, we should see leverage correlation and volatility autocorrelation as well as a variance that will be asymptotically linear to conform to a GBM-like behaviour seen at large time scales.

The Ornstein–Uhlenbeck (OU) Model

We will start with perhaps the simplest well-known stochastic volatility model: the OU process. This process can be put in the form of our model above via an exponential kernel:

$$d\sigma_t = -\alpha(\sigma - m)dt + \lambda dW_t^{(2)} \Longrightarrow \sigma(t) = m + \lambda \int_{-\infty}^{t} e^{-\alpha(t-s)} dW_s^{(2)}. \tag{4.10}$$

So we use the following in Eq. (4.8):

$$\begin{aligned} k(t,u) &= e^{-\alpha(t-u)}, \\ \mathbb{E}\left[\sigma_t \sigma_u\right] &= m^2 + \left(\frac{\lambda}{2\alpha}\right) e^{-\alpha|t-u|}. \end{aligned}$$

Let us denote $\left(\frac{\lambda}{2\alpha}\right)$ by θ. For the OU model described in Eq. (4.10) we can show that

$$\begin{aligned} \mathbb{Var}[r_T] = (m^2+\theta)T + \theta(\theta - 2\lambda\Delta\rho)\left[\frac{T}{2\alpha} - \frac{1-e^{-2\alpha T}}{(2\alpha)^2}\right] \\ + m^2(2\theta - 2\lambda\Delta\rho)\left[\frac{T}{\alpha} - \frac{1-e^{-\alpha T}}{\alpha^2}\right]. \end{aligned} \tag{4.11}$$

Consider how this behaves asymptotically. For a small-T approximation, take the Taylor series of (4.11):

$$\mathbb{Var}[r_T] \approx (m^2+\theta)T + \alpha\left[2\theta(\theta - 2\lambda\Delta\rho) + m^2(2\theta - 2\lambda\Delta\rho)\right]T^2 + \mathcal{O}(T^3). \tag{4.12}$$

For a negative ρ we will see a superlinear growth in the variance, whereas for large T,

$$\mathbb{Var}[r_T] = \mathcal{O}\left(\left(m^2 + \theta + \frac{\theta(\theta - 2\lambda\Delta\rho)}{2\alpha} + \frac{m^2(2\theta - 2\lambda\Delta\rho)}{\alpha}\right)T\right). \tag{4.13}$$

For large enough T the behaviour will become linear.

It is also worth making note of another asymptotic regime. For $\alpha \to \infty$ we can see that

$$\mathbb{V}\mathbb{a}\mathbb{r}[r_T] = \mathcal{O}((m^2 + \theta)T) \tag{4.14}$$

where we make use of big-O notation to avoid taking explicit limits as $\theta \to 0$. Intuitively, in the large α limit the characteristic scale of the volatility correlation $(1/\alpha)$ goes to zero and the process starts to resemble a GBM. This may also explain why this particular model is popular in many applications: it allows for the creation of a non-constant volatility surface in options modelling without deviating too much from a GBM. To see this, consider a special case of the OU model, the Heston stochastic volatility model ([Heston, 1993]).

It is well known that if $m = 0$ a sum of squared OU processes can be used to create a Cox–Ingersoll–Ross (CIR) model, the same process used in the Heston model. Proceeding from our single OU process, we have the following correspondence:

$$\begin{aligned} d(\sigma_t^2) &= 2\sigma d\sigma_t + \lambda^2 dt \\ &= \underbrace{(2\alpha)}_{\kappa}\Big(\underbrace{\frac{l^2}{2\alpha}}_{\theta} - \sigma^2\Big)dt + \underbrace{(2\lambda)}_{\epsilon} dW_t^{(2)}. \end{aligned}$$

If $m = 0$ then $\sigma_t \sim \mathcal{N}(0, \theta)$ and $\Delta = 0$. Making use of Eqs. (4.8) or (4.11), we can show the variance of aggregate returns in the Heston model is

$$\begin{aligned} \mathbb{V}\mathbb{a}\mathbb{r}[r_T] &= \theta T + \left(\theta^2 - \Delta\rho\theta\epsilon\right)\left(\frac{T}{\kappa} - \frac{1 - \exp(-\kappa T)}{\kappa^2}\right) \\ &\approx \theta T + \kappa\theta^2 T^2 + \mathcal{O}(T^3) \end{aligned}$$

where the second expression is the "small T" approximation. While we still see the same behaviour as in the OU model, the deviations from a GBM are less severe. For a large enough κ the differences are inconsequential in practice and the sample variance of returns will scale mostly linearly.

Rough Volatility: The Gatheral or Stein & Stein Model

Many authors have pointed out the insufficiency of the Heston/OU model in a variety of ways ([Daniel et al., 2005] and references therein). In particular, volatility autocorrelation decays too rapidly as an exponential in the model, whereas the data suggest something more akin to a power law. Some authors suggest that a sum or nested set of (potentially infinite) OU processes can account for this and would reflect the different time scales of market participants ([Bochud and Challet, 2007], [Perelló et al., 2008], [Barndorff-Nielsen and Shephard, 2001] and [Vitali et al., 2019]). Others have attempted to create an entirely new class of stochastic volatility models using tools beyond the standard Ito process. One such tool is the fractional Brownian motion (FBM) denoted by

$$W_t^H = \frac{1}{C_H}\int_{-\infty}^{t}\left((t-s)_+\right)^{H-1/2} - \left((-s)_+\right)^{H-1/2}\Big)\, dW_s$$

where C_H is a normalization constant. The parameter $H \in (0,1)$, called the Hurst index, controls the degree of mean reversion or momentum for the process. We have chosen the value of $C_H = \int_{-\infty}^{0}\left(1-s\right)^{H-1/2} - (s)^{H-1/2}\Big)\, dt' + \frac{1}{2H}$ in order to give a convenient expression for variance,

$$\mathbb{V}\mathbb{a}\mathbb{r}[W_t^H - W_u^H] = |t-u|^{2H}.$$

Assuming that $t, u > 0$, we can use this to derive the autocorrelation:

$$E[W_t^H W_u^H] = \frac{1}{2}(t^{2H} + u^{2H} - |t-u|^{2H}).$$

For $H < 1/2$ the process has negatively correlated values (i.e. mean-reverting behaviour), for $H = 1/2$ it is a standard BM and for $H > 1/2$ it has positively correlated values. For more details, see the original work introducing the FBM [Mandelbrot and Van Ness, 1968] or a more modern text such as [Holden et al., 2009].

For our purposes we will be studying a model inspired by [Comte and Renault, 1998] and [Gatheral et al., 2018]. Although it should be noted that in those references they made use of a Log-Normal model (i.e. $f(\sigma) = e^{\sigma}$ in 4.7). The model we study here will

be a more tractable "smoothed" model called the fractional Stein and Stein model after [Stein and Stein, 1991] and developed by Gatheral in [Gatheral et al., 2018]. This model is constructed via the OU process, but we replace the BM driving the process with an FBM:

$$d\sigma_t = -\alpha(\sigma - m)dt + \lambda dW_t^H \implies \sigma(t) = m + \lambda \int_{-\infty}^{t} e^{-\alpha(t-s)} dW_s^H. \tag{4.15}$$

Asymptotically at least, models of the form of (4.15) will produce the power-law behaviour of volatility we see in the data ([Cheridito et al., 2003], Theorem 2.3). Unfortunately, it is difficult to express this model in the form of Eq. (4.7) (see e.g. [Alòs and Lorite, 2021], p. 121 for a comprehensive discussion). For the often-used *truncated* FBM $\tilde{W}_t^H = \frac{1}{C_H} \int_0^t (t-s)^{H-1/2} dW_s$ we can show that

$$k(t,u) = k(t-u) = \frac{1}{C_H} \left(\frac{d}{dx}\Big|_{x=(t-u)} \int_0^x (x-z)^{\alpha} e^{\alpha z} dz \right). \tag{4.16}$$

However, this comes at the cost of a covariance function that is more difficult to work with, involving hypergeometric integrals.

Let $\tilde{m}(t) = m + e^{-\alpha t}(\sigma_0 - m)$. Then

$$\sigma(t) = m + \lambda \int_{-\infty}^{t} e^{-\alpha(t-s)} dW_s^H = \tilde{m}(t) + \lambda \int_0^t e^{-\alpha(t-s)} dW_s^H.$$

In general, looking at financial data, α in Eq. (4.15) tends to be very small. In this regime (months to a few years), we will assume α is small enough to make the following approximation:

$$\sigma_t \approx \tilde{m} + \lambda(W_t^H - W_0^H) \tag{4.17}$$

which is much easier to work with:

$$k(t,s) = \int_{-\infty}^{t} \left((t-s)_+ \right)^{H-1/2} - \left((-s)_+ \right)^{H-1/2} \right)$$
$$\mathbb{E}\left[\sigma_t \sigma_u\right] = m^2 + \lambda^2 E[W_t^H W_u^H].$$

We will also make the simplifying assumption that $\Theta = 1$, implying that we have chosen parameters guaranteeing σ_t is positive[11]. With these simplifications, we have

$$\begin{aligned}\mathbb{Var}[r_T] = m^2 \left[T + C_1^{(m)} \lambda^2 T^{2+2H} - C_2^{(m)} \rho \lambda T^{3/2+H}\right] \\ + \left[\left(\frac{\lambda^2}{1+2H}\right) T^{1+2H} + C_1 \lambda^4 T^{2+4H} - C_2 \rho \lambda^{3/2} T^{3/2+3H}\right]\end{aligned} \tag{4.18}$$

where

$$\begin{aligned} C_1^{(m)} &= \frac{1+4H+4H^2}{2(1+H)(1+2H)^2}, \\ C_1 &= \frac{2+7H+4H^2}{4(1+4H)(1+2H)^2} - \frac{\sqrt{\pi}\Gamma(2+2H)}{8(2^{4H})\Gamma(3/2+2H)} - \frac{1}{8(1+2H)^2}, \\ C_2^{(m)} &= \frac{8}{3+8H+4H^2}, \\ C_2 &= \frac{16H}{(1+2H)^2(1+6H)} + \frac{3\Gamma(1/2+H)\Gamma(1+2H)}{\Gamma(5/2+3H)}. \end{aligned}$$

This is the small-T formula for the T-period variance analogous to Eq. (4.12) in the pure OU model. We would now like to know the large-T behaviour as in Eq. (4.13). While as previously mentioned we cannot give the full model the same treatment analytically, we can draw some conclusions.

It has been shown by [Zeng et al., 2012] that there exists a stationary distribution given by

$$p(t, \sigma_t | \sigma_0) = \frac{1}{\sqrt{2\pi\gamma^2(t)}} \exp\left\{-\frac{(\sigma_t - \tilde{m}(t))^2}{2\gamma^2(t)}\right\}$$

where the long-term variance is given by $\gamma^2(t) = 2H\lambda^2 e^{-2\alpha t} \int_0^t s^{2H-1} e^{2\alpha s} ds$. If $t \to \infty$ then $\tilde{m} \to m$ and if $H \leq 1/2$ then $\gamma^2(t) = \mathcal{O}(\frac{H\lambda^2}{\alpha} t^{-(1-2H)})$. Thus, the long-term distribution for r_T will become GBM-like with a constant volatility of m. The speed of this convergence is governed by two mean reversion components: the classical OU reversion governed by α

[11] Keep in mind that for large enough time we can also take into account the OU-style mean reversion.

and the FBM reversion governed by H.

Finally, consider the regime where we send $H \to 0$ (as we did for e.q. Eq. (4.14)). Equation (4.18) is uninformative as $H \neq 0$ in any FBM, but we can make the approximation that σ is nearly constant at $\mathbb{E}[\sigma_t^2] \approx m^2 + \lambda^2$, namely the lowest-order term in (4.18). Consider Figure 4.14, where we have chosen some reasonable model values and compared $\mathbb{V}\mathbb{a}\mathbb{r}[r_T]$ to the $H = 0$ case of $\sqrt{(m^2 + \lambda^2)T}$.

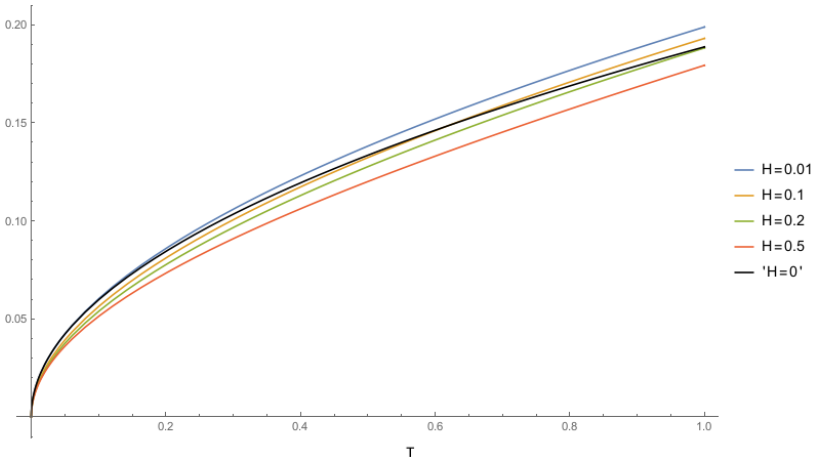


Figure 4.14: Variance for "small-time" r_T in rough model for $\lambda = 0.1, \rho = -0.5, m = 0.16$

Beyond Stochastic volatility?

In a GBM the standard deviation for both simple and logarithmic "small time" returns scale as $\sqrt{\Delta t}$ so that $\mathbb{V}\mathbb{a}\mathbb{r}[R_t] \approx \mathbb{V}\mathbb{a}\mathbb{r}[r_t] = \sigma^2 \Delta t$ until Δt grows too large so that this scaling only applies to logarithmic returns. We also saw this in (4.13) and the expression for $\gamma(t)$ in the previous sections for the OU and Gatheral models respectively. Both models eventually become GBM-like with respect to variance scaling. For long enough (> 1 year) time scales, as previously mentioned the GBM serves as a good model for equity asset returns. Indeed we will discuss how RC indices can dampen stochastic volatility resulting in smaller variances when considering the long-time emergent GBM behaviour of the index. For completeness, however, we now consider the possibility of some alternative index models

and mechanisms.

Let us denote by $S_{n\Delta t}$ and $s_{n\Delta t}$ the standard deviations of simple and log returns respectively over $n\Delta t$ time. Consider in Figure 4.15 the ratios $S_{n\Delta t}/S_{\Delta t}$ and $s_{n\Delta t}/s_{\Delta t}$ for the S&P 500 (SPX). We can see that the expected square root scaling does not hold for less than a year; the actual relationship appears to be closer to $n^{0.45}$.

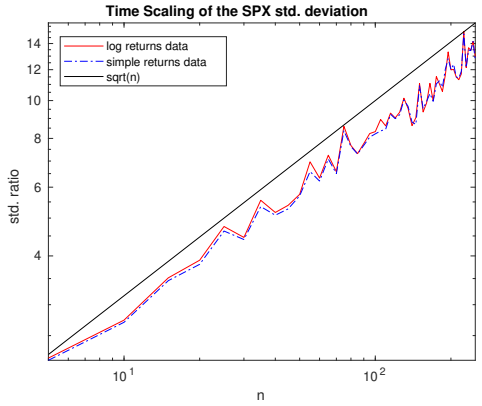


Figure 4.15: Log-Log plot of S and s ratios for $\Delta t = 1$ day

To infer that dependence between small time log returns is the culprit for the anomalous scaling in Figure 4.15, consider a counterfactual presented in Figure 4.16. In this case we have resampled across time the log $\left(\frac{P_t}{P_{t-1}}\right)$ terms in order to remove any time dependence. As we can see, the resampled return data conform to the square root scaling of a GBM, as the CLT convergence is not weakened by dependence in the summands. The difference in the standard deviation on the order of a year can be as much as 20% which will clearly be significant when considering portfolio returns.

Frustratingly, the two popular models of the form (4.7) we studied in the previous sections cannot explain the kind of anomalous scaling of the kind in Figs. (4.15) and (4.16). In the case of FBM driven model, however, a decrease in absolute variance can at least be achieved, albeit one that grows too quickly (see Figure 4.14). As we shall see in Section 4.3.3, models with this behaviour should prove satisfactory in capturing some essential features of the dependence structure of returns. We now discuss *why* this anomalous scaling was so hard to achieve and what if anything can be done.

