Distributionally Robust Performance Analysis with Applications to Mine Valuation and Risk

Christopher Dolan

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

ABSTRACT

Distributionally Robust Performance Analysis with Applications to Mine Valuation and Risk

Christopher Dolan

We consider several problems motivated by issues faced in the mining industry. In recent years, it has become clear that mines have substantial tail risk in the form of environmental disasters, and this tail risk is not incorporated into common pricing and risk models. However, data sets of the extremal climate behavior that drive this risk are very small, and generally inadequate for properly estimating the tail behavior. We propose a data-driven methodology that comes up with reasonable worst-case scenarios, given the data size constraints, and we incorporate this into a real options based model as in [10] for the valuation of mines. We propose several different iterations of the model, to allow the end-user to choose the degree to which they wish to specify the financial consequences of the disaster scenario. Next, in order to perform a risk analysis on a portfolio of mines, we propose a method of estimating the correlation structure of high-dimensional max-stable vectors. Using the techniques of [45] to map the relationship between normal correlations and max-stable correlations, we can then use techniques inspired by [4, 44, 56] to estimate the underlying correlation matrix, while preserving a sparse, positive-definite structure. The correlation matrices are then used in the calculation of model-robust risk metrics (VaR, CVAR) using the the Sample-Out-of-Sample methodology [5]. We conclude with several new techniques that were developed in the field of robust performance analysis, that while not directly applied to mining, were motivated by our studies into distributionally robust optimization in order to address mining problems.

Contents

List of Figures

List of Tables

Acknowledgements

First of all, I would like to sincerely thank my advisor, Jose Blanchet, who has been endlessly patient with me and has provided me with invaluable support and advice through this whole process. I would not have been able to do this without him.

I would like to thank my committee members, Richard Davis, Daniel Rabinowitz, Garud Iyengar, and Upmanu Lall for their time, their helpful discussions and their comments. I would also like to thank the faculty and staff, especially Dood Kalichan and Anthony Cruz for all the assistance they have provided me over the years. I am also grateful to my friends and fellow classmates at Columbia.

On a personal note, I would like to thank my parents, Robert and Antonella Dolan, and my sister, Stephanie Dolan-Foss, who have also been immeasurably patient with me over the years. This would not have been possible without you. Finally, I would like to thank Courtney Razner for her unwavering support and love.

CHAPTER 1

Introduction

This dissertation was motivated by a number of problems that were encountered in the course of a multi-year project to evaluate financial and environmental risk for mining corporations. Fundamentally, we were asked to identify environmental risks that would threaten the operating status of a mining project, quantify the financial impact of a particular risk event occurring, and incorporate these two ideas into pricing models and risk metrics. While seemingly straightforward, we found that in many ways, existing models for these problems were unsatisfactory, and did not capture many of the features seen in the real world. Moreover, certain data were very limited, especially by design (e.g., climate maxima), and made estimation error a significant problem. This motivated an approach to these problems that was focused on the idea of model robustness - because many of the features of the data were poorly understood, we did not want to burden ourselves with undue assumptions or overly complicated models, while at the same time being able to account conservatively for the divergence from a simpler model from reality.

In essence, the idea of distributionally robust calculations follows this type of "flavor": Given some data $X_1, ..., X_n$, we calibrate a simple parametric model P_{ref} where the data are distributed according to $f(x|\theta)$, where θ represents the underlying set of parameters. Given some function of concern, $g(X)$, and a distributional discrepancy $D(P, P_{ref})$, which measures (in some sense) deviations between models P and P_{ref}, we wish to calculate:

$$\max_{P:D(P,P_{ref})\leq\delta} E^P[g(X)]$$

where $E^P[g(X)]$ is the expectation of $g(X)$ under the measure P (we could also take a minimum). Estimating δ is quite important - and the methods for doing so will depend on the exact specification of the problem. For a very small size, for example, δ can be chosen as $D(P_n, P_{ref})$, where P_n is the empirical distribution, this will give the user conservative upper and lower bounds on the expectation of the function, incorporating the possible misspecification of the model. This is especially important

when trying to estimate extreme quantiles of distributions with small sample sizes. For large sample sizes, δ can be chosen according to an expert opinion, or through the use of certain asymptotics. The function g can be a pricing function in a financial application, or it can be an indicator function for an extreme event. The most common discrepancies used for $D(\cdot,\cdot)$ are the Kullback-Leibler divergence and the Renyi divergence; we will also consider Wasserstein-type distances. We will see in Chapter 2 how this can be applied to the pricing of a mine.

Since the 1980s, many projects and investments undertaken by corporations have been evaluated using the "real options" methodology [10] - essentially, the idea is that every project results in a series of cashflows based on traded (or non-traded) financial risks, and that the decision maker on a project will take actions with the intent of maximizing the value of the project; these can then be priced like an option under a risk-neutral measure. To use the example of a mine, consider the simplest idea possible - a mine operator will extract q units of a mineral at time T with price S_T per unit(following, say, a Geometric Brownian Motion) and cost C per unit. If the mineral is not extracted at time T, the lease expires and the operator can no longer extract anything. Then the mine can be viewed simply as q European options on the mineral with unit price S_T, with strike C and expiry T. More generally, a mine operator can be viewed as having Q units of mineral reserves, with a rate of extraction q, and cost of extraction C. At any time, he holds the option to abandon the mine, pay a cost to have it closed, or to operate the mine and earn profits of $q(S-C)dt$ until the Q units of reserves are depleted. At every step, he will make a decision (based on his operating strategy), to maximize the discounted future value of the mine plus current cashflows. This can be viewed as a stochastic control problem which can be used to find the optimal strategy, and hence the price of the mine.

The shortcoming of this approach lies in the fact that the risks it considers are purely financial - the market price of the underlying mineral - but it ignores the environmental risks that are perhaps the most significant driver of value in the field. An environmental disaster at a mine site will frequently reduce the value of a mine to zero, as it is left inoperable, or even less than zero when one factors in reparations and fines levied by the local government, and lawsuits from damaged parties. Moreover, these environmental disasters (namely tailings dam failures), have demonstrably occurred

at a much greater rate than predicted by models and engineering specifications. Consequently, the value of mines given by the standard real option models will be greater than predicted.

In Chapter 2, we develop a model for pricing mines that incorporates environmental disasters, up to a level of specification in the design of the mine - if a mine is built to withstand up to a 1-in-100 year climate event, failures will occur at that rate. We assume this as a user input. However, given the difficulties associated with estimating tails quantiles of distributions using small data sets, we then incorporate a date-driven methodology for estimating a robust worst-case probability of failure [7]. We assume that the threshold for a climate event with probability p of occurring in a particular year was estimated by looking at the time series of daily maxima for precipitation. The distribution of annual maxima would then be estimated by looking at the annual maxima from that time series, and calibrating an appropriate GEV model using maximum likelihood estimation, at which point the threshold would be the $1-p$ quantile of the calibrated (or reference) distribution. The builder of the mine would then build their tailings dam to an appropriate height and other specifications to be able to withstand such an event. We can then incorporate the robustification methodology discussed earlier on the time series - finding the greatest (and least) possible probability of exceeding that threshold in a given year, given the divergence between the actual data and the calibrated model. We can use this robust worst-case probability then as an input into the pricing engine, to come up with a set of lower bounds for mine that, as we shall see, perform well when used with real data.

While this pricing engine can be used to value individual entities and calculate hedge ratios, calculating risk metrics on a portfolio of mines is somewhat more complicated. While one can simulate the underlying mineral prices to reprice the portfolio, we must be able to jointly simulate the maxima of precipitation events over a wide variety of mine sites. The theory of max-stable vectors - where the marginal distributions are GEVs - has been developed in recent years, but estimation is still an issue. The most common approaches involve pairwise [38] or triple-wise [13] composite likelihoods. However, when the dimension of the locations grows large (relative to the number of observations), the performance of these estimation techniques deteriorates rapidly. For example, if the number of locations is equal to 10, the density of the corresponding multi-dimensional max-stable distribution has 10^5 terms [6, 55].

In Chapter 3, we develop a family of algorithms for estimating the underlying covariance matrix Σ of a max-stable vector of the following form (component-wise):

$$M(i) = \max_{n \geq 1} \log \{ -\log(A_n) + X_n(i) + \mu(i) \}$$

where the X_n are iid realizations of a normal random vector with zero mean and covariance matrix Σ and the A_n are an increasing sequence of sums of independent $exp(1)$ random variables. [19]. We refer to the distribution of X_n as the generative distribution, or, in a slight abuse of the language, the generative process. In order for this problem to be fungible, we must make assumptions about the structure of the covariance matrix Σ, namely, that it is sparse, and therefore, the number of parameters being estimated will be significantly less than the dimension d of the underlying data. Since the link function between the correlation of the generative process, Σ, and the correlation of the corresponding max-stable vector is unknown, but we denote it by $C_M(\Sigma)$, we use the exact simulation techniques developed by [45] in order to estimate the function $C_M(\Sigma)$. We use ideas from the theory of covariance matrix thresholding, and advances by [44, 62], to develop a penalized estimator that still satisfies the constraints of a covariance matrix - that is, it is strictly positive definite, while also being sparse, namely, we solve the optimization problem

$$\begin{aligned} \max_{\Sigma : \Sigma_{ii} = 1} \frac{1}{2} \left\| C_M(\Sigma) - S \right\|_F^2 &+ \lambda_1 \left\| H(C_M(\Sigma) - C_M(S)) \right\|_{1,\text{off}} \\ &+ \lambda_2 \left\| H(C_M(S) - C_M(\Sigma)) \right\|_{1,\text{off}} \\ \text{s.t. } \Lambda_{min}(\Sigma) &\geq \tau \end{aligned}$$

where S is the correlation matrix of the data, $\Lambda_{min}(\Sigma)$ is the minimal eigenvalue of the matrix Σ, and $H(A)_{ij} = A_{ij} I(A_{ij} > 0)$. The norm $\|\cdot\|_{1,\text{off}}$ is defined as:

$$\|A\|_{1,\text{off}} = \sum_{i \neq j} |a|_{ij}.$$

Once we have a meaningful estimator of the covariance structure of the precipitation vector, we can then proceed to calculate various risk metrics on the portfolio.

In Chapter 4, we use a novel methodology [5] called Sample-out-of-Sample to calculate risk metrics like VaR and CVaR on a portfolio of mines. The sample-out-of-sample methodology combines real observed data with reasonable stressed data in the following manner: given sample data $X_1, ..., X_n$ and stressed data $Y_1, ..., Y_n$, we create the combined data $Z_1 = X_1, ...Z_n = X_n, Z_{n+1} = Y_1, ...Z_{2n} = Y_n$. The SOS profile function $R_n(\theta)$ is defined as:

$$\begin{aligned} R_n(\theta) = \min & \sum \|X_i - Z_k\|_2^2 \, \pi(i,k) \\ \text{s.t.} \ & \sum_k \pi(i,k) = \frac{1}{n} \\ & \pi(i,k) \geq 0 \\ & \sum h(\theta, Z_k)\pi(i,k) = \theta. \end{aligned}$$

The limiting distribution of $R_n(\theta)$ is calculated for a variety of different set-ups of this framework, which can then be used to calculate robust confidence intervals in the normal fashion. When presented with a portfolio of mines, we can use a combination of historical and simulated asset price data along with historical and simulated precipitation to construct the combined data set. The mines can be repriced from the simulated data using the Taylor Expansion and Greeks (partial derivatives) calculated in the initial pricing; the simulated precipitation can be done using the exact sampling methods from [45] and the covariance estimator from Chapter 3. Using this, we can come up with a simulated set of portfolio values, and can come up with robust estimates for their behavior in the tail (VaR and CVaR) using our sample-out-of-sample methodology.

Finally, we independently consider two developments in robust performance analysis. As was mentioned earlier in this introduction, this is typically formulated as the maximization of the expectation of some function over all probability models within some range determined by a divergence measure around a baseline model. However, this fails to preserve certain natural features of the model, such as an i.i.d. model assumptions, and therefore, the bounds given may not be as tight as possible. We derive an asymptotic formula for the bounds given for a rare event in a random walk setting, and calculate the corresponding rate of convergence. While it is natural to add independence constraints to improve the performance of this approach, adding these independence constraints makes the optimization non-convex, and therefore impossible to solve for all practical

purposes. However, we develop a methodology to solve this problem in a large deviations setting for a variety of random walk problems. This is incorporated into an iterative numerical procedure which has excellent performance in simulation studies. We also demonstrate another technique that works outside of the large deviations setting, and that has very promising performance in simulation studies for random variables with finite support.

CHAPTER 2

Robust Real Options Model

The goal of this chapter is to introduce and explain a novel methodology for practical robust pricing using real option techniques. We concentrate on mining projects because there are risk elements in the mining industry that allow us to expose the need for robust pricing. In particular, the presence of what is referred to as private risk in the real options literature and environmental shocks make mining a well suited setting in order to explain our methodology.

In order to understand the elements of our methodology and provide a general overview of the contributions of this chapter we organize this introduction as follows. First we briefly summarize conventional pricing techniques which are prevalent in practice. We will expose positive and negative aspects of current practices with the intent of motivating the approach that we propose as a way to mitigate potential drawbacks in conventional techniques. In the second part of the introduction we explain the elements of our methodology and how these are applied to provide a robust valuation technique. The concept of robustness will be explained at a conceptual level.

2.0.1. Robustness as a Concept. Here, we use the term "robustness" or "robust performance analysis" to quantify the effect of modeling error as in [30] - namely, for a specific aspect of the model, we wish to bound the effect of model error for the expectation of a function of a random variable - in this case, the price of a mine as a function of the price of the underlying asset. In essence, for some random variable X, given a baseline model P^0, and a measure of divergence δ, we wish to estimate:

$$\min \text{ or } \max_P E^P[f(X)]$$

$$\text{s.t.} D(P, P_0) \leq \delta$$

where $D(P, P_0)$ is some suitable measure of distributional discrepancy such that the formulation of our optimization problem is convex. Typical choices for this are the Kullback-Leibler divergence

[42] and the Renyi α-divergence [54]. In many ways, this is similar to the ideas of robust control developed in [33, 34, 52]. It is instructive to think of this as an adversarial optimal control problem - in which the adversary changes the underlying distribution of the model within a constraint on the difference of the models, and we wish to describe this worst case change in probability distribution and quantify the impact thereof.

This is closely related, but distinct from the concept of "robustness" typically seen in the statistics literature. Robust statistics is a collection of methods that aim to reproduce many of the classical statistical methods (for both hypothesis testing and point estimation) that minimize the effects of outliers or small departures from model assumptions [36, 32]. Many of the classical methods are extremely sensitive to modeling assumptions in the sense that small deviations from the assumption of normality in the underlying distribution (or in the use of the central limit theorem for the limiting distribution of a test statistic), the distribution of the test statistic using classical methods can often diverge significantly. For example, the F-test for homogeneity of variances is highly dependent on the normality assumption of both samples. Common ideas within this literature are those of the breakdown point - essentially the percentage of data that does not come from the assumed distribution - and the influence function - essentially the directional derivative between the asymptotic limit of the test statistic under the baseline model in the direction of a mixture of the baseline distribution and a point mass at an outlier. The most common type of estimator in the field are M-estimators: generalizations of the maximum likelihood estimate in which the objective function is chosen to achieve a desirable behavior for the influence function (boundedness, for example). Perhaps the simplest examples of robust estimators are the median (for the location of a symmetric distribution) or interquartile range (for dispersion). The median has a breakdown point of 50% and the interquartile range has a breakdown point of 25%, compared to 0% for the mean and the total range, respectively.

In a sense, the objectives of these two fields of inquiry are similar - namely, grappling with the difficulties caused by model misspecification (even in a rather small sense). In the case of statistical robustness, we are concerned with creating estimators that perform well against adverse model behavior, however, in the case of the "robust performance analysis" - the function that is being tested has been predetermined - fundamentally, one must make a modeling choice for the

most robust model rather than the most robust function; instead, we are concerned with finding the distributions that produce the worst-case performance. In a certain sense, these two approaches are mirror images of one another, while also having very different purposes.

2.0.2. A quick summary of standard pricing techniques: advantages and drawbacks. Perhaps the most basic valuation methodology relates to the use of discounted cashflow. This technique, which is widely used in practice, has the benefit of being very simple to conceptualize. The disadvantage is that the underlying assumptions tend to oversimplify dynamic and random aspects of investing under uncertainty. For example, under the discounted cashflow technique one typically assumes that the manager or owner of a project will exploit a mine at a given pre-specified rate without the opportunity to react to adverse or unexpected outcomes, such as unusual fluctuations of the underlying commodity price. A more sophisticated form of cashflow analysis allows the introduction of probability distributions to recognize risk and uncertainty. In such case, the expected net present value (NPV) of future payments with a suitable discounted rate (often referred as the Internal Rate of Return, or IRR). The problem is, however, that unless this probability distribution is well calibrated (relative to marketable assets) and the manager's decision making process in recognized, the cashflow analysis might lead to pricing which might not be arbitrage free (i.e. the pricing might give rise to profit opportunities without any risk), which is not a realistic consequence in any practical pricing model.

Risk neutral pricing is a technique which is closely related to discounted cashflow with the inclusion of probability distributions introduced for modeling uncertainty [18, 20, 35]. The difference is that risk neutral pricing is built from a fundamental characterization of arbitrage free models, which is crucial to calibrate the underlying probability models used to model uncertainty. In essence, the idea of risk-neutral pricing is that there is a set of probability measures where the price of each asset is equal the discounted (at the risk-free rate) expectation of the future value of the asset under this probability measure. Namely, given a set of linear instruments, and such a probability measure, an arbitrage-free price for an derivative instruments can be calculated. In the simplest case, this is a probability measure where assets grow (on average) at the risk-free rate, and discounting is carried out at the risk-free rate. Such a characterization fully specifies the family of probability models which might be used in the discounted cashflow methodology to avoid creating an arbitrage. In

simple words, from a purely mechanical standpoint, risk neutral pricing is closely related to cash-flow analysis. By mechanical we mean that the pricing formulas are identical once the probability distribution and the discount rate are specified. The differences are the following: First, the probability distribution that is used for randomness is calibrated to avoid arbitrage opportunities in the risk neutral pricing setting (as opposed to using a statistical or an expert judgement for the actual likelihood of certain events). Second, the discount rate is always the risk free rate in the risk neutral methodology approach, whereas in the cashflow analysis setting, the rate is adjusted to reflect the risk of the investment to produce the IRR.

It is often argued in the literature that the cashflow analysis is equivalent to the risk neutral pricing approach because one can always reverse engineer the IRR to equate any given risk neutral pricing valuation. This is correct, again, mathematically, but the risk neutral pricing methodology is slightly more convenient because it implicitly assigns an IRR to every single eventuality that can occur – every single possible bet or outcome. A particular mining project will be the sum of many such events and it is conceptually easier (from the standpoint of calibration) to work with the sum of the pieces [28, 10].

Since the risk neutral methodology allows to provide arbitrage free prices to every possible uncertain outcome, then we can now incorporate the impact of managerial decisions, dynamically in time, in the face of uncertainty. This is precisely the idea behind real option valuation. It recognizes that investment opportunities (such as the development of a mining projects, among other examples) have embedded decisions which mimic the features of financial options, but whose impact have real consequences in the operation and cashflow of an investment. [2, 61].

Although the real option pricing methodology is conceptually very advantageous, in practice, it may be practically difficult to implement because there are many different probability distributions which are arbitrage free. Thus, it can be specially challenging to calibrate in situations in which there is very limited information on consequential risks. Also, where private risk involved, it is not priced by the market and is therefore particularly challenging to estimate [60, 59]. In standard cashflow analysis the IRR serves as a mechanism to summarize the market price of risk. Regardless of the potential conceptual shortcomings, earlier, the idea of having a single parameter which allows

to one to assess the impact of risk and uncertainty is quite appealing and this is a lesson which we shall keep in mind as we introduce and explain our methodology.

2.0.3. General overview of our robust pricing methodology. We have mentioned that real option pricing allows us to apply non-arbitrage principles to investment valuation. The idea is to recognize which risk factors are marketable (i.e. priced by the market) and which other factors are not directly priced by the market (for example, risks such as tailing dams failures - see Section 2.3 or social unrest) [59, 60, 8].

The key elements for the implementation of the procedure that we propose can be described as follows:

(1) We assume that there exists a basket of single assets (mines) for which their respective cost structure (the details of which are described in Section 2) is well known (not necessarily the same among these assets) and for which a transaction price (which we take as the market price) is known at a particular point in time.

(2) We proceed by identifying risk factors which can be calibrated from market information (we call these marketable factors). For example, the price of the underlying commodity (such as gold) is one of such assets. We advocate the use conventional financial models for marketable factors [37].

(3) We introduce a flexible, yet simple, model which can be used to incorporate both marketable factors and risk whose market price cannot be calibrated directly from market information (we call them private risk factors). This may include, as an example, the occurrence of tailing dam failures, or flooding at the site, or the loss of production due to other natural hazards or environmental conflicts.

(4) We calibrate the marketable risk factors using market information (such as financial options traded). We use a range of techniques, both objective (such as statistical modeling) and subjective (such as the opinion of analysts and experts) to calibrated private risk factors.

(5) We obtain a probabilistic model, which we call the *baseline model*, for both private and marketable risks combined. We use the notation P_0 to compute probabilities of events under the baseline model. We assume that these types of risks are statistically independent. Our robust methodology will correct for this assumption, as we shall explain momentarily.

(6) We apply the real option pricing methodology under the model obtained in E), for each of the companies in A). This step is carried out using dynamic programming. The mechanics are somewhat technical, but they are standard and efficient numerical methods (which we adopt) are explained in [10]. The value obtained is called the *baseline price*, and the optimal policy obtained by applying dynamic programming is called the *baseline policy*.

(7) We define a family of probability models (that is, probability measures defined on the same set of events as the baseline model) which forms a "neighborhood" around the baseline model. The size of the neighborhood is measured by a parameter $\delta \geq 0$. Such a neighborhood is called the uncertainty set around P_0 and it is denoted as $\mathcal{U}_\delta$. If $\delta = 0$ then there is no uncertainty about the baseline model and therefore $\mathcal{U}_\delta$ is a set with a single element, namely, P_0; so we write $\mathcal{U}_\delta = \{P_0\}$. We will discuss how we construct $\mathcal{U}_\delta$ in the sequel. Let us continue and explain how we calibrate δ.

We now advocate two procedures using ideas from Extreme Value Theory and Robust Modeling:

Procedure 1: Robust Pricing with Rainfall Data

(1) Assuming that the occurrences of a disaster within particular time periods, are independent and identically distributed, and that a mine has been built to withstand a 1-in-$\frac{1}{p}$ year rainfall event, we estimate from a time series of rainfall maxima the $(1-p)$ quantile of the distribution. Then, given δ specified by the distance between the empirical data and the fitted distribution, we calculate the worst case arrival probability $p^\star$ within the set $\mathcal{U}_\delta$. The methodology for this will be discussed in Section 2.4.3.

(2) We reprice the mine using a Bernoulli process with the arrival rate $p^\star$. These prices can now be compared to market prices, and used to identify possible value trades.

Procedure 2: Robust Pricing with Market Comparisons

(1) Suppose that $\delta > 0$ is fixed, and that αhas been chosen for the particular mine using the GEV procedures as above. Then the family $\mathcal{U}_\delta$ is set. We can produce an interval around

the baseline price for each of the companies in 1). We call this interval an uncertainty interval. This interval is computed for each company as follows. We fix the baseline policy obtained from 6). The upper bound of the interval is obtained by computing the maximum risk neutral price by ranging over all probabilistic models in $\mathcal{U}_\delta$. Similarly, for the lower bound of the interval, we evaluate the minimum risk neutral price over all of the probabilistic models in $\mathcal{U}_\delta$. Finally, δ is selected as the smallest value for which the market price is contained in the corresponding uncertainty interval.

(2) We obtain a series of values $\delta_1, ..., \delta_n$, where n is the size of the basket constructed in item 1). These values should be treated as realizations from a distribution. One might fit some distribution or simply use these values as an empirical sample. We might use $\delta_{(n)} = \max\{\delta_i\}_{i=1}^n$ in order to find uncertainty intervals for companies which are not inside the basket.

We now discuss how to build the set $\mathcal{U}_\delta$. The uncertainty set $\mathcal{U}_\delta$ contains all arbitrage free models which differ (according to a suitable notion of discrepancy, called Renyi divergence) at most by an amount δ. The Renyi divergence [42] is convenient because it is a non-parametric notion (i.e. we do not attempt to specify a particular form of belief because, as we mentioned earlier, we might not have enough information in certain types of private risks). Moreover, the Renyi divergence preserves arbitrage free models in the following sense: if pricing using P_0 to evaluate expected discounted cashflows gives rise to arbitrage free valuations, then any member of $\mathcal{U}_\delta$ will also produce arbitrage free valuations of investments. In fact, as one increases δ, the set $\mathcal{U}_\delta$ will virtually include all possible arbitrage free models in the sense just described [20].

In the next subsection we discuss the differences between our approach and standard real option pricing methodology. We will explain why our approach is particularly well suited for the evaluation of mining assets.

2.0.4. Advantages and drawbacks of our robust pricing methodology. As we mentioned earlier, our robust pricing methodology builds on what is known as the "Integrated Approach" in real options pricing. [8] provides a discussion and a critique of this approach. [8] concludes that the approach tends to be accurate because it recognizes the differences between marketable and private risks, but it also acknowledges that the Integrated Approach might be difficult to apply.

The main problem with applying the Integrated Approach is the choice of the calibration of the underlying baseline model. The methodology that we suggest in this paper provides a practical approach to alleviate this calibration issue to a certain extent.

The introduction of the uncertainty region $\mathcal{U}_\delta$ allows an ambiguous description of a reasonably good baseline model – as opposed to an exact description of it. We demonstrate with empirical examples that simple, yet intuitive, descriptions can be given for the baseline probabilistic description in order to obtain uncertainty intervals of practical use.

The set $\mathcal{U}_\delta$ includes models which explore variations around the baseline model in *every* direction simultaneously. These variations are controlled by a single parameter, δ, which we call the uncertainty size and is dimension-less.

The drawback of the method is, first, that it still relies on knowing a relatively large set of assets for calibration. The second problem is that, for the uncertainty intervals to be useful in practice, the baseline model should be reasonably good so that δ is reasonably small.

2.1. Robust Pricing in Binomial and Trinomial Lattices

Our objective in this section is to focus on the conceptual elements of behind our proposed approach. We therefore concentrate on simple binomial and trinomial models. Moreover, we provide a quick review of basic real option pricing methodology, to introduce terminology and notation.

2.1.0.1. *Binomial models and review of real option valuation via a simple mining model.* In order to motivate the use of real options for valuing mining investment opportunities, we present a simplified example which can be valued on a binomial lattice. We emphasize that this is an idealized situation in which the only source of randomness is a marketable risk, namely, the price of copper. This risk can be diversified by replication and reflects a complete market which means that there is only one way to calibrate the underlying risk neutral probabilities. Our discussion in this subsection borrows from the presentation in [48].

Suppose we have a T-year lease on a copper mine. At the beginning of every year t, we have the option to extract the complete reserves of the mine Q at cost per-unit C, for a profit of $(S(t)-C)\cdot Q$ where $S(t)$ is market price of copper at the beginning of year t, or pay a maintenance cost of M dollars per year.

We assume, for simplicity, that once the copper is extracted, we no longer have to pay a maintenance cost. We also assume that C is time independent.