In both Sections 4.3.2 and 4.3.2 our models possessed a stationary kernel, $k(t,u) = k(t-u)$, producing a stationary process. Indeed, in such a case we can show that

$$\mathbb{E}[\sigma^2] = \int_{-\infty}^{t} k^2(t-u)dt' = \int_0^{\infty} k^2(u)du.$$

Therefore the first term of (4.8) grows as T; and if $C(t,u) = C(t-u) = (k \star k)(t-u)$ (where $k \star k$ is the convolution of identical copies of the kernel), we have that

$$\int_0^T \int_0^T C(t,u)dtdu = T \int_{-T}^{T} C(s)\left(1 - \frac{|s|}{T}\right) ds$$

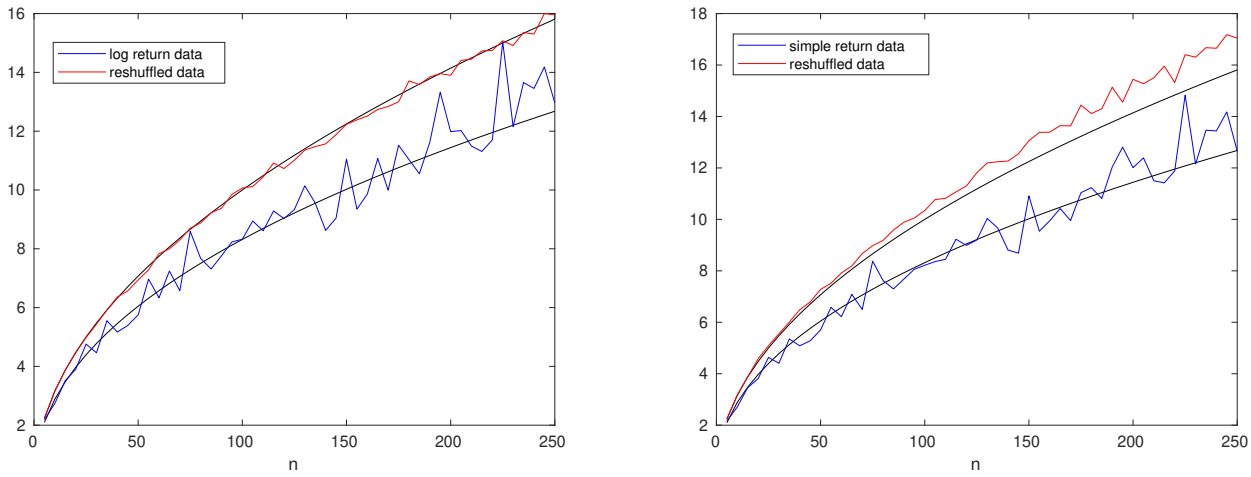
which necessarily grows at least as fast as T. So with the stationary models of the kind we have used, it is impossible to achieve the scaling of variance seen in Figure 4.15. There are two possible responses to this. First, we could abandon the form of (4.7) entirely: rather than model the volatility separately, we could just model the returns directly. So for example the following fractional GBM (studied by [Li et al., 2016] and others) would exhibit power-law behaviour for $\mathbb{E}[(r_t)^2(r_u)^2]$:

$$P_T = P_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma W_T^H\right\}$$

and we would see the anomalous scaling we desire:

$$\mathbb{V}\mathbb{a}\mathbb{r}[R_T] = VAR\left[\sigma W_T^H\right] = \sigma^2 T^{2H}.$$

Unfortunately, for Gaussian processes return and volatility autocorrelation would be linked by Lemma (D.1.1). This implies some kind of non-Gaussian rough model or multifractal model (see [Mandelbrot et al., 1997], [Bouchaud et al., 2000]). Alternatively, we could abandon models based on stochastic processes entirely. For instance, agent-based models ([Feng et al., 2012] or [Chen et al., 2013]) or even quasi-deterministic models ([Orlando et al., 2022] or [Guyon and Lekeufack, 2022]) have been able to capture many stylized features of markets.



(a) (b)

Figure 4.16: Permuted return data for (a) log and (b) simple returns more closely conforms to $n^{0.5}$ as expected vs $n^{0.45}$ (both drawn in black).

Of particular note is the recent work of [Baltussen et al., 2019]), who suggested that aggregated returns *do* appear to display serial correlations. The complex market operations used to hedge index exposure may influence the constituent securities that make up the index, with the resulting feedback leading to serial correlations. A more realistic upgrade to (4.7) may introduce a serially correlated substitute for $W^{(1)}$.

For now, it is important to show that stochastic volatility and intertemporal dependence therein can influence and explain the aggregated return variance of index (and by extension annuity) returns. When modelling RC indices for the purposes of FIAs, we can show volatility and leverage correlations of the original index are carried over to the risk control version. This will lead to issues when using the naive GBM models no matter how well one can reproduce the risk-control index calculations. Again, see Section 4.2.2 for a discussion of the relationship between risk control mechanisms and stochastic volatility in FIA indices.

4.3.3 RC index Simulation

The Role of Dependence

Thus far, we have only discussed the statistics of the SPX. However, our focus in this chapter is the SPX RC indices. As we saw in Section 4.2.1, the choice of intertemporal dependence structure of returns for the RC indices is extremely relevant to determining annuity returns. We have also discussed how the intertemporal dependence structure induced by stochastic volatility will have an impact on the aggregated return statistics of an index. So it is natural that this would be a consideration for RC indices as well.

Consider the volatility autocorrelation and leverage correlations of the SPXAV10P and the SPX together in Figure 4.17, and recall our outline of a generic RC mechanism in Section 4.2.2. It is to be expected that the RC mechanism "dampens" the volatility autocorrelation. The autocorrelation shows up in the clustering of large volatility events: periods of high volatility follow one another. The RC mechanism when encountering a high volatility event will shift index exposure to bonds lowering the volatility of subsequent returns.

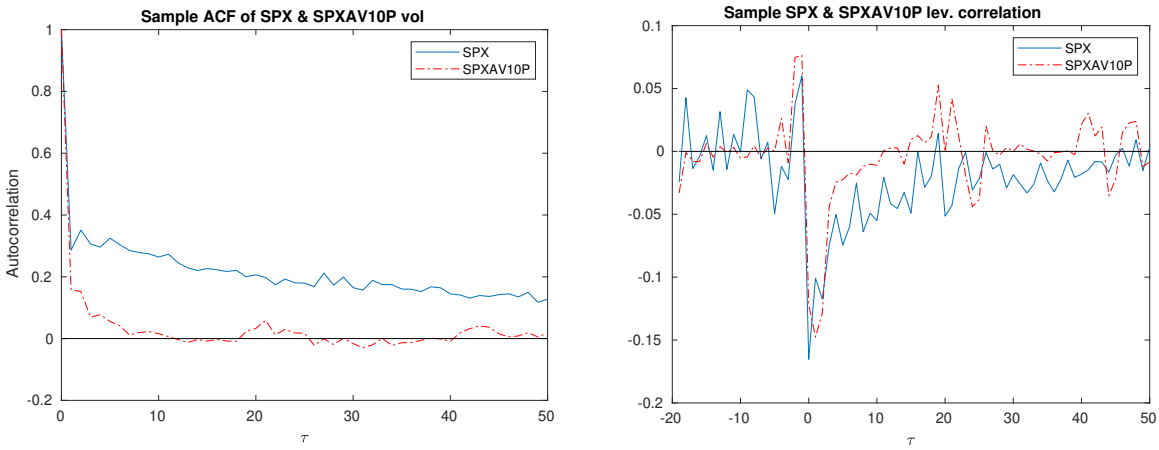


Figure 4.17: Volatility autocorrelation is damped by the RC mechanism (left) while the leverage correlation persists (right)

It is the RC index's leverage correlation that is more interesting. If the autocorrelation

is damped, we would expect to see something similar in the leverage correlation through some mechanism similar to that in Eq. (4.9). Instead, we see essentially the same pattern. It is tempting to think this is simply because of the normalization we have chosen (indeed the leverage *covariance* is lower in magnitude). To check this, we perform the following experiment: we fit an ML model for SPXAV10P as in Section 4.2.3 using the SPX and 100078US indices and consider the results when simulating the SPX via a GBM (Figure 4.18). In the language of (4.1) I is the SPXAV10P, P_t is the SPX, B_t is 100078US and f is our ML model. Considering the results in Figure 4.18, the autocorrelation is clearly even lower than the RC index and the leverage correlation is greatly reduced. There does appear to be a small amount of induced leverage, but this could easily be an artefact of the ML model imperfectly fitting the RC mechanism. Naturally, the GBM has no native autocorrelation to dampen and no leverage correlation, as volatility is not stochastic. On the other hand, the observed pattern of SPX volatility autocorrelation appears to have a power-law decay and is persistent over long periods of time (longer than the leverage correlation, which was also remarked on by [Perelló et al., 2004]).

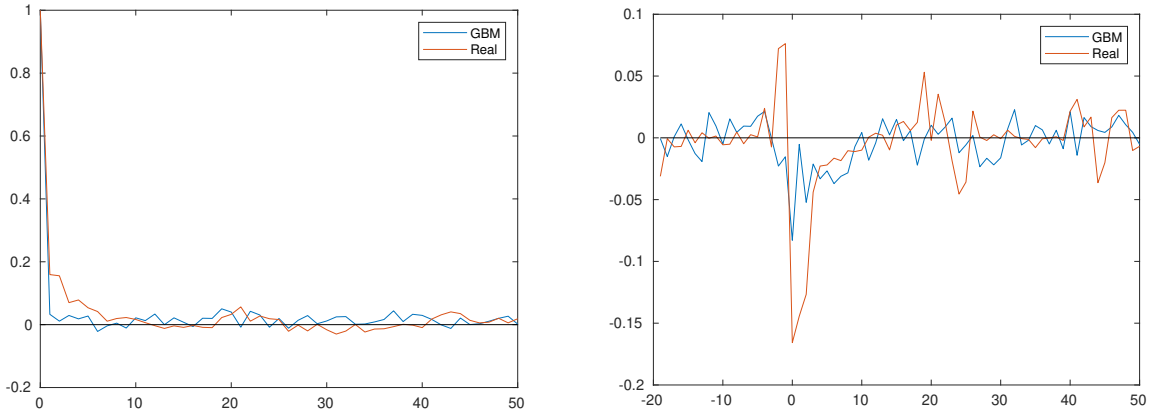


Figure 4.18: Unlike the volatility autocorrelation (left), the RC mechanism will induce a similar leverage correlation when fed GBM-generated returns.

These observations suggest that while the RC mechanism does reduce the intertemporal covariance of volatility (e.g. $\mathbb{E}[\sigma_t \sigma_u]$ in Eq. 4.8), the RC index still has a rich intertemporal

dependence of volatility and a stochastic volatility of its own inherited from the non-RC underlying index. Given the persistence of volatility autocorrelation, it stands to reason that even the most ambitious RC mechanisms would be unable to completely dampen it. This is also a practical issue: an index carrier could not purchase efficiently priced options from an index provider for a perfect RC index, as the provider mitigates volatility risk by adding a risk margin. Effectively, the hedge provider is pricing for volatility higher than the index is designed to maintain, as the trades required to perfectly conform to the RC mechanisms of (4.4) are not practical.

We now return our attention to the results of Section 4.2.4 for annuity returns using a GBM economy (CORE or otherwise). We would like to see index models with stochastic volatility become GBM-like on longer time scales, as seen in the data. We saw two examples of this in Sections 4.3.2 and 4.3.2. Through Eq. (4.8) we saw that the stochastic volatility dynamics at smaller time scales can aggregate to affect the resultant larger time-scale GBM. For instance, in Eq. (4.13) we saw that at large T the popular OU/Heston volatility dynamics show the variance to be linear in T, like a GBM, but is larger than it would be otherwise if there were no dependence structure to volatility. Indeed, if we were to use a model with no intertemporal dependence we would obtain

$$\mathbb{V}\mathrm{ar}[r_T] = \underbrace{\mathbb{E}\left[\int_0^T \sigma_s^2 ds\right]}_{\text{GBM volatility}} + \int_0^T \int_0^T \left(\frac{1}{4}\cancelto{0}{C(t,u)} - \cancelto{0}{\mathcal{L}(t-u)}\right) dtdu.$$

While the RC mechanism purports to dampen volatility, it does not completely eliminate intertemporal covariance as above. In Section 4.2.4 we used a GBM, which has no such dependence. We believe this explains the suspiciously extreme narrowing of the distributions in Section 4.2.4. At the current state of the art, as demonstrated in Section 4.2.1 we can fit the statistics of the resultant RC index directly, but if we want to produce an RC index by passing its constituent components through a model, we need an accurate model for the stochastic volatility that can interact with the RC mechanism.

In the following section we consider two modifications of our models in Sections 4.3.2 and 4.3.2, aiming to determine the necessary features of an accurate stochastic volatility

model.

Candidate models for Stochastic Volatility

If true, our hypothesis would suggest that any underlying simulation or economy must properly capture the proper decay of volatility autocorrelation such that the RC mechanism does not overpower it and deliver unrealistic results as in Section 4.2.4. We will test this using the following two models.

Model 1: An FBM model We will employ the model of [Gatheral et al., 2018], a version of which we studied in Section 4.3.2. We make the change that $f(\sigma) = e^{\sigma}$, as this is more realistic and empirically sound. In discrete daily returns the model takes the form

$$\begin{aligned} r_{t_i} &= \mu\Delta t + e^{v_{t_i}}\sqrt{\Delta t}\Delta W_{t_i}, \\ v_{t_i} &= v_{t_{i-1}} + \alpha(m - v_{t_{i-1}})\Delta t + \lambda\Delta W^H_{t_i}, \\ \rho\Delta t(\Delta t)^{2H} &= \mathbb{E}[\Delta W_{t_i}\Delta W^H_{t_i}]. \end{aligned} \tag{4.19}$$

Unfortunately, estimation within this model is severely complicated by the fractional term. We simply used the parameters found for the SPX in [Gatheral et al., 2018]. Simulation of an FBM is also difficult to achieve through some kind of Euler-style integration. We prefer to simulate the FBM through spectral methods for accuracy (see [Banna et al., 2019]); including a ρ and therefore leverage correlation is very challenging and we will have to eschew this for now, although as we shall see this will not materially affect our conclusions.

Model 2: A Moving Average Model We set $\Delta t = 1$ day such that for an N-day simulation horizon we have $t - t_0 = N\Delta t$ and $t_0 = t_{i-N} < ... < t_{i-j} < ... < t_i = t$. For a volatility kernel in the form (4.7), let $k(t_i, t_{i-j}) = k_{i-j}$ where $k_{i-j} \approx 0$ for some $j > q$. We

can then make the following approximation for a discrete daily volatility model:

$$\int_{t_0}^{t} k(t,t')dW_t^{(2)} \approx \sum_{j=1}^{q} k(t_i, t_{i-j})\Delta W_{t_j}^{(2)}.$$

Note that this is functionally equivalent to a moving average model $MA(q)$ (see [Kleiner, 1977]), and so fitting the following model with built-in MA tools is straightforward:

$$\begin{aligned} r_{t_i} &= \mu\Delta t + e^{v_{t_i}}\sqrt{\Delta t}\Delta W_{t_i}^{(1)}, \\ v_{t_i} &= m + \sum_{j=0}^{q} k_{i-j}\Delta W_{t_{i-j}}^{(2)}, \\ \rho(\Delta t)^2 &= \mathbb{E}[\Delta W_{t_i}^{(1)}\Delta W_{t_i}^{(2)}]. \end{aligned} \tag{4.20}$$

We do this for the SPX and compare the subsequent return distributions in Figure 4.19. We can see that the $MA(50)$ volatility model can reproduce real SPX returns more closely, notably in the tails of the return distribution where the deviations from normality are driven by high stochastic volatility events.