Let r be the risk-free rate, and let the annual volatility of the price of copper be σ. If we construct a binomial tree of copper prices, we can value the mining operation using a backward recursion which is known as a dynamic programming recursion.

In our binomial tree, we let $u = \exp(\sigma)$ and $d = 1/u$ the factors by which the price of copper moves up and down. That is, given the price $S(t)$ at time t, then $S(t+1) = u \cdot S$ with probability p_u or $S(t+1) = d \cdot S$ with probability $p_d = 1 - p_u$, with

$$p_u = \frac{e^r - d}{u - d}$$

and $p_d = 1 - p_u$. These are the so-called risk neutral probabilities which are obtained using a non-arbitrage argument [48]. We shall use $E_0(\cdot)$ to denote mathematical expectations associated with the use of the probabilities p_u and p_d.

At time $t = T$ (i.e. the end of the leasing period) the operator of the mine will either pay the maintenance cost, or receive the profit from the mine:

$$V(S(T), T) = \max[-M, (S(T) - C) \cdot Q].$$

At time $t = 0, ..., T-1$, the operator of the mine will maximize his value by either extracting the copper, or paying the maintenance cost if the expected continuation value less the maintenance cost is greater than the extraction value. That is,

$$\begin{aligned}
&V(S(t), t) \\
&= \max[-M + e^{-rt} E_0 \left[V(S(t), t+1)\right], Q(S(t) - C)] \\
&= \max[-M + e^{-rt}(p_u V(uS(t), t+1) + p_d V(dS(t), t+1)), Q(S(t) - C)].
\end{aligned}$$

The value of the mining operation today is $V(S(0), 0)$. Consider a mine with the following parameters:

$$S = 0.4$$

$$Q = 150,000$$

$$M = 500,000$$

$$C = 0.5$$

$$T = 10$$

$$\sigma^2 = .08$$

$$r = .1$$

$$u = 1.32$$

$$d = .75$$

$$p_u = .61$$

$$p_d = .39$$

If one were to evaluate this investment opportunity with standard NPV methods, the value of the mine would always be negative, as the current value of copper is below the extraction cost. As such, there is no internal rate of return that will give a positive price for the mine.

However, looking at the lattice in Table 2.1.0.1, the expected value from following this operating policy is $32.5mm. In Table 2.1.0.1, an up move is made by moving directly to the right, and a down move is made by moving to the right and down one step. In the last column, we show the corresponding copper prices for time $T = 10$.

2.1.0.2. *Real option valuation under incomplete markets and private risks.* We now modify the previous simple example. This time we introduce a private risk and we illustrate the use of our robust pricing methodology.

In any given year, we have a fixed, independent probability p_{jump} of there being a disaster at the mine site, in which case, the owner of the mine loses any ability to extract from the mine site, and pays a large fine L. This disaster is idiosyncratic to the mining site, and cannot be hedged in

$T=0$	1	2	3	4	5	6	7	8	9	10	$S(10)$
32.5	48.4	70.6	101.1	142.8	199.3	275.5	377.6	514.2	696.7	940.1	6.76
	17.5	27.9	42.8	63.9	93.4	134.2	189.8	265.1	366.2	501.6	3.84
		7.3	13.4	22.5	35.9	55.6	83.9	123.6	178.4	252.4	2.18
			1.1	4.0	8.7	15.9	27.4	45.1	71.8	111.0	1.24
Mine Value				-1.7	-0.7	0.9	3.6	8.2	16.3	30.6	0.7
					-2.4	-2.1	-1.7	-1.4	-1.0	-0.5	0.4
						-2.1	-1.7	-1.4	-1.0	-0.5	0.23
							-1.7	-1.4	-1.0	-0.5	0.13
								-1.4	-1.0	-0.5	0.07
									-1.0	-0.5	0.04
										-0.5	0.02

TABLE 2.1. Copper Prices and Mine Values

the financial markets with traded instruments. At time $t = T$, then, our expected payoff is:

$$V(S(T), T, p_{jump}) = \max[-M, Q(S(T) - C)](1 - p_{jump}) - p_{jump}L.$$

For time $t = 0, ..., T - 1$:

$$\begin{aligned} V(S(t), t, p_{jump}) &= (1 - p_{jump})E[V(S(t+1), t+1, p_{jump})|\text{no disaster at time t}] - p_{jump}L \\ &= (1 - p_{jump}) \max[-M + e^{-rt}(p_u V(uS(t), t+1, p_{jump}) \\ &\quad + p_d V(dS(t), t+1, p_{jump})), Q(S(t) - C)] - p_{jump}L. \end{aligned}$$

We look at the case where we expect one disaster every 40 years ($p_{jump} = .025$) and have a loss $L = \$15mm$. The value is considerably different from the value of the mine without the disasters, as we can see in Table 2.1.0.2.

One issue that arises in this method of valuation is that is can be difficult to precisely estimate the distribution of the private risks. Moreover, in this case, the value of the mine is in fact quite sensitive to the arrival rate of disasters. We propose here a methodology of putting bounds on the value, given some level of uncertainty in our probability distribution. To quantify this uncertainty,

$T=0$	1	2	3	4	5	6	7	8	9	10	$S(10)$
22.1	35.4	54.5	81.5	119.6	179.2	247.2	350.3	488.8	672.6	916.5	6.76
	11.1	19.9	32.9	51.7	78.8	117.3	171.3	246.4	250.3	488.8	3.84
		3.4	8.7	16.8	29.1	47.5	74.6	113.8	169.0	245.9	2.18
			−1.3	1.4	5.6	12.5	23.5	40.8	67.6	107.9	1.24
Mine Value				−3.5	−2.4	−0.8	1.9	6.6	14.8	29.5	0.7
					−3.8	−3.4	−2.0	−2.3	−1.6	−0.9	0.4
						−3.4	−2.9	−2.3	−1.6	−0.9	0.23
							−2.9	−2.3	−1.6	−0.9	0.13
								−2.3	−1.6	−0.9	0.07
									−1.6	−0.9	0.04
										−0.9	0.02

TABLE 2.2. Mine Valuation Tree with Disaster

we use the Kullback-Leibler divergence:

$$KL(P|P_0) = \sum_x p_0(x) \log \frac{p(x)}{p_0(x)}.$$

In this example, for the sake of simplicity, we restrict our search to probability measures where the arrivals of disasters are memoryless, and have a constant probability of arrival. In later sections, we will develop a methodology to search over all probability measures within a particular range δ of our base distribution, and develop a methodology for estimating an appropriate δ.

Starting with a base disaster probability p^0_{jump}, which is determined either through statistical methods or expert analysis, we specify a level of tolerance δ and solve the following optimization:

$$\min_{0 \le p_{jump} \le 1} V(S(0), 0, p_{jump})$$
$$\text{s/t}(T+1)(p_{jump} \log(\frac{p_{jump}}{p^0_{jump}}) + (1-p_{jump}) \log(\frac{1-p_{jump}}{1-p^0_{jump}})) \le \delta$$

using a search on p_{jump} to match the KL-divergence constraint. Note that this is actually the solution to min-max problem, as $V(S(0), 0, p_{jump})$ is the value of the mine under the optimal extraction strategy for the probability measure with p_{jump} as the intensity. We can also solve the

δ	min	max
.005	20.45	23.76
.01	19.78	24.46
.02	18.8	25.42
.05	16.97	27.3

TABLE 2.3. Upper and Lower Bounds as Function of δ

corresponding maximization problem to come up with a "confidence interval" for a reasonable range of values for the mine given our base disaster estimate. See Table 2.1.0.2 for an example.

2.2. Real Options Review

As we have mentioned earlier, the most basic and fundamental method of valuation in finance is the discounted cashflow model. For many capital budgeting decisions, however, discounted cash flow based valuation methods do not account for the flexibility in operations that many projects allow, nor do discounted cash flow methods.

The recognition by [10] that these flexibilities could be formally described as options has allowed for the development of methods for valuation and hedging for a variety of many projects using the tools of mathematical finance. In [8], this is referred to as the "classic approach" - which relies heavily on the ideas taken from classical mathematical finance, as it assumes that the risk inherent in natural resource investments, derives entirely which the uncertainty is a traded metal or mineral with a large and liquid market. In this approach, there exists a self-financing, dynamically traded replicating portfolio that can perfectly hedge the stream of cash flows and decisions made by the operator of the mine, guaranteeing an arbitrage free-price (or range of prices, depending on the model of the underlying being used).

2.2.1. Classic Approach for Mine Valuation. Fundamentally, a mine can be viewed as a collection of reserves, Q, with a unit extraction rate q and unit extraction cost C. If a mine is open, the operator collects $q(S-C)dt$ in profits during the time interval $[t, t+dt]$. If the mine is treated as static investment, where the underlying mineral is simply extracted until the reserves are depleted,

the value of the mine can be expressed as:

$$V(S) = \int_0^{Q/q} q(Se^{(r-d)t} - C)dt \tag{2.2.1}$$

where S is the current value of the mineral, r is the risk-free interest rate, and d is the lease rate of the mineral. This price can be attained through a static hedging strategy by holding a portfolio of forward contracts on the underlying. Ignoring the decisions available to the mine owner can lead to clearly underestimating the value of the mine. In particular, this is evident from the fact that the estimated value indicated in (2.2.1) can be negative whenever the cost of extraction is sufficiently large relative to the price of the underlying.

In order to accurately evaluate the value of the mine we must recognize that the following states are possible for the mine (representing decisions available to the owner):

- Open
- Closed
- Abandoned

As in [10], we let a mine have the following parameters:

- Initial Reserve level $Q(0)$
- Extraction rate q
- Unit Extraction Cost C
- Annual Maintenance Cost M
- Switching Cost Open-to-Closed K_1
- Switching Cost Closed-to-Open K_2
- Cost growth rate π
- Tax Rate t_r
- Property tax t_p

Underlying Price	Operating State	P&L
$S(t) > \phi_1(t, Q(t))$	Continue Open	$q(1-t_1)(S(t)-C)dt$
$\phi_3(t, Q(t)) < S(t) < \phi_1(t, Q(t))$	Closed	$-K_1 - Mdt$
$S(t) < \phi_3(t, Q(t))$	Abandoned	0

TABLE 2.4. Net P&L for operating decisions from $[t, t+dt]$ when Open

Let $s = (s(t) : t \geq 0)$ be a price path of the underlying resource. We define $S(t) = s(t)e^{-\pi t}$ be the cost adjusted price process. Let $Q = \{Q(t) : t \geq 0\}$ be the level of reserves at time t. Let

$$\phi(t, Q) = \begin{bmatrix} \phi_1(t, Q) \\ \phi_2(t, Q) \\ \phi_3(t, Q) \\ \phi_4(t, Q) \end{bmatrix}$$

be an operating strategy where $\phi_1(t, Q)$ is the level of S at which an open mine closes given level of reserves Q, $\phi_2(t, Q)$ is the level of S at which a closed mine opens given reserve level Q, and $\phi_3(t, Q)$ and $\phi_4(t, Q)$ are the levels of S at which the mine is abandoned for an open and closed mine, respectively. Let Φ be the set of all such strategies. See Tables 2.4 and 2.5 for the Profits and Losses (P&L) from $[t, t+dt]$ for a particular operating strategy. For a price path S, and an operating strategy ϕ, let $H(S, \phi)$ be the discounted value of all cashflows from price path S using operating strategy ϕ during the time interval $[0, T]$. The value of the mine under a risk-neutral measure P^* for the price process s is

$$V(s(0)) = \max_{\phi \in \Phi} E^{P^*}[H(S, \phi)].$$

2.2.2. Implementation. Monte Carlo simulation provides a pricing framework that can be used for a variety of models and mine specifications; we can approximate the optimal switching strategy using the Longstaff-Schwartz algorithm [46], similar to [57, 17]. Monte Carlo techniques must be used as the number of state variables is greater than two.

Underlying Price	Operating State	P&L
$S(t) > \phi_2(t, Q(t))$	Open	$-K_2 + q(1-t_1)(S(t)-C)dt$
$\phi_4(t, Q(t)) < S(t) < \phi_2(t, Q(t))$	Continue Closed	$-Mdt$
$S(t) < \phi_4(t, Q(t))$	Abandoned	0

TABLE 2.5. Net P&L for operating decisions from $[t, t+dt]$ when Closed

We present here a version of the implementation of the Longstaff-Schwartz algorithm used in [17].

Using the risk neutral dynamics, we generate N paths of the cost adjusted underlying S at a set of discrete time points equally spaced with distance Δt between consecutive points until a terminal time $T = \kappa \frac{Q}{q}$, for $\kappa > 1$. In order for the valuation to be accurate, T must greater than Q/q. The number of steps in each path is therefore $\alpha \frac{Q}{q} / \Delta t = N_T$.

We let $R(j) = jq\Delta t$, for $i = 0, ..., \frac{Q(0)}{q \cdot \Delta t} = N_Q$. We denote by $V_t((S_i(t), R(j))$ and $W_t(S_i(T), R(j))$ the realized value of the open and closed mine at the t-th time step, for price path i, and reserve level $R(j)$. The cash flow when the mine is open is:

$$CF(S(t)) \quad = \quad (1-t_r)(S(t)-C) \cdot q \cdot \Delta t.$$

At terminal time T, we assume the value of the mine is identically zero for each time after that points. Moreover, whenever the reserves are at zero, the value of the mine (open or closed) is zero, regardless of time:

$$V_t(S_i(t), 0) = W_t(S_i(t), 0) = 0, t = 1, .., N_T.$$

At time T, the operator will either have the mine open or abandon the mine, in order to maximize his profits, since the value of the mine will be zero afterwards:

$$\begin{aligned} V_{N_T}(S_i(N_T), R(j)) &= \max(CF(S_i(N_T), 0) \\ W_{N_T}(S_i(N_T, R(j)) &= \max(CF(S_i(N_T) - K_2, 0). \end{aligned}$$

Expected Value	Optimal Decision	Observed Value
$CF(S_i(t)) + f_{t,R(j-1)}(S_i(t))$	Continue Open	$V_t(S_i(t), R(j)) = CF(S_i(t)) + e^{-(r-\pi+t_p)\Delta t} V_{t+1}(S_i(t+1), R(j-1))$
$-K_1 - M\Delta t + g_{t,R(j)}(S_i(t))$	Close	$V_t(S_i(t), R(j)) = -K_1 - M\Delta t + e^{-(r-\pi+t_p)\Delta t} W_{t+1}(S_i(t+1), R(j))$
0	Abandon	$V_t(S_i(t), R(j)) = 0$

TABLE 2.6. Decision tree for Open Mine

Expected Value	Optimal Decision	Observed Value
$-K_2 + CF(S_i(t)) + f_{t,R(j-1)}(S_i(t))$	Open	$W_t(S_i(t), R(j)) = -K_2 + CF(S_i(t)) + e^{-(r-\pi+t_p)\Delta t} V_{t+1}(S_i(t+1), R(j-1))$
$-M\Delta t + g_{t,R(j)}(S_i(t))$	Continue Closed	$W_t(S_i(t), R(j)) = -M\Delta t + e^{-(r-\pi+t_p)\Delta t} W_{t+1}(S_i(t+1), R(j))$
0	Abandon	$W_t(S_i(t), R(j)) = 0$

TABLE 2.7. Decision tree for Closed Mine

After this point, we are able to estimate the continuation value of each state as a function of the price of the underlying at the previous time step.

$$\begin{aligned} e^{-(r-\pi+t_p)\Delta t} V_{t+1}(S_i(t+1), R(j)) &= f_{t,R(j)}(S_i(t)) + \epsilon_{i,1} \\ e^{-(r-\pi+t_p)\Delta t} V_{t+1}(S_i(t+1), R(j)) &= g_{t,R(j)}(S_i(t)) + \epsilon_{i,2} \end{aligned}$$

where

$$\begin{aligned} f_{t,R(j)}(x) &= \alpha_{t,j}(1) + \sum_{\ell=1}^{n} \beta_{t,j,\ell}(\ell) x^{\ell} \\ g_{t,R(j)}(x) &= \gamma_{t,j}(2) + \sum_{\ell=1}^{n} \zeta_{t,j}(\ell) x^{\ell}. \end{aligned}$$

Regressions are carried out separately for each level $R(j)$. The choice of n can vary, for a single state variable, $n = 3$ is typically sufficient. For each level $R(j)$, an operating decision will be made in order to maximize the value of the asset at each time step. We update the observed values V_t and W_t as in Tables 2.6 and 2.7.

Expected Value for Open Mine	Optimal Decision
$CF(S(0)) + f_{0,R(N_Q-1)}(S(0))$	Continue Open
$-K_1 - M\Delta t + g_{t,R(N_Q)}(S(0))$	Close
0	Abandon
Expected Value for Closed Mine	**Optimal Decision**
$-K_2 + CF(S(0)) + f_{0,R(N_Q-1)}(S(0))$	Open
$-M\Delta t + g_{0,R(N_Q)}(S(0))$	Continue Closed
0	Abandon

TABLE 2.8. Decision tree at $t = 0$

For both states, the operator compares the values in the lefthand column, chooses the maximum value, and records the observed values in the corresponding righthand column.

This process repeats until $t = 0$. At time 0,we estimate the continuation values by averaging over all the paths:

$$\begin{aligned} f_{0,R(N_Q-1)}(S(0)) &= \sum_{i=1}^{N} e^{-(r-\pi+t_p)\Delta t} V_1(S_i(1), R(N_Q - 1)) \\ g_{0,R(N_Q)}(S(0)) &= \sum_{i=1}^{N} e^{-(r-\pi+t_p)\Delta t} W_1(S_i(1), R(N_Q)). \end{aligned}$$

The value of the mine will be the expected value of the initial optimal operating decision.

See Algorithm 1 for a summary of this procedure.

This approach of using Monte Carlo simulations to price American Options, and more generally, to solve optimal stopping and optimal switching problems was first proposed by [46]. The theoretical convergence properties of the procedure were unknown until the results of [16]. The key to their insight was that there are two approximations being made: (1) The true value function is being approximated by a finite number n of basis functions and (2) Monte-Carlo simulations and least squares regression are used to estimate the value of the approximated value function. They demonstrated that the solution to the optimal stopping problem using the finite number of basis functions (i.e., approximation 1) approaches with probability 1 the solution to the optimal stopping problem with the true value function, and moreover, that for a fixed number of basis functions approximating the true value function, the Monte-Carlo and regression procedure has almost sure

convergence to the solution of the optimal stopping problem for the approximated value function. Therefore, the whole procedure has almost sure convergence in n and N. Clearly, then the choice of basis functions is very important as $n \to \infty$ is required for almost sure convergence to the true solution, and in practice, it is typical to see $n = 3$ or $n = 4$. [17] provides a method of choosing basis functions based on polynomial forms of the forward price that show excellent empirical performance for mining operations. The case of convergence of multiple-exercise options, as seen here, is somewhat more complicated. While there has been no general proof of the convergence of the algorithm for multiple-exercise options, see [12, 14, 49, 9] for further development of the theory for multiple-exercise options and empirical evidence for convergence.

2.3. Modeling Catastrophe Risk

While the real options framework provides many new insights into the optionality and true value of a mining operation, it fails to account for one major uncertainty in the cash flows of a mine: natural and man-made disasters. Perhaps the most significant of these are tailings dam failures. A tailings dam is used to create a tailings pond, where the byproducts of ore refinement are stored. The tailings pond is retained permanently behind the tailings dam. A tailings pond frequently contains toxic metals such as iron and mercury, and if, released into the environment, can cause major environmental damage and can be deadly to local human and animal populations. Increasingly, tailings dams are being viewed as a liability, in that they fail at a much higher rate than conventional dams. A tailings dam failure will cause significant economic damage to the owner of the mine, as they will be forced to suspend operations until a new tailings dam can be built, and they will be forced to pay for a large amount to clean up the damage that the failure and ensuing flooding have caused, and will incur substantial fines and legal dues.

A recent example of this was the Bento Rodrigues dam disaster in Minas Gerais, Brazil on November 5, 2015. The tailings dam was a property of Samarco, which is a joint venture between Vale and BHP Billiton, two of the largest mining conglomerates in the world. The flooding caused at 17 deaths, and about 60 million cubic meters of iron waste flowed into the nearby Doce River. The Brazilian Government suspended Samarco's activities immediately, and in January 2016, fined

Algorithm 1 To estimate the value of a mine using Longstaff-Schwarz
Given: Mine Structure $(q, Q, C, M, K_1, K_2, t_1, \pi)$, Market Parameters $(S(0), r, d, \sigma)$, and Mine State at $t = 0$

Step 1: Generate N price paths for S with increments of Δt for $N_T = \alpha \frac{Q}{q} \Delta t$, using Geometric Brownian Motion with growth rate $(r - d - \pi)$ and variance σ
Step 2: At time terminal T, estimate the value of the mine.
for $i = 1 : N$
  for $j = 1 : N_Q$
    $V_{N_T}(S_i(N_T), R(j)) = \max(CF(S_i(N_T)), 0)$
    $W_{N_T}(S_i(N_T), R(j)) = \max(CF(S_i(N_T)) - K_2, 0)$
  end loop
end loop
Step 3: Work backwards in time, estimate continuation values, make optimal choice each path, operating state and level of reserves, and then record realized values
for $t = (N_T - 1) : 1$
  for $j = 1 : N_Q$
    Fit regression models $e^{-(r-\pi+t_p)\Delta} V_{t+1}(S_i(t+1), R(j)) = f_{t,R(j)}(S_i(t)) + \epsilon_{i,1}$ and
    $e^{-(r-\pi+t_p)\Delta} W_{t+1}(S_i(t+1), R(j)) = g_{t,R(j)}(S_i(t)) + \epsilon_{i,2}$
    for $i = 1 : N$
      estimate continuation value, make the operating decision, and record values for
      $V_t(S_i(t), R(j))$ and $W_t(S_i(t), R(j))$ as in Tables 2.6 and 2.7 respectively
    end loop
  end loop
end loop
Step 4: Estimate continuation value at $t = 0$

$$f_{0,R(N_Q-1)}(S(0)) = \frac{1}{N} \sum_{i=1}^{N} e^{-(r-\pi+t_p)\Delta t} V_1(S_i(1), R(N_Q - 1))$$

$$g_{0,R(N_Q)}(S(0)) = \frac{1}{N} \sum_{i=1}^{N} e^{-(r-\pi+t_p)\Delta t} W_1(S_i(1), R(N_Q))$$

Step 5: Record optimal expected value for $t = 0$ given the mine state according to to Table 2.8

them R$20 billion (approximately $4.8 billion USD). This does not include the costs of cleaning up the disaster, which will be a minimum of $2.6 billion USD, nor settlements with any of those harmed. While this is an extreme case - it shows that the potential costs of an environmental disaster can in fact exceed the value of the mine.

We propose to incorporate the risk of natural disasters into the real options framework. We assume that disasters arrive according to some risk-neutral process, and that when a disaster happens, the operator of the mine suffers a loss, and the mine is forced closed for some period of time (or permanently). This approach to real options valuation is the "Integrated Approach" from [8],

which was originally developed by [59, 60]. In this case, there are two different types of risk - one is a market-traded, hedgeable risk, and the other is an idiosyncratic, non-market-traded risk, the so-called "private risk" which, in fact, dominates many real investments. For the former risk, we can and should use market inputs - namely, the value of the underlying, and the implied volatility from the corresponding options markets. The other risk here is the failure of the mine due to some exogenous risk factor - this cannot be hedged with a replicating portfolio of the underlying. Instead, we are forced to use a holistic approach to deal with private risks - using subjective probabilities that attempt to closely replicate the real world.

2.3.1. Disaster Model. We assume that the cost-adjusted price paths S follows a risk-neutral process, and that there is a point process $D(t)$ which describes the arrivals of disasters at the mine site. When a disaster occurs, it overrides the operating decision of the operator, incurs a large loss, and forces the mine to be closed for a certain period of time. The loss associated with a disaster (in terms of government fines and legal liability) can be expressed as either a fixed cost L or a draw from a positive random variable $L_{D(t)}$. We assume the closure time is a fixed length of time $T_C = N_D \Delta t$. Hence, we will have a new set of states for when the mine is forced closed $U_t(S(t_t), R(j), k)$ for $k = 1, ..., N_3$. For an operating policy ϕ, a price path S, and a disaster arrival path D, let $H(S, D, \phi)$ be the discounted value of the cashflows from following ϕ with S and D. Given a risk-neutral measure P^*, we can express the value of the mine as

$$E^{P*}[H(S, D, \phi)].$$

We now update the values of our mine according to Tables 2.9 and 2.10.

If the disaster occurs, it overrides the operator's choice to maximize the value of his mine. For each time period, for $l = 1, ..., N_3$. $U_t(S_i(t), R(j), k) = -Mdt + e^{-(r-\pi)\Delta t} U_{t+1}(S_i(t_j), R(j), k+1)$. When $k = 0$, the operator has the choice to re-open or abandon the mine according to the following decision tree in Table 2.11.

Expected Value	Optimal Decision	Observed Value
$CF(S_i(t)) + f_{t,R(j-1)}(S_i(t))$	Continue Open	$V_t(S_i(t), R(j)) = CF(S_i(t)) + e^{-(r-\pi+t_p)\Delta t} V_{t+1}(S_i(t+1), R(j-1))$
$-K_1 - M\Delta t + g_{t,R(j)}(S_i(t))$	Close	$V_t(S_i(t), R(j)) = -K_1 - M\Delta t + e^{-(r-\pi+t_p)\Delta t} W_{t+1}(S_i(t+1), R(j))$
0	Abandon	$V_t(S_i(t), R(j)) = 0$
	Disaster	$V_t(S_i(t), R(j)) = -L + e^{-(r-\pi+t_p)\Delta t} U_{t+1}(S_i(t+1), R(j), N_3)$

TABLE 2.9. Open Mine Operating Decisions

Expected Value	Optimal Decision	Observed Value
$-K_2 + CF(S_i(t)) + f_{t_j,R(j-1)}(S_i(t))$	Open	$W_t(S_i(t), R(j)) = -K_2 + CF(S_i(t)) + e^{-(r-\pi+t_p)\Delta t} V_t(S_i(t+1), R(j-1))$
$-M\Delta t + g_{t,R(j)}(S_i(t))$	Continue Closed	$W_t(S_i(t), R(j)) = -M\Delta t + e^{-(r-\pi+t_p)\Delta t} W_{t+1}(S_i(t+1), R(j))$
0	Abandon	$W_t(S_i(t), R(j)) = 0$
	Disaster	$W_t(S_i(t), R(j)) = -L + e^{-(r-\pi+t_p)\Delta t} U_{t+1}(S(t+1), R(j), N_D)$

TABLE 2.10. Closed Mine Operating Decisions

Expected Value	Optimal Decision	Observed Value
$-K_2 + CF(S_i(t)) + f_{t,R(j-1)}(S_i(t))$	Open	$U_t(S_i(t), R(j), 0) = -K_2 + CF(S_i(t)) + e^{-(r-\pi+t_p)\Delta t} V_{t+1}(S_i(t+1), R(j-1))$
$-M\Delta t + g_{t,R(j)}(S_i(t))$	Close	$U_t(S_i(t), R(j), 0) = -M\Delta t + e^{-(r-\pi+t_p)\Delta t} W_{t+1}(S_i(t+1), R(j))$
0	Abandon	$U_t(S_i(t), R(j), 0) = 0$
	Disaster	$U_t(S_i(t), R(j), 0) = -L + e^{-(r-\pi+t_p)\Delta t} U_{t+1}(S(t_j), R(j), N_D)$

TABLE 2.11. Forced Closed Mine Operating Decisions

Alternatively, mine closures (from a disaster) can be assumed to be permanent - in which case, $U_t(S(t), R(j)) = 0$. See Algorithm 2 for a complete description of the process with permanent mine closures.