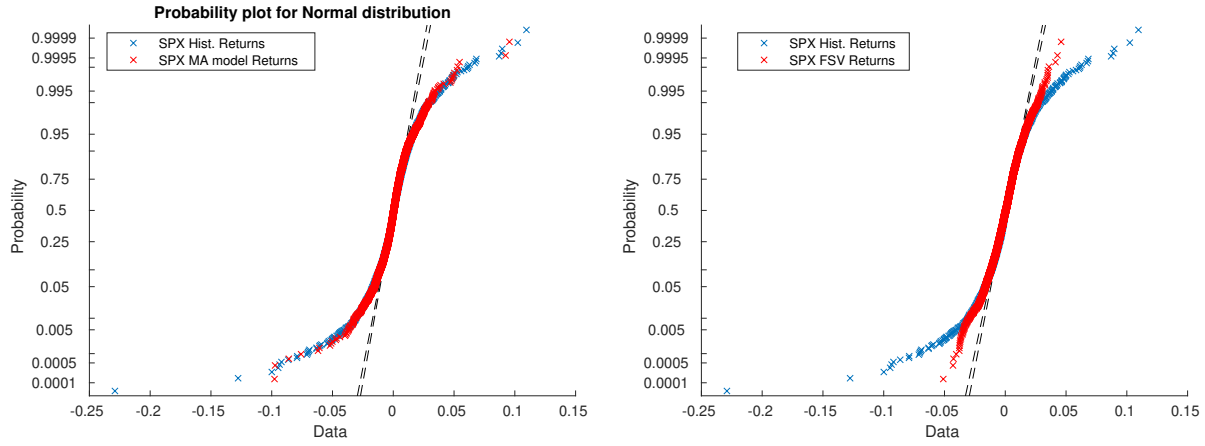


Figure 4.19: The MA model appears to reproduce the tails of the empirical distribution of returns more closely.

The question we will address in the next section is: How much does this fidelity to the

return distribution help when we suspect that the real driver of aggregated RC returns is autocorrelation? The MA model will by definition only have an exponentially decaying volatility for 50 days, whereas the FBM model may not fit the returns as well but has a persistent power-law autocorrelation that may not be overpowered by the RC mechanism.

Simulation Results

We simulated an economy with an SPX component given by Eqs. (4.19) and (4.20) and a bond component given by I00078US similar to Section 4.2.4. The results are compiled in Table 4.8. Contrary to the discussion (inspired by the results of [Gatheral et al., 2018]) in Section 4.3.2, on the scale of a decade we found that the α parameter is crucial in calibrating returns. This makes sense as $\alpha = 0.14$ gives an approximate characteristic scale of 5 years, of obvious relevance for a 10-year FIA. Recall we discussed in Section 4.2.1 the prudence (or otherwise) of using historical RC index results as a benchmark. That said, a properly calibrated FBM model can capture the standard deviation of historical FIA returns. This is despite their potentially having a higher skewness, perhaps as an artefact of no leverage correlation.

On the other hand, the MA(50) model delivers results that are essentially identical to the GBM model studied in Part 4.2. Examining Figure 4.20, we can see the extreme narrowing of the annuity return distribution for the MA(50) model annuity returns versus the more heavy-tailed FIA return distribution, even though the tails of the underlying returns better match the underlying distribution in the MA case! This would appear to validate our hypothesis: the RC mechanism interacts with the more realistic power-law decay of an FBM model in the appropriate way, whereas the MA model decays too quickly and will be unrealistically narrow in the aggregate, as in the GBM case.

4.4 Conclusion

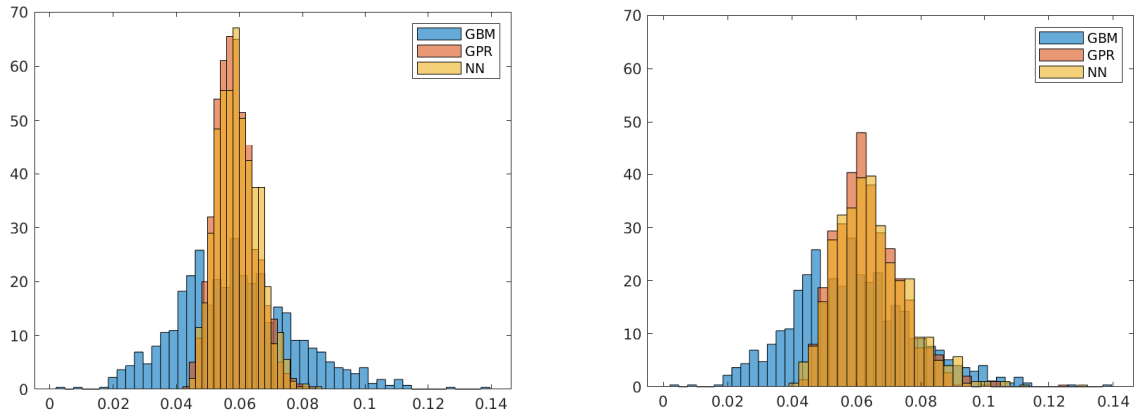
The main focus of this work was to provide a blueprint for solving the annuity decision problem practically and at scale through simulations. The number of annuity indices

		GBM	Real Returns
Direct RC index model	Mean	0.059	0.060
	SD	0.019	0.016
	Skewness	0.421	0.322

SPX model		GPR	NN
GBM	Mean	0.054	0.054
	SD	0.006	0.006
	Skewness	0.394	0.444
MA model ($\rho = -0.4$)	Mean	0.059	0.059
	SD	0.006	0.006
	Skewness	0.454	0.391
FSV ($H = 0.14$, $\alpha = 0$, $\rho = 0$)	Mean	0.071	0.072
	SD	0.025	0.029
	Skewness	2.969	2.927
FSV ($H = 0.14$, $\alpha = 0.18$, $\rho = 0$)	Mean	0.064	0.065
	SD	0.010	0.011
	Skewness	0.850	0.824

Table 4.8: Annuity return statistics using an SPX/I00078US GPR model with various stochastic volatility models for SPX

currently employed makes direct simulation problematic given the complex dependence structure between them. Furthermore, even within the comparatively simple class of RC indices, direct modelling of the index methodology remains elusive. We were, however, successful in showing that even basic ML applications can outperform more elaborate direct modelling of indices. For instance, ML models seemed to reproduce the index volatility calculations (e.g. EWMA) better than the publicly listed methodology. In fact, we do not even need the constituent underlying assets as features when using the CORE methodology. If the index components are unknown, our CORE methodology proves to be a superior approach in many cases, not simply a healthy alternative. However, outside the class of RC indices the CORE methodology tended to fail. For example, for MARC5 we had to add gold, and the Barclays index we studied appeared too complex or else had novel exposure. Extremely complex indices aside, the fact that for large classes of indices we



(a) (b)

Figure 4.20: Relevance of stochastic volatility to aggregated returns: (a) the MA model behaves very much like our original constant volatility model, whereas (b) the rough model appears more realistic.

could reproduce more reliable returns via machine learning made the exercise worthwhile as a whole. This provides a viable avenue for index forecasting where transparency of the index methodology is low.

In fact, the ML methods we used were in some sense "too accurate," reproducing the RC mechanisms with enough fidelity that a GBM-based underlying economy model was too simple. Our ML-trained index functions were strong enough to "over-react" to the random nature of GBM paths, anticipating time dependence of volatility, and when receiving none they "over-dampened" random jumps. This necessitated a more thorough investigation of the relationship between stochastic volatility (and the induced return dependence therein) and the aggregate distribution of index returns. We identified the major features of stochastic volatility (autocorrelation and leverage effects) and found a way of expressing the variance of aggregate returns with respect to them. While a full accounting of the dependence between returns would require more study, we felt that this was sufficient given the nature of RC calculations. We then explored what kinds of stochastic volatility models would work well with our ML index models and found that the "rough" autocorrelation of

volatility seems to play a crucial role in explaining the behaviour of RC indices.

This naturally leads us to the first avenue for future research: constructing a rough stochastic volatility economy that will work with our ML models to accurately model the annuity decision problem. Another natural next step is to find estimates for some of the more complex indices. Regardless of volatility models, a more informed choice of the underlying economy (e.g. adding commodities, different duration and credit quality when it comes to fixed income, and some alternatives) would provide a much richer data set to train our ML models. Absent that, more complex ML techniques may help. The novelty and complexity of some index strategies may warrant techniques such as natural language processing, adversarial learning and so on.

Aside from the technical details of modelling, there are products besides FIAs to which this work can be applied. Registered index-linked annuities, or RILAs, are very similar to FIAs but contain a loss profile (see [Moenig, 2022]). In terms of product analysis, extending our methodology to the evaluation and comparison of RILAs is straightforward but possibly even more informative. Because they offer different downside or loss profiles, while the option pricing stability arising from RC indices offers genuine benefits, they generally suffer from low transparency. Understanding how to forecast these properly could be critical for the industry.

Because RC indices have cheaper options, they also offer participation rates exceeding 100–200% that are attractive to clients and resistant to variation based on prevailing market volatility. Furthermore, the recent rise in interest rates allows for richer option budgets, upon which all index-linked annuities (including both FIAs and RILAs) depend. Therefore, it is important to understand and be able to compare indexed annuities linked to different, sometimes less familiar indices. Due to their stable and relatively low volatility, RC indices allow for more predictable option pricing, and in a rising interest rate environment they offer attractive participation rates. As it is impossible to research, deconstruct and model all of them, we believe that our methodology opens the door to another way of simulating these products while enabling the inclusion of many different structures struck on many different indices.

Chapter 5

Conclusion

Dissertation Review The first two chapters of this thesis were concerned with particular models for the losses of an insurance company, Additive Background Risk Models (ABRMs). Specifically in Chapter 2 we examined estimation in an already-studied Tweedie model and a new Stable ABRM. The same Stable model was studied in Chapter 3 in the context of risk allocation and pricing. Our Stable ABRM is the main contribution of this work: we explored its potential for modelling insurance losses as well as studying allocations, risk measures, and prices; finally we studied how to estimate in the model.

While a Stable ABRM can be estimated in a typical MLE scheme, we pointed out the impracticality of this for some insurance applications, in particular loss reserving. Large, complex, interdependent losses require a more numerically efficient means of estimation. By making use of the fact that the extreme Stable distributions (the most useful case for insurance losses) have an MGF, we were able to adapt the CGMM to create a convenient convex objective function for estimating the quantities of a Stable ABRM. By comparison with previously studied Tweedie ABRMs as a proof of concept, we are confident that a Stable ABRM can be practically used now by insurance practitioners. This opens up a whole new class of models that are useful for heavy-tailed losses and suitable for many business lines as well as catastrophe modelling.

Once the practicality of Stable ABRMs has been established through estimation, we

can concern ourselves with more classic risk-theoretic questions. Pricing risk and allocating capital under a Stable ABRM is non-trivial for the same reasons as MLE estimation: a lack of closed-form expressions such as pdfs, cdfs and so on. We studied the ES/CVar/TCE risk measure due to its ubiquity under Basel and its helpful properties. We also studied the resulting linear allocations through the CAPM-like Weighted Insurance Pricing Model. Our expression for the stable ES makes use of the Fox H-function, a special function well studied for its use in representing the Stable pdf among other distributions. We argue that the Fox H-function Mellin–Barnes integral or series representations can be used to practically compute risk capital and allocations for a Stable ABRM.

In the final chapter of this dissertation we shifted to the study of Fixed Index Annuities (FIAs): structured products offered by insurance carriers using equity exposure to hedge longevity risk. The main issue facing FIA carriers or brokers, given the large universe of products and underlying indices, is forecasting performance in a consistent way. This problem is twofold: (i) the large universe of indices makes cross-sectional forecasting difficult while (ii) the complex nature of the indices involved renders simple models such as a geometric Brownian motion (GBM) insufficient. We first used machine learning to shrink the number of needed variates down to a more manageable number of baseline indices, and then showed how more complex stochastic volatility models are required to capture the necessary intertemporal correlations in index returns that are reflected in annuity products.

Limitations and Future Work In Chapter 2 we arguably "buried the lede" by restricting ourselves to estimation in the context of Stable and Tweedie ABRMs for loss reserving. The technology presented in the CGMM can, however, be more broadly applied in other insurance and non-insurance settings. Special cases of the Tweedie such as the multivariate Gamma have appeared in many insurance settings. That being said, the CGMM objective function is very "flat" and only becomes more so as the dimensions increase; numerical methods outside the scope of the chapter need to be considered in future research.

In Chapter 3 the most obvious question is whether or not we can apply our methods (Fox H representations, linear sharing) to other risk measures. We considered for instance size bias and Esscher risk measures and they seem (at least initially) feasible.

In neither chapter did we consider the myriad of other possible applications of Stable ABRMs. Stable distributions have famously been considered in finance and practical sciences like engineering. It seems likely that fruitful applications of our work will be found if the net is cast more widely.

Many of the broad concepts and methods in Chapter 4 are not restricted to FIAs. In fact, the project could in some sense be better thought of as index modelling. Structured products linked to indices are likewise not just restricted to FIAs. However, one specific issue is there is no analytic mechanism explaining the need for rough volatility in the FIA returns. This is in part due to the optionality in FIA returns and also the difficult-to-use stochastic calculus of rough processes. The methods explored in Chapter 4 are very much near the cutting edge of current research in stochastic processes, and we hope that this index/FIA problem may serve as another laboratory for this work.

Bibliography

[BXI, 2022] (2022). Barclays Trailblazer Sectors 5 Index (factsheet). https://indices.cib.barclays/dms/Public%20marketing/BXIITBZ5_2_pager_weekly.pdf. Accessed: 24/11/2022.

[spg, 2022] (2022). S&P Risk Control 2.0 indices methodology. https://www.spglobal.com/spdji/en/documentmethodologies/methodology-sp-risk-control-2-indices.pdf. Accessed: 24/11/2022.

[Ackerberg et al., 2014] Ackerberg, D., Chen, X., Hahn, J., and Liao, Z. (2014). Asymptotic efficiency of semiparametric two-step GMM. *Review of Economic Studies*, 81(3):919–943.

[Agram and Øksendal, 2019] Agram, N. and Øksendal, B. (2019). Introduction to white noise, Hida–Malliavin calculus and applications. *arXiv preprint arXiv:1903.02936*.

[Alai et al., 2016] Alai, D. H., Landsman, Z., and Sherris, M. (2016). Multivariate Tweedie lifetimes: The impact of dependence. *Scandinavian Actuarial Journal*, 2016(8):692–712.

[Alexandrova et al., 2017] Alexandrova, M., Bohnert, A., Gatzert, N., and Russ, J. (2017). Equity-linked life insurance based on traditional products: The case of select products. *European Actuarial Journal*, 7(2):379–404.

[Alòs and Lorite, 2021] Alòs, E. and Lorite, D. G. (2021). *Malliavin calculus in finance: Theory and practice*. CRC Press.

[Artzner et al., 1999] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

[Asimit et al., 2010] Asimit, A. V., Furman, E., and Vernic, R. (2010). On a multivariate pareto distribution. *Insurance: Mathematics and Economics*, 46(2):308–316.

[Asimit et al., 2016] Asimit, A. V., Vernic, R., and Zitikis, R. (2016). Background risk models and stepwise portfolio construction. *Methodology and Computing in Applied Probability*, 18:805–827.

[Avanzi et al., 2016] Avanzi, B., Taylor, G., Vu, P. A., and Wong, B. (2016). Stochastic loss reserving with dependence: A flexible multivariate Tweedie approach. *Insurance: Mathematics and Economics*, 71:63–78.

[Babbel et al., 2010] Babbel, D. F., VanderPal, G., and Marrion, J. (2010). Real world index annuity returns. *Available at SSRN 1482023*.

[Bahnemann, 2015] Bahnemann, D. (2015). Distributions for actuaries. *CAS monograph series*, 2:1–200.

[Balbás et al., 2009] Balbás, A., Garrido, J., and Mayoral, S. (2009). Properties of distortion risk measures. *Methodology and Computing in Applied Probability*, 11(3):385.

[Baltussen et al., 2019] Baltussen, G., van Bekkum, S., and Da, Z. (2019). Indexing and stock market serial dependence around the world. *Journal of Financial Economics*, 132(1):26–48.

[Banna et al., 2019] Banna, O., Mishura, Y., Ralchenko, K., and Shklyar, S. (2019). *Fractional Brownian motion: Approximations and projections*. John Wiley & Sons.

[Barndorff-Nielsen and Shephard, 2001] Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241.

[Bee and Trapin, 2018] Bee, M. and Trapin, L. (2018). A characteristic function-based approach to approximate maximum likelihood estimation. *Communications in Statistics: Theory and Methods*, 47(13):3138–3160.

[Bernard and Boyle, 2011] Bernard, C. and Boyle, P. (2011). A natural hedge for equity-indexed annuities. *Annals of Actuarial Science*, 5(2):211–230.

[Bilodeau, 2004] Bilodeau, M. (2004). "Tail Conditional Expectations for Elliptical Distributions," Zinoviy M. Landsman and Emiliano A. Valdez, October 2003. *North American Actuarial Journal*, 8(3):118–122.

[Bochud and Challet, 2007] Bochud, T. and Challet, D. (2007). Optimal approximations of power laws with exponentials: Application to volatility models with long memory. *Quantitative Finance*, 7(6):585–589.