Algorithm 2 To estimate the value of a mine with a disaster process using Longstaff-Schwarz
Given: Mine Structure $(q, Q, C, M, K_1, K_2, t_1, \pi)$, Market Parameters $(S(0), r, d, \sigma)$, Failure intensity λ Mine State at $t = 0$

Step 1: Generate N price paths for S with increments of Δt for $N_T = \alpha \frac{Q}{q} \Delta t$, using Geometric Brownian Motion with growth rate $(r - d - \pi)$ and variance . Generate N disaster arrival paths, where $D_i(t) = 0$ indicates no disaster during the time period, and $D_i(t) = 1$ indicates a disaster. Each time period is i.i.d. and $P(D_i(t) = 0) = e^{-\lambda \Delta t}$.
Step 2: At time terminal T, estimate the value of the mine.
if $D_i(N_T) = 0$ then
    for $i = 1 : N$
        for $j = 1 : N_Q$
            $V_{N_T}(S_i(N_T), R(j)) = \max(CF(S_i(N_T)), 0)$
            $W_{N_T}(S_i(N_T), R(j)) = \max(CF(S_i(N_T)) - K_2, 0)$
        end loop
    end loop
else if $D_i(N_T) = 1$
    $V_{N_T}(S_i(N_T), R(j)) = 0$
    $W_{N_T}(S_i(N_T), R(j)) = 0$
end if
Step 3: Work backwards in time, estimate continuation values, make optimal choice each path, operating state and level of reserves, and then record realized values
for $t = (N_T - 1) : 1$
    for $j = 1 : N_Q$
        Fit regression models $e^{-(r-\pi+t_p)\Delta} V_{t+1}(S_i(t+1), R(j)) = f_{t,R(j)}(S_i(t)) + \epsilon_{i,1}$ and
        $e^{-(r-\pi+t_p)\Delta} W_{t+1}(S_i(t+1), R(j)) = g_{t,R(j)}(S_i(t)) + \epsilon_{i,2}$
        for $i = 1 : N$
            if $(D_i(t) = 0)$ then
                estimate continuation value, make the operating decision, and record values
                for $V_t(S_i(t), R(j))$ and $W_t(S_i(t), R(j))$ as in Tables 2.6 and 2.7 respectively
            else if $(D_i(t) = 1)$ then
                $V_t(S_i(t), R(j)) = 0$
                $W_t(S_i(t), R(j)) = 0$
            end if
        end loop
    end loop
end loop
Step 4: Estimate continuation value at $t = 0$

$$f_{0,R(N_Q-1)}(S(0)) = \frac{1}{N} \sum_{i=1}^{N} e^{-(r-\pi+t_p)\Delta t} V_1(S_i(1), R(N_Q - 1))$$

$$g_{0,R(N_Q)}(S(0)) = \frac{1}{N} \sum_{i=1}^{N} e^{-(r-\pi+t_p)\Delta t} W_1(S_i(1), R(N_Q))$$

Step 5: Record optimal expected value for $t = 0$ given the mine state according to to Table 2.8

Q	=	$150mm$ lb
q	=	10mmlb/year
C	=	$\$0.5/lb$
K_1	=	$\$200,000$
K_2	=	$\$200,000$
M	=	$\$500,000$/year
π	=	8%/year
r	=	10%/year
d	=	1%/year
t_r	=	50%
σ^2	=	8%/year
λ	=	1/40
L	=	$\$10mm$

TABLE 2.12. Mine Parameters

2.3.2. Results. In this section, we benchmark our results by comparing our (baseline) model with the results from [10], which were acquired using a finite difference solution to a pricing PDE. Table 2.12 shows the various parameters of the simulation and the mine structure. We model the underlying as Geometric Brownian Motion:

$$dS(t) \quad = \quad (r - d - \pi)S(t)dt + \sigma S dW(t),$$

and we model the arrival of disasters as a Poisson point process: $D(t)$ with mean parameter λ. In one case, the mine will be closed for a 2 year period, and in the other, the mine will be permanently closed in the event of a disaster. Table 2.13 shows the valuations against a variety of different starting spot prices.

2.4. Review of GEV models and robustification

We now consider a situation where the designer of a mine use climate information from the mine site to set certain construction parameters. We assume that the industry standard is to build a tailings dam to a 1-in-100 year specification, that is, it is built to withstand precipitation events up to the threshold of a single-day precipitation event that occurs (on average) once in 100 years. For pricing purposes, this would correspond to using an arrival intensity of $\lambda = .01$ as in the model

	Finite Difference		Simulation	
$S(0)$	Open	Closed	Open	Closed
0.3	1.25	1.45	1.17	1.37
0.4	4.15	4.35	4.16	4.54
0.5	7.95	8.11	7.85	7.96
0.6	12.52	12.49	12.52	12.46
0.7	17.56	17.38	17.77	17.57
0.8	22.88	22.68	22.79	22.59
0.9	28.38	28.18	28.36	28.17
1.0	34.01	33.81	34.18	33.98
	2-year Closure		Permanent Closure	
0.3	Open	Closed	Open	Closed
0.4	0	0.09	0	0.04
0.5	1.93	2.13	1.56	1.67
0.6	5.45	5.42	4.6	4.46
0.7	10.27	10.10	8.43	8.23
0.8	14.88	14.68	12.59	12.39
0.9	19.87	19.67	17.09	16.89
1.0	25.53	25.33	21.32	21.12

TABLE 2.13. Values of Mines Against Spot Price

discussed in Section 3. In order to estimate this threshold, we will calibrate a Generalized Extreme Value (GEV) model to the annual maxima of precipitation of the latitude and longitude closest to the mine location, as taken from the ECMWF data set. However, we assume that the estimates of the GEV parameters are inefficient, and we correct for the failure specification using a technique derived from [7].

2.4.1. GEV Models. The Extreme Value Theorem provides a complete classification of all distributions $G(x)$ that form the limit of

$$\lim_{n\to\infty} \frac{M_n - b_n}{a_n}$$

where M_n is the maximum of n independent samples from a random variable X, and a_n and b_n are a series of (deterministic) scaling constants. In other words, if we have:

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n\left(a_n x + b_n\right) = G(x),$$

then $G(x)$ is a member of the family of generalized extreme value (GEV) distributions.

The extremal types theorem [26, 31] states that the CDFs (cumulative distribution functions) of this family can be written as $G_\gamma(ax+b)$ where

$$G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}), \quad 1+\gamma x > 0$$

where $a > 0$, $b, \gamma \in \mathbb{R}$, which are known, respectively as the scale, location, and shape parameters. In the case where $\gamma = 0$, we take the limiting function

$$G_\gamma(x) = \exp(-\exp(-x)).$$

Any distribution $F(x)$ whose limit (in the maximum) converges to a particular $G_\gamma(x)$ is said to belong to the domain of attraction of , which is denoted as $\mathcal{F} \in \mathcal{D}(G_\gamma)$. For the following we denote the right endpoints of a distribution as $x_F^* = \sup\{x : F(x) < 1\}$. There are three particular sub-families of GEV distributions:

(1) The Frechet Distribution ($\gamma > 0$) corresponding to fat-tailed, power-law type behavior. A distribution $F \in D(G_\gamma)$ for some $\gamma > 0$, if and only if the support of the distribution F is unbounded to the right, that is, $x_F^* = \infty$ and its tail probabilities satisfy

$$1 - F(x) = \frac{L(X)}{x^{1/\gamma}}, x > 0$$

where $L(x)$ is a function that is slowly varying at ∞. In other words, for every $t > 0$, $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$.

(2) The Gumbel Distribution ($\gamma = 0$) corresponding to semi-exponentially decaying behavior. A distribution $F \in D(G_0)$ if and only if

$$\lim_{t\to x_{F-}^*} \frac{1 - F(t + xf(t))}{1 - F(x)} = \exp(-x), x \in \mathbb{R}$$

Algorithm 3 Estimate $Var_p(X)$ for $p$close to 1
Given: N independent samples $X_1, .., X_N$ of X

Let $n < N$ and let $m = \lfloor \frac{N}{n} \rfloor$
Step 1: Partition $X_1, ..X_N$ into blocks of size n, and compute the block maxima for each block to obtain $M_1, ...M_m$
Step 2: Calibrate a GEV model using an appropriate parameter estimation technique (maximum likelihood, method of moments, etc.) using the block maxima $M_1, ...M_m$ and obtain parameters $\hat{a}, \hat{b}, \hat{\gamma}$.
Step 3: (Compute the $p^n - th$ quantile of the GEV model): Solve for x such that $G_{\hat{\gamma}}(\hat{a}x + \hat{b}) = p^n$ and let x_p be the corresponding solution.
RETURN x_p

for a positive function $f(x)$.

(3) The Weibull Case ($\gamma < 0$) corresponding to bounded random variables. A distribution $F \in D(G_\gamma)$ for some $\gamma < 0$ if and only if the distribution is bounded to the right, and its tail probabilities satisfy

$$1 - F(x_F^* - \epsilon) = \epsilon^{-1/\gamma} L(\frac{1}{\epsilon}), \epsilon > 0$$

where $L(x)$ is a function that is slowly varying at ∞.

A full discussion of this theory lies beyond the scope of this paper; we encourage the reader to see the standard texts [25, 43] for more information.

2.4.2. Calibration Procedure. The calibration of max-stable distributions can be quite challenging. The most common procedure used is the Peaks-over-Threshold method. In keeping with the methodology of [7] however, we instead use a block-maxima technique [11], (see Algorithm 3). Essentially, we contend that for a sufficiently large sub-sample, the maxima of a block will well-approximate the distribution of the maximum. In this case, we look at the annual maxima of daily precipitation for a particular mine site from the ECMWF. Since this data set covers daily precipitation from 1900 to 2010, each mine site had a 111 (approximate) samples from the limiting distribution. We then use a maximum-likelihood based estimation procedure to estimate the three parameters: γ, a, b and acquire confidence intervals for the parameter γ. The 99% quantile of the calibrated distribution P_{ref} was taken to be the 1-in-100 year specification, $q_{.99}$.

2.4.3. Robustification Procedure.

2.4.3.1. *Divergence Measures.* We consider, though, that the sample size is relatively small and we are highly prone to error in our model specification. We wish instead to estimate the worst-case probability that mine fails in a particular year, looking over all possible models of this probability with a certain range. In order to quantify this distance, we utilize a formal measure of divergence between two probability measures. Consider two probability measures P and Q, where P and Q are both defined on the same $(\Omega, \mathcal{F})$, and P is absolutely continuous with respect to Q). The Radon-Nikodym derivative $\frac{dP}{dQ}$ is then well-defined. For any $\alpha > 1$, the Renyi divergence of degree α is defined as:

$$D_\alpha(P, Q) = \frac{1}{\alpha - 1} E_Q \left[\left(\frac{dP}{dQ} \right)^\alpha \right].$$

For every α, $D_\alpha(P, Q) = 0$ if and only if $P = Q$. Furthermore, the map $\alpha \to D_\alpha$ is nondecreasing, and left continuous. The special case where $\alpha = 1$ is the celebrated Kullback-Leibler divergence:

$$D_1(P, Q) = E^Q \left[\frac{dP}{dQ} \log \left(\frac{dP}{dQ} \right) \right].$$

2.4.3.2. *Robust Bounds on Probabilities.* We denote by P_{ref} the reference probability measure chosen using standard estimation procedures, as in Algorithm 3. Since this estimate is prone to error, we consider all models that are nearby P_{ref} according to the Renyi divergence for a chosen α. Given a desired level of distance δ, and a quantile of interest x_p, we consider an optimization of the form:

$$V_\alpha(\delta) = \sup\{P(M > x_p) : D_\alpha(P, P_{ref}) \leq \delta\}. \tag{2.4.1}$$

As in [7], we choose α to satisfy the equation:

$$\frac{\alpha}{\alpha - 1} |\gamma_{ref}| = |\gamma_{ref}| + \epsilon$$

where ϵ is half the width of the confidence interval obtain when estimating γ_{ref}. This corresponds to having the worst case distribution found from carrying out 2.4.1 for every value of x_p belonging to the domain of attraction with a shape parameter corresponding to the upper limit of the confidence interval. Likewise, we can estimate the level of potentially model misspecification δ by comparing the P_{ref} to the realized distribution of maximums, that is

$$\delta = D_\alpha(P, P_{ref})$$

and estimating α-divergence using the k-nearest neighbor algorithm of [53], with the realized values $M_1, ...M_m$ as the sample for P and random samples from P_{ref}. The k-nearest neighbor algorithm can be summarized (in the univariate case) like so:

(1) Let $X_1, ..., X_n$ be a sample from P_{ref}, and $M_1, ..., M_m$. Let $\rho_k(i)$ be the Euclidean distance of the k-th nearest neighbor of X_i in the sample $X_1, ..., X_n$ and let $\nu_k(i)$ be the Euclidean distance of the k-th nearest neighbor of X_i in the sample $M_1, ..., M_m$.

(2) We calculate an asymptotically unbiased estimator of the α-divergence in the sense of [15] with:

$$\hat{R}_\alpha(P_{emp}, P_{ref}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(n-1)\rho_k(i)}{m\nu_k(i)}\right)^{1-\alpha} B_{k,\alpha}$$

where $B_{k,\alpha} = \frac{\Gamma(k)^2}{\Gamma(k-\alpha+1)\Gamma(k+\alpha-1)}$.