[Bouchaud et al., 2003] Bouchaud, J.-P., Potters, M., et al. (2003). *Theory of financial risk and derivative pricing: From statistical physics to risk management*. Cambridge University Press, 2nd edition.

[Bouchaud et al., 2000] Bouchaud, J.-P., Potters, M., and Meyer, M. (2000). Apparent multifractality in financial time series. *The European Physical Journal B: Condensed Matter and Complex Systems*, 13(3):595–599.

[Boyle and Tian, 2008] Boyle, P. and Tian, W. (2008). The design of equity-indexed annuities. *Insurance: Mathematics and Economics*, 43(3):303–315.

[Breiman, 1992] Breiman, L. (1992). *Probability*. SIAM.

[Bun et al., 2017] Bun, J., Bouchaud, J.-P., and Potters, M. (2017). Cleaning large correlation matrices: Tools from random matrix theory. *Physics Reports*, 666:1–109.

[Byczkowski et al., 1993] Byczkowski, T., Nolan, J. P., and Rajput, B. (1993). Approximation of multidimensional stable densities. *Journal of Multivariate Analysis*, 46(1):13–31.

[Carmean, 1995] Carmean, C. C. (1995). Environmental and asbestos liabilities: A growing concern for the insurance industry. *Environmental Claims Journal*, 8(1):131–140.

[Carr and Cherubini, 2022] Carr, P. and Cherubini, U. (2022). Generalized compounding and growth optimal portfolios reconciling Kelly and Samuelson. *The Journal of Derivatives*, 30(2):74–93.

[Carrasco et al., 2014] Carrasco, M., CIREQ, CIRANO, and Florens, J.-P. (2014). On the asymptotic efficiency of GMM. *Econometric Theory*, pages 372–406.

[Carrasco and Florens, 2000] Carrasco, M. and Florens, J.-P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, pages 797–834.

[Chen et al., 2013] Chen, J.-J., Zheng, B., and Tan, L. (2013). Agent-based model with asymmetric trading and herding for complex financial systems. *PlOS ONE*, 8(11):e79531.

[Cheridito et al., 2003] Cheridito, P., Kawaguchi, H., and Maejima, M. (2003). Fractional Ornstein–Uhlenbeck processes. *Electronic Journal of Probability*, 8:1–14.

[Clark and Dickson, 2021] Clark, S. P. and Dickson, M. (2021). Fixed indexed annuities: Dissecting performance expectations. *Retirement Management Journal*, 10(1):36–45.

[Comte and Renault, 1998] Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323.

[Cont, 2001] Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223.

[Daniel et al., 2005] Daniel, G., Joseph, N. L., and Brée, D. S. (2005). Stochastic volatility and the goodness-of-fit of the Heston model. *Quantitative Finance*, 5(2):199–211.

[Davis and Etheridge, 2006] Davis, M. and Etheridge, A. (2006). Louis Bachelier's *Theory of speculation*: Translated by Mark Davis and Alison Etheridge.

[De Jong, 2012] De Jong, P. (2012). Modeling dependence between loss triangles. *North American Actuarial Journal*, 16(1):74–86.

[Domingos, 2020] Domingos, P. (2020). Every model learned by gradient descent is approximately a kernel machine. *arXiv preprint arXiv:2012.00152*.

[Durrett, 2019] Durrett, R. (2019). *Probability: Theory and examples*, volume 49. Cambridge University Press.

[Eaton et al., 1971] Eaton, M. L., Morris, C., and Rubin, H. (1971). On extreme stable laws and some applications. *Journal of Applied Probability*, 8(4):794–801.

[Embrechts et al., 2013] Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: For insurance and finance*, volume 33. Springer Science & Business Media.

[Embrechts et al., 2002] Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. *Risk management: value at risk and beyond*, 1:176–223.

[England and Verrall, 2002] England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3):443–518.

[Fama and French, 2004] Fama, E. F. and French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of economic perspectives*, 18(3):25–46.

[Feller, 1957] Feller, W. (1957). *An introduction to probability theory and its applications*, volume 1. Wiley, 2nd edition.

[Feng et al., 2012] Feng, L., Li, B., Podobnik, B., Preis, T., and Stanley, H. E. (2012). Linking agent-based models and stochastic models of financial markets. *Proceedings of the National Academy of Sciences*, 109(22):8388–8393.

[Feuerverger and McDunnough, 1981] Feuerverger, A. and McDunnough, P. (1981). On some Fourier methods for inference. *Journal of the American Statistical Association*, 76(374):379–387.

[Fröhlich and Weng, 2018] Fröhlich, A. and Weng, A. (2018). Parameter uncertainty and reserve risk under Solvency II. *Insurance: Mathematics and Economics*, 81:130–141.

[Furman et al., 2018] Furman, E., Kuznetsov, A., and Zitikis, R. (2018). Weighted risk capital allocations in the presence of systematic risk. *Insurance: Mathematics and Economics*, 79:75–81.

[Furman et al., 2021] Furman, E., Kye, Y., and Su, J. (2021). Multiplicative background risk models: Setting a course for the idiosyncratic risk factors distributed phase-type. *Insurance: Mathematics and Economics*, 96:153–167.

[Furman and Landsman, 2005] Furman, E. and Landsman, Z. (2005). Risk capital decomposition for a multivariate dependent gamma portfolio. *Insurance: Mathematics and Economics*, 37(3):635–649.

[Furman and Landsman, 2010] Furman, E. and Landsman, Z. (2010). Multivariate Tweedie distributions and some related capital-at-risk analyses. *Insurance: Mathematics and Economics*, 46(2):351–361.

[Furman and Zitikis, 2008a] Furman, E. and Zitikis, R. (2008a). Weighted premium calculation principles. *Insurance: Mathematics and Economics*, 42(1):459–465.

[Furman and Zitikis, 2008b] Furman, E. and Zitikis, R. (2008b). Weighted risk capital allocations. *Insurance: Mathematics and Economics*, 43(2):263–269.

[Furman and Zitikis, 2010] Furman, E. and Zitikis, R. (2010). General stein-type covariance decompositions with applications to insurance and finance. *ASTIN Bulletin: The Journal of the IAA*, 40(1):369–375.

[Furman and Zitikis, 2017] Furman, E. and Zitikis, R. (2017). Beyond the pearson correlation: Heavy-tailed risks, weighted gini correlations, and a gini-type weighted insurance pricing model. *ASTIN Bulletin: The Journal of the IAA*, 47(3):919–942.

[Gaillardetz and Lakhmiri, 2011] Gaillardetz, P. and Lakhmiri, J. Y. (2011). A new premium principle for equity-indexed annuities. *Journal of Risk and Insurance*, 78(1):245–265.

[Gaillardetz et al., 2012] Gaillardetz, P., Li, H. Y., and MacKay, A. (2012). Equity-linked products: Evaluation of the dynamic hedging errors under stochastic mortality. *European Actuarial Journal*, 2(2):243–258.

[Gatheral et al., 2018] Gatheral, J., Jaisson, T., and Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6):933–949.

[Gisler and Wüthrich, 2008] Gisler, A. and Wüthrich, M. V. (2008). Credibility for the chain ladder reserving method. *ASTIN Bulletin: The Journal of the IAA*, 38(2):565–600.

[Groetsch, 2007] Groetsch, C. W. (2007). Integral equations of the first kind, inverse problems and regularization: A crash course. In *Journal of Physics: Conference Series*, volume 73 (1). IOP Publishing.

[Guyon and Lekeufack, 2022] Guyon, J. and Lekeufack, J. (2022). Volatility is (mostly) path-dependent. *Available at SSRN: https://ssrn.com/abstract=4174589 or http://dx.doi.org/10.2139/ssrn.4174589*.

[Hänggi, 1985] Hänggi, P. (1985). Stochastic processes applied to physics. *Philadelphia: Heiden*.

[Hansen, 1982] Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.

[Heston, 1993] Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343.

[Hida, 1976] Hida, T. (1976). Analysis of Brownian functionals. In *Stochastic Systems: Modeling, Identification and Optimization, I*, pages 53–59. Springer.

[Hinton, 1990] Hinton, G. E. (1990). Connectionist learning procedures. In *Machine learning*, pages 555–610. Elsevier.

[Hochreiter, 1998] Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107–116.

[Holden et al., 2009] Holden, H., Øksendal, B., Ubøe, J., and Zhang, T. (2009). *Stochastic partial differential equations: A modeling, white noise functional approach*. Springer Science & Business Media.

[Huang et al., 2017] Huang, H. H., Zhang, S., and Zhu, W. (2017). Limited participation under ambiguity of correlation. *Journal of Financial Markets*, 32:97–143.

[Israelov and Klein, 2016] Israelov, R. and Klein, M. (2016). Risk and return of equity index collar strategies. *The Journal of Alternative Investments*, 19(1):41–54.

[Jørgensen, 1987] Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 49, (2):127–145.

[Kim and Francis, 2013] Kim, D. and Francis, J. C. (2013). *Modern portfolio theory: Foundations, analysis, and new developments*. John Wiley & Sons.

[Kleiner, 1977] Kleiner, B. (1977). Time series analysis: Forecasting and control.

[Klenke, 2013] Klenke, A. (2013). *Probability theory: A comprehensive course*. Springer Science & Business Media.

[Kotchoni, 2012] Kotchoni, R. (2012). Applications of the characteristic function-based continuum GMM in finance. *Computational Statistics & Data Analysis*, 56(11):3599–3622.

[Koutrouvelis, 1980] Koutrouvelis, I. A. (1980). Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, 75(372):918–928.

[Kuznetsov, 2013] Kuznetsov, A. (2013). On the Convergence of the Gaver–Stehfest Algorithm. *SIAM Journal on Numerical Analysis*, 51(6):2984–2998.

[Landsman and Valdez, 2003] Landsman, Z. M. and Valdez, E. A. (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7(4):55–71.

[LeBaron, 2001] LeBaron, B. (2001). Stochastic volatility as a simple generator of apparent financial power laws and long memory. *Quantitative Finance*, 1(6):621.

[Li et al., 2016] Li, M., Gençay, R., and Xue, Y. (2016). Is it Brownian or fractional Brownian motion? *Economics Letters*, 145:52–55.

[Lilley and Frean, 2005] Lilley, M. and Frean, M. (2005). Neural networks: A replacement for Gaussian processes? In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 195–202. Springer.

[Lin and Tan, 2003] Lin, X. S. and Tan, K. S. (2003). Valuation of equity-indexed annuities under stochastic interest rates. *North American Actuarial Journal*, 7(4):72–91.

[Lin et al., 2009] Lin, X. S., Tan, K. S., and Yang, H. (2009). Pricing annuity guarantees under a regime-switching model. *North American Actuarial Journal*, 13(3):316–332.

[Liu and Zeng, 2017] Liu, J. and Zeng, X. (2017). Correlation ambiguity and under-diversification. *Available at SSRN 2692692*.

[Mack, 1991] Mack, T. (1991). A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin: The Journal of the IAA*, 21(1):93–109.

[Mack, 1993] Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin: The Journal of the IAA*, 23(2):213–225.

[MacKay et al., 2017] MacKay, A., Augustyniak, M., Bernard, C., and Hardy, M. R. (2017). Risk management of policyholder behavior in equity-linked life insurance. *Journal of Risk and Insurance*, 84(2):661–690.

[Madigan and Metzner, 2003] Madigan, K. M. and Metzner, C. S. (2003). Reserving for asbestos liabilities. In *CAS Forum*, pages 173–212. Citeseer.

[Mainik and Rüschendorf, 2010] Mainik, G. and Rüschendorf, L. (2010). On optimal portfolio diversification with respect to extreme risks. *Finance and Stochastics*, 14(4):593–623.

[Mandelbrot, 1967] Mandelbrot, B. (1967). The variation of some other speculative prices. *The Journal of Business*, 40(4):393–413.

[Mandelbrot et al., 1997] Mandelbrot, B. B., Fisher, A. J., and Calvet, L. E. (1997). A multifractal model of asset returns. *Cowles Foundation discussion paper*.

[Mandelbrot and Van Ness, 1968] Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437.

[Marri and Moutanabbir, 2022] Marri, F. and Moutanabbir, K. (2022). Risk aggregation and capital allocation using a new generalized archimedean copula. *Insurance: Mathematics and Economics*, 102:75–90.

[Masoliver and Perelló, 2006] Masoliver, J. and Perelló, J. (2006). Multiple time scales and the exponential Ornstein–Uhlenbeck stochastic volatility model. *Quantitative Finance*, 6(5):423–433.

[McAlea et al., 2016] McAlea, E. M., Mullins, M., Murphy, F., Tofail, S. A., and Carroll, A. G. (2016). Engineered nanomaterials: Risk perception, regulation and insurance. *Journal of Risk Research*, 19(4):444–460.

[McNeil et al., 2005] McNeil, A. J., Frey, R., Embrechts, P., et al. (2005). *Quantitative risk management: Concepts, techniques and tools*, volume 3. Princeton University Press.

[Melnick and Tenenbein, 1982] Melnick, E. L. and Tenenbein, A. (1982). Misspecifications of the normal distribution. *The American Statistician*, 36(4):372–373.

[Merton, 1969] Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, pages 247–257.

[Merz and Wüthrich, 2009] Merz, M. and Wüthrich, M. V. (2009). Combining chain-ladder and additive loss reserving method for dependent lines of business. *Variance*, 3(2):270–291.

[Merz et al., 2013] Merz, M., Wüthrich, M. V., and Hashorva, E. (2013). Dependence modelling in multivariate claims run-off triangles. *Annals of Actuarial Science*, 7(1):3–25.

[Miles et al., 2019] Miles, J., Furman, E., and Kuznetsov, A. (2019). Risk aggregation: A general approach via the class of generalized gamma convolutions. *Variance*.

[Moenig, 2022] Moenig, T. (2022). It's RILA time: An introduction to registered index-linked annuities. *Journal of Risk and Insurance*, 89(2):339–369.

[Moenig and Xu, 2022] Moenig, T. and Xu, C. (2022). Valuing lifetime withdrawal guarantees in RILAs. *Available at SSRN 4021905*.

[Moore, 2009] Moore, K. S. (2009). Optimal surrender strategies for equity-indexed annuity investors. *Insurance: Mathematics and Economics*, 44(1):1–18.

[Moore and Young, 2003] Moore, K. S. and Young, V. R. (2003). Pricing equity-linked pure endowments via the principle of equivalent utility. *Insurance: Mathematics and Economics*, 33(3):497–516.

[NAFA, 2022] NAFA (2022). Helping consumers understand FIAs. https://nafa.com/wp-content/uploads/NAFA_EduSeries_WhyFIA.pdf. Accessed: 12/13/2022.

[NAIC, 2022] NAIC (2022). NAIC model laws. https://content.naic.org/cipr-topics/naic-model-laws. Accessed: 24/11/2022.

[Nešlehová et al., 2006] Nešlehová, J., Embrechts, P., and Chavez-Demoulin, V. (2006). Infinite mean models and the LDA for operational risk. *Journal of Operational Risk*, 1(1):3–25.

[Nielsen, 2015] Nielsen, M. A. (2015). *Neural networks and deep learning*, volume 25. Determination Press, San Francisco.

[Nikolić et al., 2018] Nikolić, B., Toland, T., and Baboolal, D. (2018). Accumulation value of fixed annuities (MYGA & FIA): Understanding yields by product design [white paper]. *Cannex Financial Exchanges Limited*.

[Nolan, 2003] Nolan, J. (2003). *Stable distributions: Models for heavy-tailed data*. Birkhauser, New York.

[Nolan, 1997] Nolan, J. P. (1997). Numerical calculation of stable densities and distribution functions. *Communications in statistics. Stochastic models*, 13(4):759–774.

[Nolan et al., 2001] Nolan, J. P., Panorska, A. K., and McCulloch, J. H. (2001). Estimation of stable spectral measures. *Mathematical and Computer Modelling*, 34(9-11):1113–1122.

[Novikov, 1965] Novikov, E. A. (1965). Functionals and the random-force method in turbulence theory. *Journal of Experimental and Theoretical Physics*, 20(5):1290–1294.

[Orlando et al., 2022] Orlando, G., Bufalo, M., and Stoop, R. (2022). Financial markets' deterministic aspects modeled by a low-dimensional equation. *Scientific Reports*, 12(1):1–13.

[Pai and Ravishanker, 2020] Pai, J. and Ravishanker, N. (2020). Livestock mortality catastrophe insurance using fatal shock process. *Insurance: Mathematics and Economics*, 90:58–65.

[Panjer, 2002] Panjer, H. H. (2002). Measurement of risk, solvency requirements and allocation of capital within financial conglomerates. *Institute of Insurance and Pension Research, University of Waterloo, research report*, pages 1–15.

[Panjer, 1998] Panjer, H. H. e. a. (1998). *Financial economics: With applications to investments, insurance, and pensions*. Actuarial Foundation: Schaumburg, IL.

[Paulson and Goin, 2021] Paulson, M. and Goin, G. (2021). Controllable factors and uncontrollable risks in a retirement spending plan: Using a registered indexed linked annuity to mitigate risk exposure. *Retirement Management Journal*, 10(1):59–66.

[Perelló et al., 2004] Perelló, J., Masoliver, J., and Bouchaud, J.-P. (2004). Multiple time scales in volatility and leverage correlations: A stochastic volatility model. *Applied Mathematical Finance*, 11(1):27–50.

[Perelló et al., 2008] Perelló, J., Sircar, R., and Masoliver, J. (2008). Option pricing under stochastic volatility: The exponential Ornstein–Uhlenbeck model. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(06):P06010.

[Perry, 1982] Perry, P. R. (1982). The time–variance relationship of security returns: Implications for the return-generating stochastic process. *The Journal of Finance*, 37(3):857–870.

[Peters, 2019] Peters, O. (2019). The ergodicity problem in economics. *Nature Physics*, 15(12):1216–1221.

[Pitman and Pitman, 2016] Pitman, E. and Pitman, J. (2016). A direct approach to the stable distributions. *Advances in Applied Probability*, 48(A):261–282.

[Que and Belkin, 2016] Que, Q. and Belkin, M. (2016). Back to the future: Radial basis function networks revisited. In *Artificial intelligence and statistics*, pages 1375–1383. PMLR.

[Rachev et al., 2011] Rachev, S. T., Kim, Y. S., Bianchi, M. L., and Fabozzi, F. J. (2011). *Financial models with Lévy processes and volatility clustering*, volume 187. John Wiley & Sons.

[Rachev et al., 2005] Rachev, S. T., Stoyanov, S. V., Biglova, A., and Fabozzi, F. J. (2005). An empirical examination of daily stock return distributions for US stocks. *Data analysis and decision support*, pages 269–281.

[Ross, 2013] Ross, S. A. (2013). The arbitrage theory of capital asset pricing. In *Handbook of the fundamentals of financial decision making: Part I*, pages 11–30. World Scientific.

[Safran and Shamir, 2017] Safran, I. and Shamir, O. (2017). Depth-width tradeoffs in approximating natural functions with neural networks. In *International conference on machine learning*, pages 2979–2987. PMLR.

[Samorodnitsky, 2017] Samorodnitsky, G. (2017). *Stable non-Gaussian random processes: stochastic models with infinite variance*. Routledge.

[Samuelson, 1973] Samuelson, P. A. (1973). Mathematics of speculative price. *Siam Review*, 15(1):1–42.

[Schmidhuber, 2015] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61:85–117.

[Schmidt, 2006] Schmidt, K. D. (2006). *Optimal and additive loss reserving for dependent lines of business*. Techn. Univ., Inst. für Mathematische Stochastik.

[Schneider, 1986] Schneider, W. (1986). Stable distributions: Fox function representation and generalization. In *Stochastic processes in classical and quantum systems*, pages 497–511. Springer.

[Schwartz and Whitcomb, 1977] Schwartz, R. A. and Whitcomb, D. K. (1977). The time–variance relationship: Evidence on autocorrelation in common stock returns. *The Journal of Finance*, 32(1):41–55.

[scikit-learn Developers, a] scikit-learn Developers. Gaussian process regression (GPR). https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.GaussianProcessRegressor.html. Accessed: 24/11/2022.

[scikit-learn Developers, b] scikit-learn Developers. Multi-layer Perceptron regressor. https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html. Accessed: 24/11/2022.

[Scott, 2013] Scott, M. (2013). Applied stochastic processes. *Lecture Notes*.

[Seal, 1980] Seal, H. L. (1980). Survival probabilities based on Pareto claim distributions. *ASTIN Bulletin: The Journal of the IAA*, 11(1):61–71.

[Semenikhine et al., 2018] Semenikhine, V., Furman, E., and Su, J. (2018). On a multiplicative multivariate gamma distribution with applications in insurance. *Risks*, 6(3):79.

[Springer, 1979] Springer, M. D. (1979). *The algebra of random variables*. Wiley.

[Srokowski and Kamińska, 2004] Srokowski, T. and Kamińska, A. (2004). Stochastic equation for a jumping process with long-time correlations. *Physical Review E*, 70(5):051102.

[Stein and Stein, 1991] Stein, E. M. and Stein, J. C. (1991). Stock price distributions with stochastic volatility: An analytic approach. *The Review of Financial Studies*, 4(4):727–752.

[Stoyanov et al., 2006] Stoyanov, S. V., Samorodnitsky, G., Rachev, S., and Ortobelli Lozza, S. (2006). Computing the portfolio conditional value-at-risk in the alpha-stable case. *Probability and Mathematical Statistics*, 26(1):1–22.

[Su and Furman, 2017a] Su, J. and Furman, E. (2017a). Multiple risk factor dependence structures: Copulas and related properties. *Insurance: Mathematics and Economics*, 74:109–121.

[Su and Furman, 2017b] Su, J. and Furman, E. (2017b). Multiple risk factor dependence structures: Distributional properties. *Insurance: Mathematics and Economics*, 76:56–68.

[Sun and Wu, 2009] Sun, H. and Wu, Q. (2009). A note on application of integral operator in learning theory. *Applied and Computational Harmonic Analysis*, 26(3):416–421.

[Taylor, 2009] Taylor, G. (2009). Chain ladder for Tweedie distributed claims data. *Variance*, 3(1):96–104.

[Tegnér and Poulsen, 2018] Tegnér, M. and Poulsen, R. (2018). Volatility is log-normal—but not for the reason you think. *Risks*, 6(2):46.

[Tucker et al., 1968] Tucker, H. G. et al. (1968). Convolutions of distributions attracted to stable laws. *The Annals of Mathematical Statistics*, 39(5):1381–1390.

[Uchaikin and Zolotarev, 2011] Uchaikin, V. V. and Zolotarev, V. M. (2011). *Chance and stability: Stable distributions and their applications*. Walter de Gruyter.

[VanderPal, 2008] VanderPal, G. (2008). Equity index annuities. *Journal of Personal Finance*, 7(2).

[VanderPal, 2011] VanderPal, G. (2011). Real-world index annuity returns. *Journal of Financial Planning*, 24(3):50.

[Venter, 2004] Venter, G. G. (2004). Capital allocation survey with commentary. *North American Actuarial Journal*, 8(2):96–107.

[Viswanathan et al., 2003] Viswanathan, G., Fulco, U., Lyra, M., and Serva, M. (2003). The origin of fat-tailed distributions in financial time series. *Physica A: Statistical Mechanics and its Applications*, 329(1–2):273–280.

[Vitali et al., 2019] Vitali, S., Budimir, I., Runfola, C., and Castellani, G. (2019). The role of the central limit theorem in the heterogeneous ensemble of Brownian particles approach. *Mathematics*, 7(12):1145.

[Wang and Uhlenbeck, 1945] Wang, M. C. and Uhlenbeck, G. E. (1945). On the theory of the Brownian motion II. *Reviews of Modern Physics*, 17(2-3):323.

[Watson, 2022] Watson, J. (2022). When indices are cut: What withdrawals teach about risk-control index design. https://blogs.cfainstitute.org/investor/2021/11/15/when-indices-are-taken-away-what-withdrawals-teach-about-risk-control-index-design/. Accessed: 10/11/2022.

[Williams and Rasmussen, 2006] Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. MIT Press: Cambridge, MA.

[Wu et al., 2012] Wu, Y., Wang, H., Zhang, B., and Du, K.-L. (2012). Using radial basis function networks for function approximation and classification. *International Scholarly Research Notices*, 2012.

[Wüthrich, 2003] Wüthrich, M. V. (2003). Claims reserving using Tweedie's compound Poisson model. *ASTIN Bulletin*, 33(2):331–346.

[Wüthrich and Merz, 2008] Wüthrich, M. V. and Merz, M. (2008). *Stochastic claims reserving methods in insurance*, volume 435. John Wiley & Sons.

[Wüthrich and Merz, 2022] Wüthrich, M. V. and Merz, M. (2022). *Statistical foundations of actuarial learning and its applications*. Springer Nature.

[Xiao et al., 2021] Xiao, Y., Tang, Y., and Lee, C. F. (2021). Impacts of time aggregation on Beta value and R^2 estimations under additive and multiplicative assumptions: Theoretical results and empirical evidence. In *Handbook Of Financial Econometrics, Mathematics, Statistics, And Machine Learning*, pages 3947–3984. World Scientific.

[Zeng et al., 2012] Zeng, C., Chen, Y., and Yang, Q. (2012). The fBm-driven Ornstein–Uhlenbeck process: Probability density function and anomalous diffusion. *Fractional Calculus and Applied Analysis*, 15(3):479–492.

[Zhang et al., 2022] Zhang, S.-Q., Wang, F., and Fan, F.-L. (2022). Neural network Gaussian processes by increasing depth. *IEEE Transactions on Neural Networks and Learning Systems*.

[Zhang and Dukic, 2013] Zhang, Y. and Dukic, V. (2013). Predicting multivariate insurance loss payments under the Bayesian copula framework. *Journal of Risk and Insurance*, 80(4):891–919.

[Zhou et al., 2018] Zhou, M., Dhaene, J., and Yao, J. (2018). An approximation method for risk aggregations and capital allocation rules based on additive risk factor models. *Insurance: Mathematics and Economics*, 79:92–100.

[Zolotarev, 1986] Zolotarev, V. M. (1986). *One-dimensional stable distributions*, volume 65. American Mathematical Society.

APPENDICES

Appendix A

Background on Distributions

A.1 Dispersion Models

Let $Y \sim \mathcal{N}(\mu, \sigma)$ and consider the Normal distribution function:

$$f_Y(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} \tag{A.1}$$

Notice that $(y-\mu)^2$ is a typical notion of distance. The Euclidian metric is a natural notion of distance over $\mathbb{R}$ with which to describe errors; but how to extend it to $\mathbb{R}^+$, $\mathbb{Z}$, $[0,\infty)$, $\mathbb{S}$ and so on? That is, we would like to make substitutions of form $(x-\mu)^2 \to d(y;\mu)$ and $\frac{1}{\sqrt{2\pi\sigma^2}} \to a(y;\sigma^2)$ that define new distributions such as

$$f_Y(y|\mu,\sigma) = a(y;\sigma^2) \exp\left\{-\frac{1}{2\sigma^2} d(y;\mu)\right\}. \tag{A.2}$$

Bent Jørgensen used just this approach ([Jørgensen, 1987] and references therein) to create *exponential dispersion models* (EDMs). Naturally, $d(y;\mu)$ will be proportional to the log-likelihood of the "μ" parameter (whatever that may represent). In this way, the theory of generalized linear models can be easily extended to non-Normal, non-Euclidean settings. It is not obvious, however, that for each d there will be an appropriate normalization a to

produce a distribution. Put another way, given $d(y;\mu)$, can we find an $a(y;\sigma^2)$ that solves the following integral equation of the first kind:

$$1 = \int a(y;\sigma^2) \exp\left\{-\frac{1}{2\sigma^2} d(y;\mu)\right\} dy. \tag{A.3}$$

In general, determining the existence and uniqueness of such a solution is hard and replete with technical issues (see Section 2.4.4). One strategy is to start with a solution and "invert" this to find the distribution. Consider a distribution $a(y;\lambda)$ with known cumulant function[1] denoted by $\kappa(\theta)$ and defined as the logarithm of the MGF:

$$\lambda\kappa(\theta) = \log\left\{\int e^{\theta z} a(y;\lambda) dy\right\}. \tag{A.4}$$

We can easily construct a distribution called the *additive exponential dispersion model*, given by $Z \sim ED^*(\lambda,\theta))$:

$$p_Z(z|\theta,\lambda) = \exp\left\{z\theta - \lambda\kappa(\theta)\right\} a(z;\lambda) \tag{A.5}$$

Clearly this solves (A.3), since

$$\int \exp\left\{z\theta - \lambda\kappa(\theta)\right\} a(z;\lambda) dz = e^{-\lambda\kappa(\theta)} \int e^{z\theta} a(z;\lambda) dz = 1.$$

The transformation $(Y,\mu,\sigma^2) = (\frac{Z}{\lambda}, \kappa'(\theta), \frac{1}{\lambda})$ gives the *reproductive* exponential distribution model. Note that

$$M_Z(\tau) = \exp\left\{\lambda[\kappa(\theta+\tau) - \kappa(\theta)]\right\} \tag{A.6}$$

and

$$M_Y(\tau) = \exp\left\{\lambda[\kappa(\theta+\tau/\lambda) - \kappa(\theta)]\right\}, \tag{A.7}$$

implying that

$$\mathbb{E}[Y] = \mu = \kappa'(\theta) \equiv g(\theta) \text{ and } \mathbb{V}\mathrm{ar}[Y] = g'(\theta).$$

[1] Known as the partition function in physics and statistical mechanics.

The reproductive form of the pdf (given by $Y \sim ED(\mu, \sigma^2)$) is then

$$f_Y(y|\mu, \sigma^2) = \exp\left\{-\frac{1}{2\sigma^2}\left(-2\ell(y, \mu)\right)\right\} \tilde{a}(y; \sigma^2) \tag{A.8}$$

where $-\ell(y, \mu) = yg^{-1}(\mu) - \kappa(g^{-1}(\mu))$ is the negative log-likelihood[2] and $\tilde{a}(y; \sigma^2) = a(y\lambda; 1/\lambda)$. Having come full circle, we can regard -2ℓ as $d(y, \mu)$. Thus, maximizing likelihood is equivalent to minimizing the distance of some residual or error according to a metric determined by the partition function of the distribution – a highly elegant and useful construction! Note particularly that as long as we have a distribution and corresponding MGF we can construct a model of the kind in (A.5). Indeed, this is a very rich family of distributions found across the statistical sciences.

A.2 Tweedie Models

There is an interesting subclass of exponential dispersion models with useful properties and many well-known special cases. We begin by defining the unit variance function $V(\mu)$:

$$\mathbb{V}\mathbb{a}\mathbb{r}[Y] = \sigma^2 V(\mu). \tag{A.9}$$

It can be shown that $V(\mu) = g'(g^{-1}(\mu))$, uniquely determines the distribution. The *Tweedie Exponential Dispersion models* are characterized by the unit variance function $V(\mu) = \mu^p$, or equivalently as being the only EDMs closed under scale transforms. That is, if

$$cED(\mu, \sigma) = ED(c\mu, c^{2-p}\sigma^2)$$

then the EDM is Tweedie (denoted $Y \sim Tw_p(\mu, \sigma^2)$). Notably, Tweedie models are also infinitely divisible (a concept discussed in the next appendix).

The parameter p in the unit variance is called the Tweedie power parameter. Solving

[2] Assuming the range of τ is the same as the domain of p; otherwise we need to add terms.

the ODE defining the Tweedie unit variance allows us to find the $\kappa(\theta)$ that generates the Tweedie models. Let $\alpha = (p-2)/(p-1)$; then

$$\kappa_p(\theta) = \begin{cases} \frac{\alpha-1}{\alpha}\left(\frac{\theta}{\alpha-1}\right)^{\alpha}, & \text{for } p \neq 1, 2 \\ -\log(-\theta), & \text{for } p = 2 \\ e^{\theta}, & \text{for } p = 1. \end{cases} \tag{A.10}$$

Comparing this characterization to (A.6) and (A.7), we can easily recover a few common distributions:

- $p = 0$ is the Normal distribution,
- $p = 1$ is the Poisson distribution,
- $1 < p < 2$ yields the Compound Poisson–Gamma distribution, and
- $p = 2$ is the Gamma distribution.

There are also some more exotic distributions: Inverse Gaussian ($p = 3$) and some "extreme stable" distributions ($p > 3$ or $p < 0$). It should be noted however that the latter distributions are only *generated* by Stable distributions and will only truly be Stable if $\theta = 0$.