(3) The Renyi divergence of degree of α can be found by substituting:

$$\hat{D}_\alpha(P_{emp}, P_{ref}) = \frac{1}{\alpha}\log \hat{R}_\alpha(P_{emp}, P_{ref}).$$

At this point, we can now solve for the worst-case probability, $p_{\theta_x} = \theta_x p_x$, where θ_x is the solution to:

$$P_{ref}(x,\infty)\phi_\alpha(\theta_x) + P_{ref}(-\infty, x)\phi_\alpha\left(\frac{1-\theta_x P_{ref}(x,\infty)}{P_{ref}(-\infty,x)}\right) = \bar{\delta} \tag{2.4.2}$$

where

$$\phi_\alpha(x) = \begin{cases} x^\alpha & ,\alpha > 1 \\ x\log x & \alpha = 1 \end{cases}$$

$$\bar{\delta} = \begin{cases} \exp((\alpha-1)\delta & \alpha > 1 \\ \delta & \alpha = 1 \end{cases}$$

2.4.3.3. *Climate Risk Robustification Procedure.* We can now apply these techniques to produce robust bounds for the price of a mine. We look at the rainfall data for a site prior to the mine's construction, and calibrate a GEV distribution to the annual block maxima. We can then calculate an appropriate α-divergence using the methods suggested in the previous section. Given a tolerance

Algorithm 4 Estimate $(V_{up}, V_{down})$for a mine
Given failure level p, Pricing Engine $G(p)$ for the mine, and annual block maxima for precipitation at the mine site: $M_1, ..M_m$

Step 1: Calibrate a reference GEV distribution P_{ref}. Find $(1-p)$th quantile x_p of P_{ref}.
Step 2: Choose α such that $\frac{\alpha}{\alpha-1}|\gamma_{ref}| = |\gamma_{ref}| + \epsilon$ where ϵ is $\frac{1}{2}$ of the width of the 95% confidence interval for γ_{ref} obtained from the estimation procedure.
Step 3: Calculate δ as the estimate of $D_\alpha(P, P_{ref})$ using the k-nearest neighbors algorithm.
Step 4: Find $(p_w^*, p_b^*) = \sup / \inf\{P(M > x_p) : D_\alpha(P, P_{ref}) \leq \delta\}$
Step 5: Find best and worst case prices $(V_w^*, V_b^*) = (G(p_w^*), G(p_b^*))$
Return $(G(p_w^*), G(p_b^*))$

threshold for the an extremal climate event, 1-in-$\frac{1}{p}$ year, we can then calculate a worst-case and best-case robust threshold (p_w^*, p_b^*) for the annual probability of failure for a mine, using the optimization procedure (2.4.2). The mine can then be re-priced using the real options methodology in from Section 2.3.1, using the disaster probabilities (p_w^*, p_b^*). We summarize this procedure in Algorithm 4.

2.4.4. Calibration Procedure. Alternatively, we recommend a calibration procedure for the robustification parameter δ, once a GEV model has been fit and α has been chosen. Seeing actual values of mines is rare; it generally only occurs when the owner of the mine sells their stake (or a part thereof).

We assume that mines are priced somewhat consistently across time by the market as a function of spot price, interest rates, lease rates, volatility etc. For a sample of mines, we use a fixed baseline disaster process (or we can make the baseline disaster processes bespoke for each mine, using a very coarse estimate from failures at similar mines). Then we price the mines using the robustification procedure above, and find the value of δ that covers all of the mines, or perhaps covers some percentage of the sample (say 90%). More formally, for a set of mines with market prices M_i observed on date d_i with sport S_i, interest rate r_i, lease rate d_i, and market implied volatility (from a 3-month at-the-money, or ATM, options) σ_i, and letting $\mathcal{U}_\delta$ be the set of all probability measures such that $KL(P|P*) \leq \delta$, let

$$\begin{aligned}
[P_{min,i}(\delta), P_{max,i}(\delta)] &= \min / \max_{P \in \mathcal{U}_\delta} E^P[H(S_i, D_i, \phi)] \\
\delta_i &= \inf\{\delta \geq 0 : P_{min,i}(\delta) = M_i \text{ or } P_{max,i}(\delta) = M_i\} \\
\delta &= \max\{\delta_1, ..., \delta_n\}.
\end{aligned}$$

Mine	Date	Transaction Value	Interest	Reserves	Capacity	Cost
Pinto Valley	4/1/13	650	100%	1683	62	236
Candelaria	11/1/14	1800	80%	4422	148	185
Los Bronces	8/1/12	890	4%	62591	400	141
Condestable	3/1/12	218	46%	423	20	181
Palabora	12/1/12	489	75%	355	51	24
Carmen	6/1/11	368	46%	609	56	175
Batu Hijau	5/1/11	247	7%	3730	109	194
OK Tedi	1/1/11	335	18%	2215	101	173
Las Cruces	11/1/10	552	30%	1746	72	143

TABLE 2.14. Transactions

2.4.5. Results. We now consider a selection of M&A (Mergers and Acquisitions) transactions of copper mines from 2010 through 2014. In order to qualify, the transaction had to have a value of over $\$200,000,000$ and the project had to derive over 90% of its revenues from copper, as well as having available cost and and reserve data available.

We used to following market inputs: For the risk-free interest rate, we used the ten year treasury rate on the first market day of the month of the transaction. Implied volatility was the 90 day at the money volatility on the first market day of the month of the transaction. The lease rate was assumed to be a constant 2%, and the tax rate was assumed to be 38% (these were values were chosen in consultation with a market expert). The base model had shock arrivals were simulated from a Poisson Process with an average of one arrival every 40 years. Closures were assumed to be permanent with no additional losses. See Table 2.14 for the transactions and Table 2.15 for market data. We record the theoretical prices without shocks, with baseline shocks, and with the robust shocks in Table 2.16. See Table 2.17 for the robustification parameters used in the procedure.

Mine	Copper Price	Implied Volatility	10-yr Treasury
Pinto Valley	7509.75	17.6	1.83
Candelaria	6761.5	16.8	2.34
Los Bronces	7418.24	23.1	1.52
Condestable	8625	29.1	2.03
Palabora	7996.5	24.3	1.63
Carmen	9066.9	31.4	2.96
Batu Hijau	8931.7	26.8	3.31
OK Tedi	9553.2	33.9	3.36
Las Cruces	8289.7	30.2	2.66

TABLE 2.15. Market Data

	No Shocks	Shocks	Robust	Transaction Value	Calibrated Delta
Pinto Valley	831.4	791.8	492.4	650	.055
Candelaria	1936.5	1836.9	1489.1	1800	.05
Los Bronces	1331.5	1061.1	374.3	890	$<.001$
Condestable	234.2	224.2	141.3	218	$<.001$
Palabora	517	511.7	467.5	489	.07
Carmen	395.6	387.9	370.3	368	.06
Batu Hijau	308.7	287.3	183	247	.02
OK Tedi	600.9	575.4	385.2	335	.3
Las Cruces	671.8	640.8	544.1	552	.17

TABLE 2.16. Theoretical Prices

2.5. Valuation of Antamina Copper Mine

2.5.1. Standard Model. We present in this section a valuation of the Antamina Copper Mine using the Standard Real Options model - that is, without any sort of failure process. In this framework, a mine can be described using the following parameters:

- An initial cost of extraction C per unit
- An initial level of reserves Q
- A fixed rate of extraction q
- The maintenance cost for a closed mine M

	α	δ	θ
Pinto Valley	1.11	.23	12
Candelaria	1.89	.16	5.4
Los Bronces	1.16	.16	9.2
Condestable	3.76	1.89	13.6
Palabora	1.15	.21	10.7
Carmen	1.81	.05	3.5
Batu Hijau	1.25	.16	8.2
OK Tedi	1.05	.19	11.3
Las Cruces	2.48	.19	4.5

TABLE 2.17. Robustification Parameters

- The costs of switching between open and closed states K_1 and K_2
- A rate of cost growth π
- A tax rate t_r

The following values are extracted from the market:

- The cost of the underlying asset S
- The risk-free interest rate r
- The borrowing cost for the underlying asset d
- The volatility of the underlying asset σ

See Table 2.18 for the values used to price the mine. We vary S and σ to give some intuition about the behavior of the model. The mine is then valued as an option using the optimal extraction strategy given that is has three states:

(1) Open, in which the owner receives cashflow $q(S-c)dt$
(2) Closed, in which the owner pays $-Mdt$
(3) Abandoned, where the value of the mine goes to zero but cannot be re-opened. It is possible to set a renumeration cost for this state as well.

In order to price the mines, we varied S from $\$2.5mm/kt$ to $\$7mm/kt$ and σ from 10% to 25%. Prices are in Table 2.19.

2.5.2. Valuation using Robust Real Options Model with Failures.

Parameter	Value
C	$\$1.66mm/kt$
Q	$5600(kt)$
q	$400(kt/yr)$
M	$\$135mm/yr$
K_1, K_2	0
π	3%
t_r	40%
r	2.2%
d	1.5%

TABLE 2.18. Mine Valuation Parameters

	σ			
$S(0)$	0.1	0.15	0.2	0.25
2.5	2004.2	2192	2334.7	2487.2
3	3413.3	3523.2	3678.1	3730.7
3.5	4897.3	4958.3	5065.4	5058.7
4	6429.5	6434	6505.1	6485.1
4.5	7896.4	7945.2	7965.4	8025.6
5	9431.5	9453.1	9204.1	9347.9
5.5	10950	10931	10810	10650
6	12457	12380	12241	12177
6.5	14016	13656	13532	13536
7	15464	15398	15060	15003

TABLE 2.19. Mine Prices in mm USD

2.5.2.1. *Robustifying Precipitation-Based Disaster Risk.* We observe the time series of precipitation at the mine site, and calibrate a GEV model as would be done by the engineers building such a mine. As construction was started in 1998, we look at the annual precipitation maxima from 1977-1997 to calibrate our reference GEV model, as well as the case where the mine was fitted with the complete rainfall data going back to 1900. We look also look at the cases where the mine was

	1977-1997	1900-1997
γ	.16	$-.17$
σ	.05	.067
μ	.68	.65
$q_{.99}$	1.03	.87
$q_{.98}$	.96	.845

TABLE 2.20. Estimated GEV Parameters

	1977-1997	1900-1997
α	1.38	2.75
δ	.21	.31
worst case misspecification for 1 in 50 year event	.124	.08
worst case misspecification for 1 in 100 year event	.089	.062

TABLE 2.21. GEV distribution robustification

built to a 1-in-100 year specification, as well as 1-in-50 year specification. See Table 2.20 for the calibration of the GEV models on the two data sets.

Once the data has been chosen, we estimate the α-divergence of the data from the calibrated model (α is chosen using the confidence intervals from the estimation process for shape parameter γ), and estimate a worst-case probability of exceedance over all models within this distance from the reference (calibrated) distribution. We report the robustification parameters and the worst-case exceedances for both levels and both training sets of data in Table 2.21.

2.5.2.2. *Pricing Using the Robust Real Options Model.* We now incorporate failure events into the real options model. The failure events are simulated from a Poisson process with the intensity q_p^* where p is the risk assumed by the mine's engineers, in this case $p = \frac{1}{50}$ or $p = \frac{1}{100}$. The values for q_p^* can be seen in Table 2.21 for the two different reference distribution calibrations. When the failure event occurs, the value of the mine goes to zero and no future cashflows can be extracted from it. It is possible to include in this framework a further penalty, but because of the difficult of estimating such a cost, we omit it for now. The prices we obtain from this methodology using the 1977-1997 data set and the 1900-1997 data set are summarized in Tables 2.22 and 2.23 respectively. As we can see, since the mine is effectively deep in the money for all of these spot values, the mine

$p = .02/q_p^* = .124$	Sigma			
S	0.1	0.15	0.2	0.25
2.5	1074	1141.2	1171.2	1265.8
3	1810.9	1835	1887.1	1914.6
3.5	2579.7	2595.2	2642	2650.1
4	3333.9	3368.6	3370.5	3387.4
4.5	4119	4126.4	4124.1	4169.1
5	4843.4	4866	4871.1	4897.1
5.5	5646.1	5603.6	5631.3	5621.9
6	6359.9	6386.4	6336.4	6376.2
6.5	7151	7128.8	7173.4	7152.9
7	7912.8	7901.7	7939.7	7913.8
$p = .01/q_p^* = .089$	Sigma			
S	0.1	0.15	0.2	0.25
2.5	1251.9	1341.8	1445	1492
3	2121.3	2176.8	2214.9	2343
3.5	3005.1	3060.1	3098.2	3091.5
4	3901.2	3958.4	3925.7	4082.6
4.5	4835	4818.2	4802.3	4910.4
5	5767.8	5753.4	5719.2	5805.6
5.5	6651.6	6582.3	6627	6725.7
6	7522.4	7485.3	7532.4	7542.9
6.5	8488.2	8347.4	8438.3	8341.6
7	9297.1	9312.8	9247.7	9273.3

TABLE 2.22. Robustified Lower Bounds of Mine Prices calibrated from 1977-1997

price is in fact, rather linear with regards to spot, with the option premium decaying slightly with increased volatility for the deepest in-the-money mines.