A.3 Sums of Independent Variables and Stable Variables

> It can be said without exaggerating that contemporary probability theory has acquired [its independence, authority and methods] in the course of the solution of the problem of approximating distributions of sums of independent random variables
>
> V. M. Zolotarev

First appearing in Paul Lévy's 1925 *Calcul des probabilités*, Stable variables were studied by such renowned figures as Kolmogorov and Feller. Vladimir M. Zolotarev, a student of Kolmogorov, built much of his career on the probability theory of sums and other asymptotic schemes of random variables (see [Zolotarev, 1986] and [Uchaikin and Zolotarev, 2011]). There are several challenges in working with stable distributions, not least the nonexistence of "nice" closed-form distribution functions. This is likely at the root of their historical unpopularity compared to other, similar distributions. These challenges also carry over to the multivariate class of Stable distributions, introduced in Section 3.2.

In this section we will outline the key results on Stable distributions. We will start by examining some results regarding sums of random variables generally, before discussing the GCLT. Much of this material can similarly be found in [Breiman, 1992], [Durrett, 2019] or [Feller, 1957].

Infinite Divisibility

Definition A.3.1 (Infinite Divisibility)**.** *Consider* $j = 1, ..., n$ *and* $X_{n,j}$ *i.i.d. random variables and sums of the form*

$$Z_n = X_{n,1} + \cdots + X_{n,n} \tag{A.11}$$

The weak limit of $Z_n \to Z$ is said to be an infinitely divisible random variable.

Denote the CDF of Z_n by F_{Z_n}. Remarkably there is a fairly straightforward result characterizing (weak) limits of (A.11), the proof of which can be found in any standard probability theory text such as [Klenke, 2013] or the ones previously mentioned.

Theorem A.3.1 (Lévy–Khintchine formula)**.** *G is a weak limit of F_{Z_n} as $n \to \infty$ iff its CF is of the following form:*

$$\int e^{itz} dG(z) = \exp\left\{ita - bt^2 + \int_{x\neq 0} (e^{itx} - 1 - it\sin(x)) dM(x)\right\} \tag{A.12}$$

where $a, b \geq 0$ and M, called the Lévy measure[3]*, satisfies the following:*

1. $M(\{0\}) = 0$
2. $\int_{0<x<1} x^2 dM(x) < \infty$

Any random variable with characteristic function of the form (A.12) is said to be *infinitely divisible*, as it can be expressed as an infinite sum of other random variables. Examples include the Student-t, Gamma and compound Poisson distributions. Let us add another example with this property in mind: the Stable random variables.

Univariate Stable Distribution

Definition A.3.2 (Stable random variable, first definition)**.** *A random variable X is said to have a Stable distribution if for $n \geq 2$, $\exists c_n \in \mathbb{R}^+, d_n \in \mathbb{R}$ such that*

$$X_1 + ... + X_n \overset{d}{=} c_n X + d_n \tag{A.13}$$

[3]There are some equivalent ways of characterizing the integrand w.r.t. the Lévy measure; we chose the one from [Zolotarev, 1986] for ease of reference.

where the X_i are independent copies of X.

Such a distribution will be intimately connected to the question of sums of i.i.d. random variables. Consider the following standardized sum of random variables Y_i:

$$\tilde{Z}_n = \frac{Y_1 + \cdots + Y_n}{p_n} - q_n. \tag{A.14}$$

A random variable X is the weak limit of $\tilde{Z}_n$ if and only if it is Stable. To see this, assume that the Y_i's are stable; clearly the limit will also be Stable by (A.13). Conversely, assume $\tilde{Z}_n$ converges to X. Let $S_{1,n} = X_1 + \cdots + X_n$, $S_{2,n} = X_{n+1} + \cdots + X_{2n}$ and so on. Consider the following rearrangement:

$$\tilde{Z}_{n,k} = \frac{S_{1,n} + \cdots + S_{k,n}}{p_{n,k}} - q_{n,k}. \tag{A.15}$$

We have that

$$\frac{p_{n,k}}{p_n}\tilde{Z}_{n,k} + \left(\frac{p_{n,k}}{p_n}q_{n,k} - kq_n\right) = \tilde{Z}_n + \cdots + \tilde{Z}_n \tag{A.16}$$

where the right-hand side is a sum of k $\tilde{Z}_n$'s. Taking limits on both sides, we see that if the weak limit exists, it satisfies definition A.3.2. We can thus identify the weak limits of standardized sums of i.i.d. random variables with Stable random variables:

Definition A.3.3 (Stable random variable, second definition)**.** *A random variable X is said to have a Stable distribution if for some i.i.d. Y_i's and*

$$\tilde{Z}_n = \frac{Y_1 + \cdots + Y_n}{p_n} - q_n. \tag{A.17}$$

X has a distribution function in the weak limit of $F_{\tilde{Z}_n}$ as $n \to \infty$.

Clearly any Stable variable is infinitely divisible. The converse is not true: there are infinitely divisible distributions that are not Stable. Comparing (A.17) and (A.11) we can see that infinite divisibility is weaker, as the summands may depend on n. For instance, Gamma random variables are infinitely divisible but not Stable, since a sum of n Gamma's with shape a/n will converge to another Gamma with shape a. That is, as n grows, the summands, while i.i.d., are not the same.

While both definitions A.3.2 and A.3.3 are fairly easy to interpret, we desire a way of working analytically with Stable variables. That is, we would like to identify the distribution and find a density if it exists. Unfortunately for most cases we can only find a closed-form expression for the CF, and even then there is a great deal of machinery involved.

Following [Breiman, 1992], one can take the characteristic function of both sides of A.17:

$$\sum_{i=1}^{n} \ln(\phi_{X_i}) = itd_n + \ln(\phi_{c_n X}). \tag{A.18}$$

Substituting (A.12) into (A.18) we get a functional equation. Solving it in the non-Normal case we get

$$M(x) = \begin{cases} -k_1 x^{-\alpha} \quad , & x > 0 \\ k_2(-x)^{-\alpha}, & x < 0. \end{cases} \tag{A.19}$$

Evaluating (A.12) yields an expression for the characteristic function. For $\alpha \neq 1$ [4],

$$\int_{x\neq 0} (e^{itx} - 1 - it\sin(x))dH(x) = -|t|^{\alpha}(1 - i\beta\tan(\alpha\pi/2)\text{sign}(t)) \tag{A.20}$$

[4]Unless otherwise stated we will exclusively work in the $\alpha \neq 1$ case. The $\alpha = 1$ case corresponds to the Cauchy distribution. Dealing with the Cauchy distribution directly is easy enough, whereas we need to deal with a whole new set of considerations when deriving it from (A.12).

where $\beta = (k_1 - k_2)/(k_1 + k_2)$. Including a scale and location parameter (and modifying c_n and d_n accordingly), a Stable random variable X has CF

$$\phi_X(t) = e^{-|t|^\alpha \xi_t + it\mu} \text{ with } \xi_t = \sigma^\alpha(1 - ia\beta \operatorname{sgn}(t))$$

or if we include the $\alpha = 1$ case,

$$\phi_X(t) = \begin{cases} \exp(-|\sigma t|^\alpha(1 - i\beta(\operatorname{sign} t)a) + it\mu), & \alpha \neq 1 \\ \exp(-|\sigma t|(1 + i\beta\frac{2}{\pi}(\operatorname{sign} t)\ln|t|) + it\mu), & \alpha = 1 \end{cases} \tag{A.21}$$

denoted $X \sim S_\alpha(\mu, \sigma, \beta)$ with $a = \tan(\frac{\pi\alpha}{2})$ when $\alpha \neq 1$. The μ and σ are the location and scale parameters, and are equal and proportional to the mean and variance respectively when they exist; $\alpha \in (0, 2]$ is the tail parameter. For $\alpha = 2$ we have a Normal distribution, for $\alpha < 2$ a Pareto-tailed distribution with exponent α. The value β is a skewness parameter and if $\alpha < 1$ and $\beta = \pm 1$ the support of X is either $[\mu, \infty)$ or $(-\infty, \mu]$.

The GCLT (due to [Uchaikin and Zolotarev, 2011]) then follows very naturally:

Theorem A.3.2. *(The Generalized Central Limit Theorem) Consider the sequence of centred and normalized sums of i.i.d. random variables Y_i with Pareto tails such that*

$$1 - F_{Y_i}(y) \sim k_1 y^{-\alpha} \text{ and } F_{Y_i}(y) \sim k_2|y|^{-\alpha}.$$

Define

$$\tilde{Z}_n = \frac{Y_1 + ... + Y_n}{p_n} - q_n$$

and for $\alpha \neq 1, 2$[5] set

$$p_n^\alpha = \frac{2\Gamma(\alpha)\sin(\alpha\pi/2)}{\pi(k_1+k_2)}n \text{ and } q_n = \mathbb{E}[Y_i] \text{ (if it exists, zero otherwise).}$$

[5] In the Normal case $B_n = \sqrt{n}$. See [Uchaikin and Zolotarev, 2011] for the Cauchy case.

Then $f_{Z_n} \to f_S$ weakly where f_S is a standardized Stable distribution. Thus,

$$Z_n \xrightarrow{dist.} S_\alpha(1, \beta, 0)$$

where $\beta = (k_1 - k_2)/(k_1 + k_2)$ once again.

Properties Now that we have a characteristic function for a general univariate Stable distribution, we note a few properties before moving on to the multivariate case.

- Given $X_1 \sim S_\alpha(\sigma_1, \beta_1, \mu_1)$ and $X_2 \sim S_\alpha(\sigma_2, \beta_2, \mu_2)$, we have $X_1 + X_2 \sim S_\alpha(\sigma, \beta, \mu)$ where

$$\begin{aligned} \mu &= \mu_1 + \mu_2, \\ \sigma &= (\sigma_1^\alpha + \sigma_2^\alpha)^{\frac{1}{\alpha}}, \\ \beta &= \frac{\beta_1 \sigma_1^\alpha + \beta_2 \sigma_2^\alpha}{(\sigma_1^\alpha + \sigma_2^\alpha)}. \end{aligned} \tag{A.22}$$

- As noted before, the tails of Stable distributions are asymptotically Pareto laws with index α. So first and second moments only exist for $\alpha > 1$ and $\alpha = 2$ respectively. And if $\alpha < 1$ and $\beta = \pm 1$, the support is either $[\mu, \infty)$ or $(-\infty, \mu]$.

- Stable pdfs do not generally exist in closed form. There are, however, three notable exceptions[6]:

  - Normal $(\alpha = 2)$
  - Cauchy (t with d.o.f. $= 1$) $(\alpha = 1, \beta = 0)$
  - Lévy $(\alpha = \frac{1}{2}, \beta = 1)$

[6] Another approach to the Stable CF due to [Pitman and Pitman, 2016] suggests that since these are special cases, a natural choice of solution to (A.18) is $\phi_X(t) = e^{-|t|^\alpha \xi_t}$. With some work involving the theory of slowly varying functions and the properties of the CF we can derive the correct ξ_t and an identical expression to (A.21).

- While closed-form pdfs do not exist for general α, the class of special functions known as the Fox H-functions can be used to represent Stable pdfs [Schneider, 1986]. Suppose $X \sim S_\alpha(\sigma, \beta, 0)$. For $x \geq 0$, $r = \sqrt{1 + (\alpha\beta)^2}$ and $\gamma = \frac{1}{2} - \frac{\tan^{-1}(ab)}{\alpha\pi}$ we have

$$f_X(x) = \frac{1}{\alpha\pi\sigma r^{\frac{1}{\alpha}}} H_{2,2}^{1,1}\left[\frac{x}{\sigma r^{\frac{1}{\alpha}}}\middle|\begin{matrix}(1-\frac{1}{\alpha},\frac{1}{\alpha}) & (1-\gamma,\gamma)\\ (0,1) & (1-\gamma,\gamma)\end{matrix}\right].$$

  Note that while H-functions are only defined for positive x, we have that

$$\begin{aligned} f_X(-x|\alpha,\sigma,\beta,\mu) &= f_X(x|\alpha,\sigma,-\beta,-\mu),\\ f_X(x|\alpha,\sigma,\beta,\mu) &= \frac{1}{\sigma} f_X\left(\frac{x-\mu}{\sigma}\middle|\alpha,1,\beta,0\right). \end{aligned}$$

  H-functions are defined through a Mellin–Barnes integral; we include more details in Appendix B.1. Additionally see Section 3.6 for details on computing the values of H-functions.

Appendix B

Fox H-Functions

B.1 The Fox H-function

Here we give a brief introduction to H-functions. For more details, refer to [Springer, 1979].

Definition B.1.1 (The Fox H-function)**.** *For $0 \leq m \leq q$, $0 \leq n \leq p$ and $A_j, B_j > 0$,*

$$
\begin{aligned}
H_{p,q}^{m,n}\left[z \left| \begin{matrix} (a_1, A_1) & (a_2, A_2) & \dots & (a_p, A_p) \\ (b_1, B_1) & (b_2, B_2) & \dots & (b_q, B_q) \end{matrix} \right. \right] \\
&= \mathcal{M}_z^{-1} \left\{ \frac{(\prod_{j=1}^{m} \Gamma(b_j + B_j s))(\prod_{j=1}^{n} \Gamma(1 - a_j - A_j s))}{(\prod_{j=m+1}^{q} \Gamma(1 - b_j - B_j s))(\prod_{j=n+1}^{p} \Gamma(a_j + A_j s))} \right\} \\
&= \frac{1}{2\pi i} \int_C \frac{(\prod_{j=1}^{m} \Gamma(b_j + B_j s))(\prod_{j=1}^{n} \Gamma(1 - a_j - A_j s))}{(\prod_{j=m+1}^{q} \Gamma(1 - b_j - B_j s))(\prod_{j=n+1}^{p} \Gamma(a_j + A_j s))} z^{-s}\, ds
\end{aligned}
$$

where C separates the poles of products of Gamma functions in the numerator such that it lies respectively to the right and left of

$$
z = \frac{b_j + k}{B_j} \qquad\qquad z = \frac{a_j - 1 - k}{A_j}
$$

for all j and $k = 1, 2, 3....$

For $z \in \mathbb{R}$ the function is only defined for $z > 0$.

Important Properties:

1. For $c > 0$,

$$H_{p,q}^{m,n}\left[z^c\left|\begin{matrix}(a_1,A_1) & \dots & (a_p,A_p)\\ (b_1,B_1) & \dots & (b_q,B_q)\end{matrix}\right.\right] = \frac{1}{c}H_{p,q}^{m,n}\left[z\left|\begin{matrix}(a_1,\frac{A_1}{c}) & \dots & (a_p,\frac{A_p}{c})\\ \\ (b_1,\frac{B_1}{c}) & \dots & (b_q,\frac{B_q}{c})\end{matrix}\right.\right].$$

   For $c < 0$, use the fact that

$$H_{p,q}^{m,n}\left[\frac{1}{z}\left|\begin{matrix}(a_1,A_1) & \dots & (a_p,A_p)\\ (b_1,B_1) & \dots & (b_q,B_q)\end{matrix}\right.\right] = H_{q,p}^{n,m}\left[z\left|\begin{matrix}(1-b_1,B_1) & \dots & (1-b_q,B_q)\\ (1-a_1,A_1) & \dots & (1-a_p,A_p)\end{matrix}\right.\right].$$

2. For $d \in \mathbb{R}$,

$$z^d H_{p,q}^{m,n}\left[z\left|\begin{matrix}(a_1,A_1) & \dots & (a_p,A_p)\\ (b_1,B_1) & \dots & (b_q,B_q)\end{matrix}\right.\right] = H_{p,q}^{m,n}\left[z\left|\begin{matrix}(a_1+dA_1,A_1) & \dots & (a_p+dA_p,A_p)\\ (b_1+dB_1,B_1) & \dots & (b_q+dB_1,B_q)\end{matrix}\right.\right].$$

3. The Laplace Transform is also an H-function:

$$\mathcal{L}\left\{H_{p,q}^{m,n}\left[cz\left|\begin{matrix}(a_1,A_1) & \dots & (a_p,A_p)\\ (b_1,B_1) & \dots & (b_q,B_q)\end{matrix}\right.\right]\right\}(r)$$
$$= \frac{1}{c}H_{p,q}^{m,n}\left[\frac{r}{c}\left|\begin{matrix}(1-b_1-B_1,B_1) & \dots & (1-b_q-B_q,B_q)\\ (1-a_1-A_1,A_1) & \dots & (1-a_p-A_p,A_p)\end{matrix}\right.\right].$$

B.2 H-Function Representation of One-Dimensional Stable Densities

Originally shown by [Schneider, 1986] (using a different parameterization of stable distributions) we rederive the Fox H-function representation of a univariate stable distribution in our notation.