Since the greatest degree of convexity is demonstrated around the "strike price", i.e., the cost of production, we record the behavior of the mine value around this area (i.e. for much lower prices) to illustrate the convexity demonstrated in the real options models, as well as the values of mines where the cost of production exceeds the cost of copper. See Figures 2.1 and 2.2 for a visualization

$p = .02/q_p^* = .08$	Sigma			
S	0.1	0.15	0.2	0.25
2.5	1319.6	1405.7	1491.8	1548.4
3	2207.6	2275.8	2317.5	2385.9
3.5	3128.9	3209	3231.4	3288.4
4	4058.5	4120.6	4197.4	4152.5
4.5	5070.4	5067.9	5077.2	5087
5	5964.9	5967	6021.8	6010.8
5.5	6966.2	6961	6930.9	6960.6
6	7896.3	7946.8	7772.2	7899.8
6.5	8821.8	8776.2	8730.5	8869.8
7	9758.1	9825	9694.5	9690.6
$p = .01/q_p^* = .062$	Sigma			
S	0.1	0.15	0.2	0.25
2.5	1425.8	1538.8	1669.9	1708.7
3	2427.6	2486.4	2555.9	2608.9
3.5	3442.2	3503.5	3522.5	3620.3
4	4494.7	4514.4	4561.1	4600.9
4.5	5544.4	5517.7	5522.2	5664.1
5	6591.4	6559.7	6576.5	6543.8
5.5	7681.2	7656.5	7621.2	7562.4
6	8682.4	8730	8678.1	8592.5
6.5	9772.8	9721.8	9683.2	9558.7
7	10766	10711	10780	10716

TABLE 2.23. Robustified Lower Bounds of Mine Prices calibrated from 1900-1997

of the behavior of the model around this area. See Table 2.24 for a complete tabulation of the maximum rainfall data from 1900-1998.

2.6. Discussion

In this section, we address a possible critique of this method, namely, that there may be issues with the approach suggested in Algorithm 4for the calibration of the underlying GEV distribution,

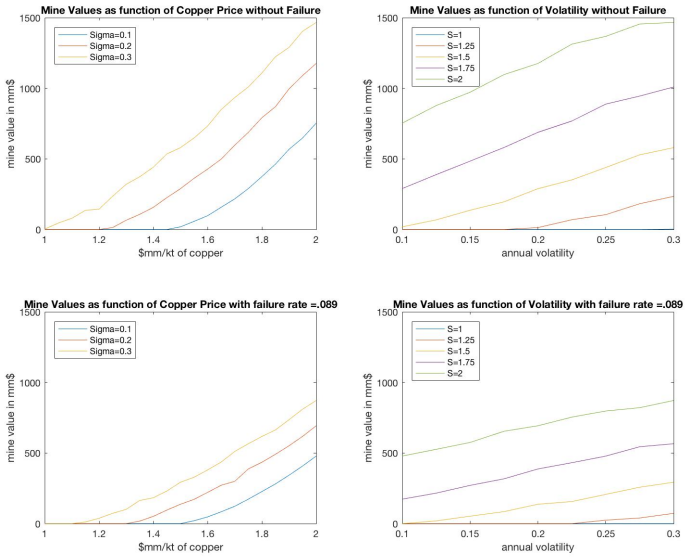

FIGURE 2.1. Price of Mine vs Spot and Volatility

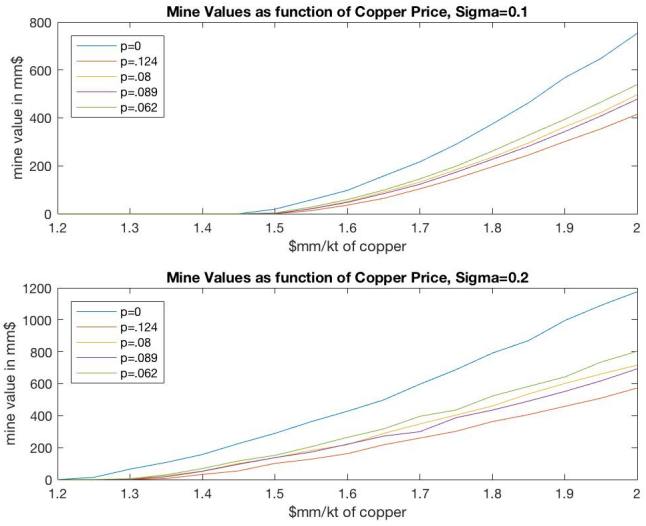

FIGURE 2.2. Comparison of Mine Prices with and Without Failures

as it ignores potential issues with trending and general temporal dependence within the time series at a mine site. We wish to emphasize, however, that this process can be viewed as a "black box" - if we know that the procedure that was actually used by the mine engineers to be estimate the failure threshold of the dam, then it is indeed appropriate to use that procedure to extract the

Year	Maximum	Year	Maximum	Year	Maximum	Year	Maximum
1900	1.69	1926	1.61	1952	1.59	1978	1.63
1901	1.66	1927	1.72	1953	1.82	1979	1.77
1902	1.69	1928	1.66	1954	1.63	1980	1.65
1903	1.62	1929	1.71	1955	1.69	1981	1.71
1904	1.58	1930	1.72	1956	1.74	1982	1.65
1905	1.74	1931	1.65	1957	1.69	1983	1.92
1906	1.56	1932	1.63	1958	1.73	1984	1.73
1907	1.73	1933	1.67	1959	1.67	1985	1.85
1908	1.75	1934	1.76	1960	1.57	1986	1.68
1909	1.62	1935	1.62	1961	1.71	1987	1.64
1910	1.71	1936	1.62	1962	1.78	1988	1.67
1911	1.60	1937	1.65	1963	1.61	1989	1.62
1912	1.64	1938	1.82	1964	1.63	1990	1.74
1913	1.57	1939	1.63	1965	1.64	1991	1.85
1914	1.49	1940	1.71	1966	1.82	1992	1.74
1915	1.72	1941	1.70	1967	1.72	1993	1.69
1916	1.53	1942	1.65	1968	1.75	1994	1.75
1917	1.71	1943	1.79	1969	1.65	1995	1.73
1918	1.64	1944	1.67	1970	1.66	1996	1.71
1919	1.72	1945	1.63	1971	1.61	1997	1.68
1920	1.64	1946	1.74	1972	1.66	1998	1.77
1921	1.67	1947	1.80	1973	1.60		
1922	1.67	1948	1.71	1974	1.70		
1923	1.62	1949	1.60	1975	1.71		
1924	1.63	1950	1.69	1976	1.78		
1925	1.68	1951	1.64	1977	1.62		

TABLE 2.24. Rainfall Data

(appropriately de-trended) stationary values, fit the reference distribution, and calculate the Renyi divergence from the (de-trended) empirical data. The simulations used to price the mine can then be performed using a corrected simulation process that will incorporate the temporal dependence in the failure process by increasing the location parameter for each time period. However, in the case that the threshold was estimated using a shorter time series, it is likely that trending and temporal dependence were not appropriately accounted for. In order to demonstrate the efficacy of even the naive robustification procedure (without accounting for trending), we conduct the following experiment:

Mine	Robust Probability	Trending Probability
Alumbrera	.0874	.0005
Bagdad	.0366	.0163
Bolivar	.0649	.0752
Centinela Oxide	.0451	.0371
Chapada	.0566	.0558
Chino	.0401	.0220
Copper Mountain	.1503	.0054
Cozamin	.0603	.0532
El Abra	.1260	.0094
Gibralter	.0502	.0214
Hucklberry	.0400	.0377
Los Pelambres	.0348	.0279
Michilla	.0465	.0008
Milpillas	.0297	.1365
Minto	.1102	.0035
Mount Polley	.1425	.0033
Piedras Verdes	.0516	.0170
Robinson	.0994	.0137
Safford	.0581	.0199
Sierrita	.0471	.0255
Tyrone	.0701	.022
Zaldivar	.0962	0

TABLE 2.25. Comparison of Robust Failure Probabilities with Trending Failure Probabilities

(1) For the 22 mine sites used in Chapter 4, we use 30 years of data to fit a naive model and calculate a failure threshold from the 99% quantile of the reference distribution, and a robust (annualized) probability of failure.

(2) For the same sites, we look at a 60 year time series of data. We de-trend the same data set by fitting a regression model of the precipitation maxima against time for each mine site, and subtract the corresponding expected value from each point to create a stationary series. We then use maximum likelihood to estimate parameters for a GEV distribution for each mine site.

(3) Incorporating this calibration, and the corresponding trend, we calculate the probability of failure over the life of a 20 year mine, and then convert this probability of failure to an annualized probability. See Table 2.25 for a comparison. In the vast majority of cases, the robust probability exceeds this level.

CHAPTER 3

Estimation of Sparse Covariance Matrices For Max-Stable Fields

3.1. Introduction

3.1.1. Motivation. For pricing and hedging purposes, estimating disaster risk at only a particular location may be sufficient to price using the real option model discussed in Chapter 2. In order to understand the complete risk profile of a particular portfolio of mines, it is necessary to estimate the joint distribution of the extremes of climate events for the whole portfolio of mines. When extremal climate events occur, there is a great deal of a probability of contagion, particularly within a specific region. It is known that the correlation structure within the tails of climate-driven processes is much different than that of the regular, daily observations [58]. Of course, by the nature of extremes, the number of observations will be very low, which will be an issue when the dimension of the problem grows; that is, when the size of the portfolio is large, the number of correlation parameters that needs to be estimated will be far greater than the number of observations. Fortunately, recognizing this, many reasonably large operators of mines will have some form of geographic diversification within their portfolio of mines; but owing to the advantages of having familiarity with the regulatory structure, geology, and working conditions in a particular area, it is reasonable to assume that there will be some "blocks" within the portfolio of mines in similar locations. This leads to essentially a block structure (modulo a permutation) for the covariance matrix of precipitation across mines, enforcing a natural sparsity constraint that we can exploit to efficiently estimate the underlying covariance structure of the maxima.

In general, even when the dimensionality of the problem is not too large, the problem of estimating a parametric form of max-stable is extremely difficult, owing in part to the aforementioned scarcity of data due to the nature of extreme events, and also due to the general intractability of the likelihood. For an arbitrary model, in order to evaluate the likelihood, it is necessary to compute a sum over all the partitions of the set $\{1, ..., d\}$ where d is the dimension of the data. The cardinality of this set grows super-exponentially, and is currently completely intractable for dimension greater

than 12 or 13. For a more complete discussion of this, see [23, 22, 13]. The authors of [23] suggest a stochastic EM algorithm approach that overcomes this problem in certain cases; however, for a broad class of models, the lack of a closed-form (or similarly tractable) solution for arbitrary higher-order partial derivatives of the exponent measure make this methodology impractical to implement. For example, it is ideal for a multivariate logistic max-stable distribution, but for models with an underlying normal process, it is ill-suited. Moreover, we are not able to exploit the natural sparsity found in most realistic portfolio compositions.

The most common approach to fitting a max-stable vector is the pairwise likelihood approach, proposed by [51], which is done by combining the bivariate densities of specific pairs of observations. Although this method is computationally tenable and has many attractive properties of the MLE (namely consistency), however, there is a severe loss of efficiency that becomes more apparent in higher dimensions [39]. A number of more efficient methods, triplewise and higher-order composite likelihoods are investigated by [29, 39, 38, 22], but the trade-off between computational efficiency and statistical efficiency is poor, and the correct method of choosing higher-order components is unclear.

3.1.2. Formulation of the Problem. We consider a sequence of independent p-dimensional Gaussian random vectors (X_n), with mean 0 and covariance Σ_{COV}, and a renewal sequence $A_n = \tau_1 + ... + \tau_n$, $n \geq 1$, where (τ_i) is an i.i.d. sequence of exponential random variables with expectation 1, independent of (X_n). The X_n can be viewed as realizations of a Gaussian random field, sampled at a discrete set of p points. Under certain conditions, we will provide a method of estimating the Σ for realizations of the random variable M in $\mathbb{R}^p$, where, the i-th entry of M, namely $M(i)$, satisfies:

$$M(i) = \sup_{n \geq 1} \{ - \log A_n + X_n(i) \}$$

when the dimension p of X and M, is high. The vector M is max-stable in the sense of [19]. This process has Gumbel marginals and given that is has an actual covariance structure (unlike its heavy-tailed analogues), it will provide a method to use concepts such as sparsity in our estimates.

In general, the link function between the covariance matrix of the underlying normal field and the corresponding max-stable field is unknown. Moreover, since sample data will consist of block maxima, data sets will typically consist of a relatively small number of observations. As the

dimension of the data grows, we potentially run into issues with overdetermination; that is, the number of parameters being estimated approaches or exceeds the number of observations. We are able to remedy these two issues as follows:

(1) Using the exact simulation technique for max-stable fields developed by [45], we are able to estimate the relationship between the correlation of the underlying normal field and the resulting max-stable field.

(2) Once the correlation matrix of the underlying process has been estimated, we are then able to use techniques from correlation matrix estimation, like those in [44, 62], which enforce sparsity in the correlation matrix but ensure that the estimated matrix is positive-definite.

3.2. Notation and Preliminaries

We will denote the covariance matrix of X as Σ_{COV}, the correlation matrix of X as Σ, and the correlation matrix of M as Σ_M.

For each pair of coordinates, the correlation of between $M(i)$ and $M(j)$ will be determined entirely by the correlation between $X(i)$ and $X(j)$. We define the function c_M as the function that determines this transformation. That is, if ρ is the correlation between $X(i)$ and $X(j)$, then $c_M(\rho)$ is the correlation between $M(i)$ and $M(j)$. We define the corresponding transformation for correlation matrices as $C_M(\cdot)$, where $C_M(\Sigma)$ is the matrix with $c_M(\Sigma(i,j))$ on all the off-diagonal elements (the diagonal elements remain 1, clearly). Since we do not have a closed-form solution for this, we produce an empirical approximation to the function $c_M(\rho)$ using the exact simulation algorithm of [45]. We denote by $c_M^{-1}(\cdot)$ the inverse of $c_M(\cdot)$, and $C_M^{-1}(\cdot)$ is $c_M^{-1}(\cdot)$ applied component-wise on the off-diagonal elements of a matrix. Since this algorithm can be run uniformly in ρ, we can use the same random numbers to generate samples for every value of ρ. For practical purposes, this was done on a grid and gaps were filled in with piecewise linear interpolation. See Figure 3.1 for a graphical representation of this function.