Theorem B.2.1. *For* $X \sim S_\alpha(\sigma, \beta, 0)$, $x \geq 0$, $r = \sqrt{1+(a\beta)^2}$, $\phi = \tan^{-1}(a\beta)$ *and* $\gamma = \frac{1}{2} - \frac{\phi}{\alpha\pi}$,

$$f_X(x) = \frac{1}{\alpha\pi\sigma r^{\frac{1}{\alpha}}} H_{2,2}^{1,1}\left[\frac{x}{\sigma r^{\frac{1}{\alpha}}}\middle|\begin{matrix}\left(1-\frac{1}{\alpha},\frac{1}{\alpha}\right) & (1-\gamma,\gamma)\\ (0,1) & (1-\gamma,\gamma)\end{matrix}\right]$$

Proof. Let $\xi_t = 1 - i\beta(\operatorname{sign} t)a$, $\xi = 1 - i\beta a$ and $\bar{\xi} = 1 + i\beta a$. We begin by inverting the characteristic function:

$$\begin{aligned}
f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}e^{|t|^\alpha \xi_t}dt \\
&= \frac{1}{2\pi}\int_0^{\infty} e^{-ixt}e^{|t|^\alpha \xi}dt + \frac{1}{2\pi}\int_{-\infty}^{0} e^{-ixt}e^{|t|^\alpha \bar{\xi}}dt \\
&= \operatorname{Re}\left[\frac{1}{\pi}\int_0^{\infty} e^{-ixt}e^{t^\alpha \xi}dt\right] \\
&= \frac{1}{\pi}\operatorname{Re}\left[\mathcal{L}\left\{e^{-t^\alpha\xi}\right\}(ix)\right] \\
&= \frac{1}{\pi}\operatorname{Re}\left[\frac{1}{\alpha\xi^{\frac{1}{\alpha}}}H_{1,1}^{1,1}\left[\frac{ix}{\xi^{\frac{1}{\alpha}}}\middle|\begin{matrix}\left(1-\frac{1}{\alpha},\frac{1}{\alpha}\right)\\ (0,1)\end{matrix}\right]\right] \\
&\quad \text{by Lemma 3.5.1} \\
&= \frac{1}{\alpha\pi\sigma r^{\frac{1}{\alpha}}}H_{2,2}^{1,1}\left[\frac{x}{\sigma r^{\frac{1}{\alpha}}}\middle|\begin{matrix}\left(1-\frac{1}{\alpha},\frac{1}{\alpha}\right) & (1-\gamma,\gamma)\\ (0,1) & (1-\gamma,\gamma)\end{matrix}\right] \\
&\quad \text{by Lemma 3.5.2, where } \theta_1 = \frac{\phi}{\alpha}, \nu_2 = \nu_1 = -1, \theta_2 = \left(\frac{\pi}{2} - \frac{\phi}{\alpha}\right).
\end{aligned}$$

■

Appendix C

Proofs of Lemmas

C.1 Proof of Lemma 3.5.1

Proof. First note that being an exponential function, the Mellin transform of e^{-bt^α} is a Gamma function:

$$\mathcal{M}\left\{e^{-bt^\alpha}\right\}(s) = \int_0^\infty t^{s-1}e^{-bt^\alpha}dt.$$

Making the substitution $u = bt^\alpha$, we obtain $\frac{dt}{t} = \frac{du}{\alpha u}$, so that

$$\begin{aligned}\mathcal{M}\left\{e^{-bt^\alpha}\right\}(s) &= \int_0^\infty \left(\frac{u}{b}\right)^{s/\alpha} e^{-u}\frac{du}{\alpha u} \\ &= \frac{1}{\alpha b^{\frac{s}{\alpha}}}\int_0^\infty u^{s/\alpha-1}e^{-u}du \\ &= \frac{1}{\alpha b^{\frac{s}{\alpha}}}\Gamma\left(\frac{s}{\alpha}\right).\end{aligned}$$

Making use of the Mellin-Barnes Representation,

$$e^{-bt^\alpha} = \frac{1}{\alpha}\mathcal{M}^{-1}\left[\Gamma\left(\frac{s}{\alpha}\right)b^{-\frac{s}{\alpha}}\right] = \frac{1}{\alpha}H_{0,1}^{1,0}\left[b^{\frac{1}{\alpha}}t\,\middle|\,\begin{matrix}-\\(0,\frac{1}{\alpha})\end{matrix}\right].$$

Using Property 2 of Fox H-functions from Appendix B.1,

$$\begin{aligned} t^j e^{-bt^\alpha} &= \frac{t^j}{\alpha} H_{0,1}^{1,0}\left[b^{\frac{1}{\alpha}} t \,\middle|\, \begin{matrix} - \\ \left(0, \frac{1}{\alpha}\right) \end{matrix} \right] \\ &= \frac{(b^{\frac{1}{\alpha}} t)^j}{\alpha b^{\frac{j}{\alpha}}} H_{0,1}^{1,0}\left[b^{\frac{1}{\alpha}} t \,\middle|\, \begin{matrix} - \\ \left(0, \frac{1}{\alpha}\right) \end{matrix} \right] \\ &= \frac{1}{\alpha b^{\frac{j}{\alpha}}} H_{0,1}^{1,0}\left[b^{\frac{1}{\alpha}} t \,\middle|\, \begin{matrix} - \\ \left(\frac{j}{\alpha}, \frac{1}{\alpha}\right) \end{matrix} \right]. \end{aligned}$$

Finally, using the Laplace transform (Property 3 in Appendix B.1) yields

$$\mathcal{L}\left\{ \frac{1}{\alpha b^{\frac{j}{\alpha}}} H_{0,1}^{1,0}\left[b^{\frac{1}{\alpha}} t \,\middle|\, \begin{matrix} - \\ \left(\frac{j}{\alpha}, \frac{1}{\alpha}\right) \end{matrix} \right] \right\}(x) = \frac{1}{\alpha b^{\frac{j+1}{\alpha}}} H_{1,1}^{1,1}\left[\frac{x}{b^{\frac{1}{\alpha}}} \,\middle|\, \begin{matrix} \left(1 - \frac{j+1}{\alpha}, \frac{1}{\alpha}\right) \\ (0,1) \end{matrix} \right].$$

■

C.2 Proof of Lemma 3.5.2

Proof. Define

$$\begin{aligned} z_1^{\nu_1} &= r_1^{\nu} e^{i\theta_1 \nu_1} \\ z_2^{\nu_2} &= r_2^{\nu_2} e^{i\theta_2 \nu_2} \end{aligned}$$

and recall the Euler reflection formula:

$$\Gamma(1-z)\Gamma(z) = \frac{\pi}{\sin(\pi z)}.$$

We have

$$
\begin{aligned}
& z_1^{\nu_1} H_{1,1}^{1,1}\left[z_2^{\nu_2} x \,\middle|\, \begin{matrix}(a_1, A_1)\\ (b_1, B_1)\end{matrix}\right] \\
&\quad = \frac{z_1^{\nu_1}}{2\pi i}\int_C \Gamma(b_1 + B_1 s)\Gamma(1 - a_1 - A_1 s))(z_2^{\nu_2} x)^{-s}\, ds \\
&\quad = \frac{1}{2\pi i}\int_C \Gamma(b_1 + B_1 s)\Gamma(1 - a_1 - A_1 s))\left(r_1^{\nu_1} e^{i(\theta_1\nu_1 - \theta_2\nu_2 s)}\right)(r_2^{\nu_2} x)^{-s}\, ds \\
&\quad = \frac{r_1^{\nu_1}}{2\pi i}\int_C \Gamma(b_1 + B_1 s)\Gamma(1 - a_1 - A_1 s))\sin\left(\pi\left(\frac{1}{2} + \frac{\theta_1\nu_1}{\pi} - \frac{\theta_2\nu_2}{\pi}s\right)\right)(r_2^{\nu_2} x)^{-s}\, ds \\
&\qquad + i\frac{r_1^{\nu_1}}{2\pi i}\int_C \Gamma(b_1 + B_1 s)\Gamma(1 - a_1 - A_1 s))\sin\left(\pi\left(\frac{\theta_1\nu_1}{\pi} - \frac{\theta_2\nu_2}{\pi}s\right)\right)(r_2^{\nu_2} x)^{-s}\, ds
\end{aligned}
$$

and using the definition of the Fox H-function and the reflection formula,

$$
\begin{aligned}
&= \pi r_1^{\nu_1} H_{2,2}^{1,1}\left[r_2^{\nu_2} x \,\middle|\, \begin{matrix}(a_1, A_1) & \left(\frac{1}{2} - \frac{\theta_1\nu_1}{\pi}, \frac{\theta_2\nu_2}{\pi}\right)\\ (b_1, B_1) & \left(\frac{1}{2} - \frac{\theta_1\nu_1}{\pi}, \frac{\theta_2\nu_2}{\pi}\right)\end{matrix}\right] \\
&\quad + i\pi r_1^{\nu_1} H_{2,2}^{1,1}\left[r_2^{\nu_2} x \,\middle|\, \begin{matrix}(a_1, A_1) & \left(1 - \frac{\theta_1\nu_1}{\pi}, \frac{\theta_2\nu_2}{\pi}\right)\\ (b_1, B_1) & \left(1 - \frac{\theta_1\nu_1}{\pi}, \frac{\theta_2\nu_2}{\pi}\right)\end{matrix}\right]
\end{aligned}
$$

■

C.3 Proof of Modified Jordan's Lemma in Section 3.6

Proof. Consider the CW curve $\Upsilon : s = c + Re^{-i\theta}, \quad \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

$$
\begin{aligned}
&\left|\int_\Upsilon g(s)z^{-s}ds\right| \\
&= \left|\int_\Upsilon g(s)e^{-s\ln(z)}ds\right| \\
&= e^{-c\ln(z)}\left|\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} g(c+Re^{-i\theta})\exp\{-\ln(z)[R\cos(\theta) - iR\sin(\theta)]\}(-iR)e^{-i\theta}d\theta\right| \\
&\leq RM_R e^{-c\ln(z)}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{-\ln(z)R\cos(\theta)}d\theta \\
&= 2RM_R e^{-c\ln(z)}\int_{-\frac{\pi}{2}}^{0} e^{-\ln(z)R\cos(\theta)}d\theta \\
&\leq 2RM_R e^{-(c+R)\ln(z)}\int_{-\frac{\pi}{2}}^{0} e^{-2R\ln(z)\theta/\pi}d\theta \\
&\qquad \text{(since on } [-\pi/2, 0] \text{ we have } -\cos(\theta) < -1 - 2\theta/\pi) \\
&= M_R e^{-(c+R)\ln(z)}\left(\frac{\pi}{\ln(z)}\left(e^{R\ln(z)} - 1\right)\right) \\
&\leq M_R\frac{\pi}{\ln(z)}e^{-c\ln(z)}.
\end{aligned}
$$

■

Appendix D

Novikov's Theorem

D.1 Miscellaneous Proofs

Lemma D.1.1. *For normally distributed random variables X_1 and X_2, we have that*

$$\mathbb{C}\mathbb{o}\mathbb{v}[X_1^2, X_2^2] = 2(\mathbb{E}[X_1, X_2]^2 - \mathbb{E}[X_1]^2\mathbb{E}[X_2]^2)$$

Proof. Recall that the normal conditional moments are

$$\mathbb{V}\mathbb{a}\mathbb{r}(X_2|X_1) = \mathbb{V}\mathbb{a}\mathbb{r}[X_2] - \frac{\mathbb{C}\mathbb{o}\mathbb{v}[X_1, X_2]^2}{\mathbb{V}\mathbb{a}\mathbb{r}[X_1]}$$

and

$$\mathbb{E}[X_2|X_1] = \mathbb{E}[X_2] + \frac{\mathbb{C}\mathbb{o}\mathbb{v}[X_1, X_2]}{\mathbb{V}\mathbb{a}\mathbb{r}[X_1]}(X_1 - \mathbb{E}[X_1]).$$

We show first that $\mathbb{C}\mathrm{ov}[(X_1 - \mathbb{E}[X_1])^2, (X_2 - \mathbb{E}[X_2])^2] = 2\mathbb{C}\mathrm{ov}[X_1, X_2]^2$:

$$
\begin{aligned}
&\mathbb{C}\mathrm{ov}[(X_1 - \mathbb{E}[X_1])^2, (X_2 - \mathbb{E}[X_2])^2] \\
&\quad = \mathbb{E}[(X_1 - \mathbb{E}[X_1])^2 (X_2 - \mathbb{E}[X_2])^2] - \mathbb{V}\mathrm{ar}[X_1]\mathbb{V}\mathrm{ar}[X_2] \\
&\quad = \mathbb{E}[(X_1 - \mathbb{E}[X_1])^2 \mathbb{E}[(X_2 - \mathbb{E}[X_2])^2 | X_1]] - \mathbb{V}\mathrm{ar}[X_1]\mathbb{V}\mathrm{ar}[X_2] \\
&\quad = \mathbb{E}\left[(X_1 - \mathbb{E}[X_1])^2 \left(\mathbb{V}\mathrm{ar}(X_2|X_1) + \mathbb{E}[X_2|X_1]^2 - 2\mathbb{E}[X_2]\mathbb{E}[X_2|X_1] + \mathbb{E}[X_2]^2\right)\right] \\
&\qquad - \mathbb{V}\mathrm{ar}[X_1]\mathbb{V}\mathrm{ar}[X_2] \\
&\quad = -\frac{\mathbb{C}\mathrm{ov}[X_1, X_2]^2}{\mathbb{V}\mathrm{ar}[X_1]}\mathbb{E}[(X_1 - \mathbb{E}[X_1])^2] + \left(\frac{\mathbb{C}\mathrm{ov}[X_1, X_2]}{\mathbb{V}\mathrm{ar}[X_1]}\right)^2 \mathbb{E}[(X_1 - \mathbb{E}[X_1])^4] \\
&\qquad\qquad \text{(making the required substitutions and simplifying)} \\
&\quad = 2\mathbb{C}\mathrm{ov}[X_1, X_2]^2.
\end{aligned}
$$

We can now prove the lemma. Starting from

$$\mathbb{C}\mathrm{ov}[X_1^2, X_2^2] = \mathbb{E}[X_1^2 X_2^2] - \mathbb{E}[X_1^2]\mathbb{E}[X_2^2]$$

and using normal cumulants or Isserlis' theorem on the right-hand side, after some calculations we get

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}[X_1^2, X_2^2] &= \mathbb{C}\mathrm{ov}[(X_1 - \mathbb{E}[X_1])^2, (X_2 - \mathbb{E}[X_2])^2] \\
&\quad + \underbrace{2\mathbb{E}[X_1]\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])^2] + 2\mathbb{E}[X_2]\mathbb{E}[(X_1 - \mathbb{E}[X_1])^2 (X_2 - \mathbb{E}[X_2])]}_{\text{Normal CoSkewness} = 0} \\
&\quad + 4\mathbb{E}[X_1]\mathbb{E}[X_2]\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])] \\
&\quad = 2\mathbb{C}\mathrm{ov}[X_1, X_2]^2 + 4\mathbb{E}[X_1]\mathbb{E}[X_2]\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])] \\
&\quad = 2(\mathbb{E}[X_1, X_2]^2 - \mathbb{E}[X_1]^2\mathbb{E}[X_2]^2)
\end{aligned}
$$

■

Proof of Theorem 4.3.1

Proof. We compute

$$\begin{aligned}\mathbb{V}\mathbb{a}\mathbb{r}[X] &= \mathbb{V}\mathbb{a}\mathbb{r}\left[\mu T - \int_0^T \frac{\sigma_t^2}{2}dt + \int_0^T |\sigma_t| dW_t\right] \\ &= \mathbb{V}\mathbb{a}\mathbb{r}\left[\int_0^T |\sigma_t| dW_t\right] + \mathbb{V}\mathbb{a}\mathbb{r}\left[\int_0^T \frac{\sigma_t^2}{2}dt\right] + 2\mathbb{C}\mathbb{o}\mathbb{v}\left[\int_0^T \frac{\sigma_t^2}{2}dt, \int_0^T |\sigma_t| dW_t\right].\end{aligned}$$

1st term:

$$\begin{aligned}\mathbb{V}\mathbb{a}\mathbb{r}\left[\int_0^T |\sigma_t| dW_t\right] &= \mathbb{E}\left[\left(\int_0^T |\sigma_t| dW_t\right)^2\right] - \mathbb{E}\left[\left(\int_0^T |\sigma_t| dW_t\right)\right]^2 \\ &= \mathbb{E}\left[\int_0^T |\sigma_t|^2 dt\right] - 0^2 \\ &= \mathbb{E}\left[\int_0^T \sigma_t^2 dt\right].\end{aligned}$$

2nd term:

$$\begin{aligned}\mathbb{V}\mathbb{a}\mathbb{r}\left[\int_0^T \frac{\sigma_t^2}{2}dt\right] &= \mathbb{E}\left[\left(\int_0^T \frac{\sigma_t^2}{2}dt\right)^2\right] - \mathbb{E}\left[\int_0^T \frac{\sigma_t^2}{2}dt\right]^2 \\ &= \mathbb{E}\left[\int_0^T \int_0^T \frac{\sigma_t^2\sigma_u^2}{4}dtdu\right] - \left(\int_0^T \mathbb{E}\left[\frac{\sigma_t^2}{2}\right]dt\right)^2 \\ &= \frac{1}{4}\int_0^T \int_0^T \left(\mathbb{E}\left[\sigma_t^2\sigma_u^2\right] - \mathbb{E}\left[\sigma_t^2\right]\mathbb{E}\left[\sigma_u^2\right]\right)dtdu \\ &= \frac{1}{4}\int_0^T \int_0^T C(t,u)dtdu.\end{aligned}$$

where $C(t,u) = COV[\sigma_t^2, \sigma_u^2]$. If the correlation structure is stationary (i.e. $C(t,u) = C(|t-u|)$) we can simplify this further into a single integral.

3rd term: Here we employ Theorem D.2.1. If

$$|\sigma_u| = \left| \int_{-\infty}^{u} k(u, \tilde{u}) dW_{\tilde{u}} \right| \quad \text{and} \quad \sigma_t^2 = \left(\int_{-\infty}^{t} k(t, \tilde{t}) dW_{\tilde{t}} \right)^2$$

we have $h(x) = |x|$ and $g(x) = x^2$. We also account for the difference in $W^{(2)}$ and $W^{(1)}$ by introducing a ρ term:

$$\begin{aligned} \mathbb{E}\left[\left(\int_0^T \sigma_t^2 dt \right) \left(\int_0^T |\sigma_u| dW_u \right) \right] &= 2\rho\lambda \int_0^T \int_0^T k(t,u) H(t-u) \mathbb{E}\left[\sigma_t |\sigma_u|\right] dt du \\ &= 2\rho\lambda\Theta \int_0^T \int_0^T k(t,u) H(t-u) \mathbb{E}\left[\sigma_t \sigma_u\right] dt du \end{aligned}$$

where $\Theta = P(\sigma_u > 0) - P(\sigma_u \leq 0)$. It only remains to show that $\mathbb{E}\left[\sigma_t |\sigma_u|\right] = \Theta \mathbb{E}\left[\sigma_t \sigma_u\right]$. But

$$\mathbb{E}\left[\sigma_t |\sigma_u|\right] = \int_0^t k(t,t') k(s,t') H(s-t') \mathbb{E}[\operatorname{sgn}(\sigma_u)] dt' = \mathbb{E}[\operatorname{sgn}(\sigma_u)] \mathbb{E}\left[\sigma_t \sigma_u\right] = \Theta \mathbb{E}\left[\sigma_t \sigma_u\right].$$

■

D.2 Some Notes on Novikov's Theorem and White Noise Analysis

> [We] are aware of the fact that in the mathematical literature [...] the notion of a random (or stochastic) process has been defined in a much more refined way. This allows [us] for instance to determine in certain cases the probability that a random function $y(t)$ is of bounded variation, or continuous or differentiable, [...]. However, it seems to us that these investigations have not helped in the solution of problems of direct physical interest, and we will, therefore, not try to give an account of them.
>
> – M.C. Wang and G.E. Uhlenbeck ([Wang and Uhlenbeck, 1945])

Building the bridge from theory to application can be quite challenging. As a branch of mathematics, white noise analysis can be traced back to [Hida, 1976] but earlier versions were originally conceived as a means of analyzing particle motion under turbulent flow (see the work of Novikov, [Novikov, 1965]). These techniques of the so-called random force method were fortunately rediscovered for finance by J.P. Bouchaud ([Bouchaud et al., 2003]). The results presented here can be easily derived using notions of functional derivatives in the sense most often used by theoretical physicists. We provide a proof at the end of this appendix that allows us to sidestep this approach, but the result has been widely used for some time. We first offer some historical background for the interested reader.

In deriving the leverage correlation term in Theorem (4.3.1) we made use of a theorem of Novikov. In particular, we showed that

$$\mathbb{E}\left[\left(\int_0^T \sigma_t^2 dt\right)\left(\int_0^T |\sigma_u| dW_u\right)\right] = 2\rho\lambda\Delta \int_0^T \int_0^T k(t,u)H(t-u)\mathbb{E}\left[\sigma_t|\sigma_u|\right] dtdu.$$

Novikov's theorem is a common tool in the random force method of statistical physics even today ([Scott, 2013]). The method relies on the manipulation of the derivative of Brownian motion and functional calculus therein. Physicists typically write $dW_t = \xi_t dt$ where ξ_t is called *white noise*, purportedly the derivative of Brownian motion W_t. In this way, they write SDEs in the same way as deterministic differential equations:

$$dX_t = \mu dt + \sigma dW_t \to \frac{dX(t)}{dt} = \mu + \sigma\xi(t).$$

As a function of time, ξ_t changes with each realization. Therfore the solution to the above is best thought of as a functional:

$$X[\xi(s)] = \mu t + \sigma \int_0^t \xi(s) ds.$$

Physicists frequently apply typical deterministic functional calculus (a.k.a. the calculus of variations) to this style of functional. For now we will ignore all the requisite rigour for functional calculus, such as the spaces we are working over, existence results and so on. Recall the Gateaux derivative:

$$D_\phi X = \lim_{\epsilon \to 0} \frac{1}{\epsilon}(X[\xi(s) + \epsilon\phi(s)] - X[\xi(s)]).$$

Using inner product notation for operators, we can write $D_\phi X = \langle A, \phi \rangle = \int A\phi ds$. If we take the Gateaux derivative in the "direction" of the Dirac Delta $\delta(s-t)$, we obtain $D_\delta X = \langle A, \delta(s-t) \rangle = A(t)$. Physicists identify $A(t)$ with the Fréchet derivative of X and write ([Hänggi, 1985])

$$A(t) = \frac{\delta X[\xi(s)]}{\delta \xi(t)}.$$

Using this notation, Novikov's theorem takes the form

$$\mathbb{E}[F[\xi(s)]\xi(t)] = \mathbb{E}\left[\frac{\delta F[\xi(s)]}{\delta \xi(t)}\right]$$

for some functional F. In this way, our result would be expressed in the following way:

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_0^T \sigma_t^2 dt\right)\left(\int_0^T |\sigma_u| dW_u\right)\right] \\
&\quad = \mathbb{E}\left[\int_0^T \int_0^T \sigma_t^2 |\sigma_u| \xi_u dt du\right] \\
&\quad = \int_0^T \int_0^T \mathbb{E}\left[\sigma_t^2 |\sigma_u| \xi_u^{(1)}\right] dt du \\
&\quad = \rho \int_0^T \int_0^T \mathbb{E}\left[2\sigma_t |\sigma_u| \frac{\delta \sigma_t}{\delta \xi_u^{(2)}} + \sigma_t^2 \frac{\delta |\sigma_u|}{\delta \xi_u^{(2)}}\right] dt du \\
&\quad = \rho\lambda \int_0^T \int_0^T \mathbb{E}\left[2\sigma_t |\sigma_u| k(t,u) H(t-u) + \sigma_t^2 \mathrm{sgn}(\sigma_u) k(u,u) H(0)\right] dt du \\
&\quad = 2\rho\lambda \int_0^T \int_0^T k(t,u) H(t-u) \mathbb{E}\left[\sigma_t |\sigma_u|\right] dt du \qquad (\text{since } H(0) = 0) \\
&\quad = 2\rho\lambda\Delta \int_0^T \int_0^T k(t,u) H(t-u) \mathbb{E}\left[\sigma_t \sigma_u\right] dt du.
\end{aligned}$$

As is often the case in both mathematics and physics, the rigour of these techniques followed the application. It is easy enough to make the theory rigorous in the deterministic case. By picking an appropriate space of functions (continuous, compact etc.), the Gateaux derivative is a functional $D_\phi X : \phi \to \mathbb{R}$ that can be represented by $D_\phi X = \int \phi dm$ where m is some measure. As long as m is absolutely continuous with respect to the Lebesgue measure λ, the Radon–Nikodym $\frac{dm}{d\lambda}$ can serve as our analogue of $\frac{\delta X}{\delta \xi}$, the Fréchet derivative. However, even this incomplete story is significantly complicated by the introduction of stochastic functions. To begin with, it is not even clear what kind of object should serve as the white noise.

This issue lies at the heart of Hida white noise theory and the associated Hida–Malliavin (a.k.a. stochastic functional) calculus ([Agram and Øksendal, 2019] and references therein). While far outside the scope of this appendix, rich mathematical theories such as these make the above discussion rigorous. That being said, we could not find a modern reference proving the Novikov theorem using Hida–Malliavin techniques, perhaps unsurprisingly since presumably the original physics derivation can be made rigorous using this theory.

However, for the sake of completeness a proof is required if we are to make use of Theorem 4.3.1. To that end, we now provide a proof using standard techniques (Ito integrals, Stein's Lemma) familiar to most readers with a mathematical finance or actuarial background. We do not claim it is necessarily an original result, but rather a contemporary verification.

Theorem D.2.1 (Novikov)**.** *For $h, g \in C^1(\mathbb{R})$ and continuous $k(\cdot,\cdot)$, define*

$$h_u = h\left(\int_{-\infty}^{u} k(u,\tilde{u})dW_{\tilde{u}}\right), \quad g_t = g\left(\int_{-\infty}^{t} k(t,\tilde{t})dW_{\tilde{t}}\right) \quad \text{and} \quad \dot{g}_t = \dot{g}\left(\int_{-\infty}^{t} k(t,\tilde{t})dW_{\tilde{t}}\right).$$

If

$$\mathbb{E}\left[\sup_{u\in[0,T]}|h_u|\right] < \infty \quad \text{and} \quad \mathbb{E}[\sup_{t\in[0,T]}|g_t|] < \infty \tag{D.1}$$

then

$$\mathbb{E}\left[\left(\int_0^T h_u dW_u\right)\left(\int_0^T g_t dt\right)\right] = \int_0^T\int_0^T \mathbb{E}\left[h_u \dot{g}_t k(t,u)\theta(t-u)\right] dtdu. \tag{D.2}$$

Proof. Consider a partition π with

$$\Delta u_i = [u_{i-1}, u_i], \quad \Delta t_i = [t_{i-1}, t_i] \quad \text{and} \quad \Delta W_i = (W_{u_i} - W_{u_{i-1}}).$$

We have

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^T h_u dW_u\right)\left(\int_0^T g_t dt\right)\right] &= \mathbb{E}\left[\lim_{\|\pi\|\to\infty}\left(\sum_{\Delta u_i\in\pi} h_{u_{i-1}}\Delta W_i\right)\left(\sum_{\Delta t_i\in\pi} g_{t_{i-1}}\Delta t_i\right)\right] \\
&= \mathbb{E}\left[\lim_{\|\pi\|\to\infty}\sum_{\Delta u_i\in\pi}\sum_{\Delta t_i\in\pi} h_{u_{i-1}}g_{t_{i-1}}\Delta W_i\Delta t_i\right] \\
&= \lim_{\|\pi\|\to\infty}\sum_{\Delta u_i\in\pi}\sum_{\Delta t_i\in\pi}\mathbb{E}\left[h_{u_{i-1}}g_{t_{i-1}}\Delta W_i\right]\Delta t_i
\end{aligned}$$

where we can establish the last equality by dominated convergence given Eq. (D.1).

Consider a new partition $\overline{\pi}$ for the interior of h and g:

$$\begin{aligned}\mathbb{E}\left[h_{u_{i-1}}g_{t_{i-1}}\Delta W_i\right] &= \mathbb{E}\left[h\left(\int_{-\infty}^{u_{i-1}} k(u,\tilde{u})dW_{\tilde{u}}\right) g\left(\int_{-\infty}^{t_{i-1}} k(t,\tilde{t})dW_{\tilde{t}}\right)\Delta W_i\right] \\ &= \mathbb{E}\left[h_{u_{i-1}}\right]\mathbb{E}\left[\lim_{\|\overline{\pi}\|\to\infty} g\left(\sum_{\substack{\Delta t_j\in\overline{\pi}\\ j\le i}} k(t,\tilde{t}_{j-1})\Delta W_j\right)\Delta W_i\right].\end{aligned}$$

By continuity of g and Stein's Lemma,

$$\begin{aligned}&\mathbb{E}\left[\lim_{\|\overline{\pi}\|\to\infty} g\left(\sum_{\substack{\Delta t_j\in\overline{\pi}\\ j\le i}} k(t,\tilde{t}_{j-1})\Delta W_j\right)\Delta W_i\right] \\ &\quad= \lim_{\|\overline{\pi}\|\to\infty}\mathbb{E}\left[g\left(\sum_{\substack{\Delta t_j\in\overline{\pi}\\ j\le i}} k(t,\tilde{t}_{j-1})\Delta W_j\right)\Delta W_i\right] \\ &\quad= \lim_{\|\overline{\pi}\|\to\infty}\mathbb{E}\left[\dot{g}\left(\sum_{\substack{\Delta t_j\in\overline{\pi}\\ j\le i}} k(t,\tilde{t}_{j-1})\right)k(t,t_{i-1})\theta(j-i+1)\Delta u_i\right] \\ &\quad= \mathbb{E}\left[\dot{g}\left(\int_{-\infty}^{t_{i-1}} k(t,\tilde{t})dW_{\tilde{t}}\right)k(t,t_{i-1})\theta(t_{i-1}-u_{i-1})\Delta u_i\right]\end{aligned}$$

and so

$$\begin{aligned}&\mathbb{E}\left[\left(\int_0^T h_u dW_u\right)\left(\int_0^T g_t dt\right)\right] \\ &\quad= \lim_{\|\pi\|\to\infty}\sum_{\Delta u_i\in\pi}\sum_{\Delta t_i\in\pi}\mathbb{E}\left[h_{u_{i-1}}g_{t_{i-1}}\Delta W_i\right]\Delta t_i \\ &\quad= \lim_{\|\pi\|\to\infty}\sum_{\Delta u_i\in\pi}\sum_{\Delta t_i\in\pi}\mathbb{E}\left[h_{u_{i-1}}\dot{g}_{t_{i-1}}k(t,t_{i-1})\theta(t_{i-1}-u_{i-1})\right]\Delta t_i\Delta u_i \\ &\quad= \int_0^T\int_0^T \mathbb{E}\left[h_u\dot{g}_t k(t,u)\theta(t-u)\right]dtdu\end{aligned}$$

which is Eq. (D.2) as required. ■

Appendix E

Miscellaneous Tables

		DY									
AY	Premium	1	2	3	4	5	6	7	8	9	10
1	62467	16864	15508	9341	3537	1853	1184	500	308	338	50
2	59821	14528	17727	8747	4149	2252	715	325	261	255	
3	62968	14241	13763	7512	5207	2068	1674	219	421		
4	64453	14765	14323	8426	6513	3144	1067	913			
5	71185	16395	17038	9826	6381	4037	1839				
6	82793	18136	21582	13415	8519	4583					
7	100826	24727	24037	15181	7105						
8	98358	24749	24501	11830							
9	76653	23063	21035								
10	71326	20083									

Table E.1: Auto data example of [Zhang and Dukic, 2013]: Personal auto line ($1,000s)

		DY									
AY	Premium	1	2	3	4	5	6	7	8	9	10
1	42847	5407	9015	4641	3384	1695	1262	1425	373	241	6
2	38829	6279	8725	6172	4494	2110	919	447	202	69	
3	43001	7256	8667	4778	4262	2884	1427	889	493		
4	41840	5028	5317	4697	3795	2871	1100	657			
5	44525	5721	6097	6389	3802	4306	862				
6	50923	7413	9385	7772	5850	3383					
7	56601	10868	12337	7966	8531						
8	54609	10143	14193	8070							
9	47204	9596	12235								
10	42412	9076									

Table E.2: Auto data example of [Zhang and Dukic, 2013]: Commercial auto line ($1,000s)