We also make use of the follow notation: for a matrix A, $\Lambda_{min}(A)$ and $\Lambda_{max}(A)$ are the smallest and largest eigenvalues of A, respectively. We let $\|A\|_1 = \sum_{i,j} |A(i,j)|$, $\|A\|_2^2 = \Lambda_{max}\left(A^T A\right)$, $\|A\|_F^2 = \sum_{j,k} A(j,k)^2$. We define $\|A\|_{1,\text{off}}$ as the ℓ_1-norm on the equivalence classes of matrices with the same off-diagonal elements, namely: $\|A\|_{1,\text{off}} = \sum_{i \neq j} |A(i,j)|$. Likewise, we use $\|A\|_{\infty,\text{off}}$

as the $\ell - \infty$ normal on the equivalence classes of matrices with the same off-diagonal elements, that is $\|A\|_{\infty,\text{off}} = \max_{i \neq j} |A(i,j)|$. We use S_M to denote the sample correlation matrix of the random variables M. Finally, we let $H : \mathbb{R}^{p \times p} \to \mathbb{R}^{p \times p}$ be defined as the function that returns the matrix with the coordinates:

$$H(A)(i,j) = A(i,j) 1_{\{A(i,j)>0\}}.$$

We consider two estimators of sparse correlation matrices. First, we consider the naive problem, the Soft-Thresholding Operator (STO), proposed by [56], given a correlation matrix S with ℓ_1 penalty:

$$\hat{\Sigma}^{STO}(S,\lambda) = \text{argmin}_{\Sigma} \frac{1}{2} \|S - \Sigma\|_F^2 + \lambda \|\Sigma\|_{1,\text{off}} \tag{3.2.1}$$

which has the closed-form solution:

$$\hat{\Sigma}^{STO}(S,\lambda) = \begin{cases} \text{sign}(S(j,k)) \cdot \max\{|S(j,k)| - \lambda, 0\} & , j \neq k \\ S(j,k) & \text{otherwise} \end{cases}.$$

While simple and efficient, it contains no guarantee of positive-definiteness.

We next consider the EC2 estimator of [44], which is an expansion of the STO method by adding a positive eigenvalue constraint.

$$\begin{aligned} \hat{\Sigma}^{EC2}(S,\lambda,\tau) = \text{argmin}_{\Sigma(j,j)=1} & \frac{1}{2} \|S - \Sigma\|_F^2 + \lambda \|\Sigma\|_{1,\text{off}} \\ \text{s.t. } & \tau \leq \Lambda_{min}(\Sigma). \end{aligned} \tag{3.2.2}$$

This is solved algorithmically using an iterative procedure. See the Appendix for details.

3.3. Estimating Σ for the max-stable vector

Let $M_1, ...M_N$ be a set of sample data from the category of max-stable vectors introduced in the introduction. Let S_M be the sample correlation matrix of the M_i.

We note that the problem of estimating the correlation matrix Σ is equivalent to estimating the covariance matrix Σ_{COV}, as we can standardize the data to match the mean and variance of a standard Gumbel distribution. We present here three estimators that are asymptotically consistent and efficiently estimate the correlation structure of the underling Gaussian distribution.

Estimator 1: STO for Max-Stable Correlations (STO-MS). The most basic estimator that we consider is a modified version of the soft thresholding estimator. We consider the following optimization:

$$\hat{\Sigma}^{STO-MS}\left(S_M, \lambda\right) = \text{argmin}_{\Sigma(j,j)=1} \frac{1}{2} \left\| C_M(\Sigma) - S_M \right\|_F^2 + \lambda \left\| C_M(\Sigma) - C_M(I) \right\|_{1,\text{off}}$$

which has the following closed-form solution:

$$\hat{\Sigma}^{STO-MS}\left(S_M, \lambda\right) = \begin{cases} c_M^{-1}\left(\text{sign}\left(S_M\left(j,k\right) - C_M(0)\right) \cdot \max\left\{\left|S_M(j,k)\right| - \lambda, 0\right\}\right) & , j \neq k \\ 1 & \text{otherwise} \end{cases}.$$

Estimator 2: STO for Max-Stable Correlations in Normal Space (STO-NS). The next estimator we consider is soft thresholding performed in the space of normal correlations:

$$\hat{\Sigma}^{STO-MSN}\left(S_M, \lambda\right) = \text{argmin}_{\Sigma(j,j)=1} \frac{1}{2} \left\| \Sigma - C_M^{-1}(S_M) \right\|_F^2 + \lambda \left\| \Sigma \right\|_{1,\text{off}}$$

which has the following closed-form solution:

$$\hat{\Sigma}^{STO-MSN} S_M, \lambda) = \begin{cases} \text{sign}(c_M^{-1}(S_M(j,k))) \cdot \max\{|c_M^{-1}(S_M(j,k))| - \lambda, 0\} & , j \neq k \\ 1 & \text{otherwise} \end{cases}.$$

Estimator 3: STO with Two Sided Shrinkage (STO-SN). The next estimator shrinks on both sides to account for the difference in behavior of the transformation function $c_M(\cdot)$ for positive and negative values:

$$\begin{aligned} \hat{\Sigma}^{STO-TS}(S_M, \lambda_1, \lambda_2) = \text{argmin}_{\Sigma(j,j)=1} \frac{1}{2} \left\| S_M - C_M(\Sigma) \right\|_F^2 + \lambda_1 \left\| H(C_M(\Sigma) - C_M(I)) \right\|_{1,\text{off}} \\ + \lambda_2 \left\| H(C_M(I) - C_M(\Sigma)) \right\|_{1,\text{off}} \end{aligned}$$

which has the closed-form solution:

$$\hat{\Sigma}^{STO-TS}(S_M,\lambda_1,\lambda_2)=\begin{cases} c_M^{-1}(\text{sign}(S_M(j,k)-c_M(0)) & \\ \quad \cdot \max\{|S_M(j,k)|-\lambda_1,0\})1_{\{S_{ij}>c_m(0)\}} & \\ \quad +c_M^{-1}(\text{sign}(S_M(j,k)-c_M(0)) & \\ \quad \cdot \max\{|S_M(j,k)|-\lambda_2,0\})1_{\{S_{ij}<c_m(0)\}} & ,j\neq k \\ 1 & \text{otherwise} \end{cases}.$$

Estimator 4: EC2 With Transformed Correlation in Normal Space (EC2-NS). Our first method inspired by the EC2 estimator [44] is simply to apply the EC2 estimator on the transformation $C_M^{-1}(S_M)$ of the sample covariance matrix S_M into the space of normal coordinates:

$$\hat{\Sigma}^{EC2-NS}(S_M,\lambda,\tau)=\text{argmin}_{\Sigma(j,j)=1}\frac{1}{2}\left\|\Sigma-C_M^{-1}(S_M)\right\|_F^2+\lambda\left\|\Sigma\right\|_{1,\text{off}}$$
$$\text{s.t. } \tau\leq\Lambda_{min}(\Sigma) \tag{3.3.1}$$

where $\lambda>0$ is a regularization parameter, and $\tau>0$ is the minimal eigenvalue, which will guarantee the positive definiteness of the solution. The drawback of the EC2-NS estimator is that since it penalizes differences from zero in normal space, shrinkage is much more sharp for negative correlations, that is, when shrinkage is applied in the space of normal correlations, which may not be a desired property.

Estimator 5: EC2 on Transformed Correlation in Max-Stable Space (EC2-MS). The next method considers shrinkage (and penalties) in the space of the max-stable coordinates, namely, we solve the following minimization problem:

$$\hat{\Sigma}^{EC2-MS}(S_M,\lambda,\tau)=\text{argmin}_{\Sigma(j,j)=1}\frac{1}{2}\left\|C_M(\Sigma)-S_M\right\|_F^2+\lambda\left\|C_M(\Sigma)-C_M(I)\right\|_{1,\text{off}}$$
$$\text{s.t. } \tau\leq\Lambda_{min}(C_M(\Sigma)) \tag{3.3.2}$$

While this estimator applies shrinkage in the space of the actual statistical error, that is, the error in the correlations of the max-stable field, the shrinkage now cannot account differences in the shrinkage that occurs the space of normal correlations. Moreover, there is no guarantee that the estimator $\hat{\Sigma}$ is positive-definite.

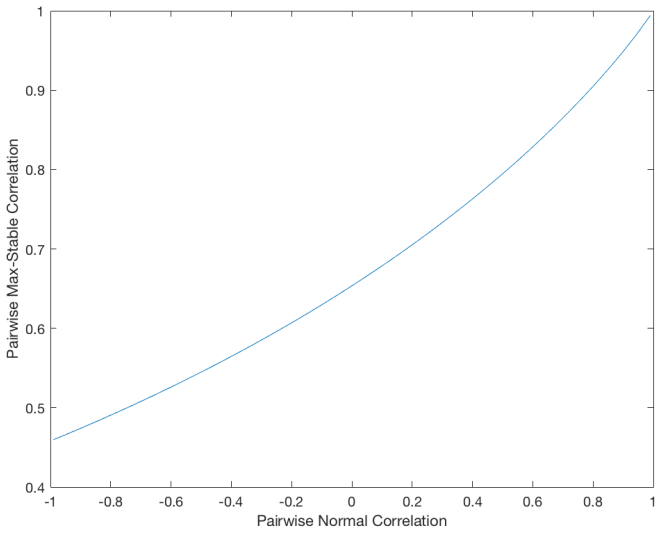

FIGURE 3.1. Normal Correlations vs Max-Stable Correlations

Estimator 6: EC2 with Separate Shrinkage and Normal Eigenvalue Constraint (EC2-SN). The next method corrects for the issues of the two previous estimators. Shrinkage is applied in the space of max-stable correlations, with different shrinkage rates being applied to correlations that correspond to positive and negative correlations in the space of normal correlations. Moreover, the eigenvalue constraint is applied to the normal correlation matrix, to ensure that the estimated matrix of interest is positive-definite:

$$\begin{aligned}\hat{\Sigma}^{EC2-SN}(S_M, \lambda_1, \lambda_2, \tau) = \text{argmin}_{\Sigma(j,j)=1} \frac{1}{2} \|S_M - C_M(\Sigma)\|_F^2 + \lambda_1 \|H(C_M(\Sigma) - C_M(I))\|_{1,\text{off}} \\ + \lambda_2 \|H(C_M(I) - C_M(\Sigma))\|_{1,\text{off}} \\ \text{s.t. } \tau \leq \Lambda_{min}(\Sigma). \end{aligned} \tag{3.3.3}$$

As we shall see, the performance of the EC2SN dominates the performance of the other estimators discussed.

3.4. Theoretical Results

3.4.1. Results for EC2-MS Estimator. We denote by $\mathcal{U}_{\tau,C_M(I)}(q, c_0(p), K)$ class of matrices such that

$$\left\{ \sigma : \Sigma : |\sigma(i,j)| \leq K \text{ for all } i \neq j, \max_i \sum_{j=1}^{p} |\sigma(i,j) - c_M(0)|^q \leq c_0(p) \right\}$$

as in [4]. See the Appendix for more details.

Let M be the max stable field generated from a sequence of i.i.d. Gaussian fields $\tilde{Z}_n$ with mean 0 and variance 1 everywhere, and renewal processes A_n, such that

$$M(t) = \max_{n \geq 1} \left\{ -\log(A_n) + \tilde{Z}_n(t) \right\}.$$

And we consider sampling this at two particular points, denoted $M(1)$ and $M(2)$, where $E[\tilde{Z}_n(1)\tilde{Z}_n(2)] = \rho$, where $\rho \in [-1+\epsilon, 1-\epsilon]$, where $\epsilon > 0$. We can therefore represent $\tilde{Z}_n(1)$ and $\tilde{Z}_n(2)$ as

$$\begin{aligned} \tilde{Z}_n(1) &= Z_n(1) \\ \tilde{Z}_n(2) &= \rho Z_n(1) + \sqrt{1-\rho^2} Z_n(2) \end{aligned}$$

where $Z_n(1)$ and $Z_n(2)$ are independent sequences of i.i.id standard normal random variables.

We introduce some further notation:

$$n^\star = \text{argmax}_{n=1}^{N} \left[-\log(A_n) + \rho Z_n(1) + \sqrt{1-\rho^2} Z_n(2) \right] \tag{3.4.1}$$

$$f(\rho) = \left(1, \frac{\rho}{(1-\rho^2)^{1/2}} \right)^T \tag{3.4.2}$$

$$\begin{aligned} g(Z_n, \rho) &= Z_n(1) + \frac{\rho}{(1-\rho^2)^{\frac{1}{2}}} Z_n(2) \\ &= f(\rho)^T \cdot Z_n \end{aligned} \tag{3.4.3}$$

where N is a finite number such that $M(2) = \max_{n=1}^{N} [-\log(A_n) + \rho Z_n(1) + \sqrt{1-\rho^2} Z_n(2)]$ which is independent of ρ and can be generated as in [45].

LEMMA 1. *Let $M(1)$ and $M(2)$ be samples from a max-stable field as defined above. Let $\rho \in [-1+\epsilon, 1-\epsilon]$ be the correlation of the processes $\tilde{Z}_n(1)$ and $\tilde{Z}_n(2)$ which generate $M(1)$ and $M(2)$. Then $E[\frac{\partial}{\partial \rho} M(1)M(2)] = \frac{\partial}{\partial \rho} E[M(1)M(2)] < \infty$.*

PROOF. Now we consider the derivative of $M(1)M(2)$ with respect to ρ:

$$\frac{\partial}{\partial\rho}M(1)M(2) = \frac{\partial}{\partial\rho}\sum_{n=1}^{N} M(1)(-\log(A_n) + \rho Z_n(1) + \sqrt{1-\rho^2}Z_n(2)) \tag{3.4.4}$$

$$= M(1)\sum_{n=1}^{N} I(n = n^\star)g(Z_n, \rho) \tag{3.4.5}$$

where $n^\star$ and $g(Z_n, \rho)$ are as in (3.4.1) and (3.4.3), respectively. Since $Z_n(1)$ and $Z_n(2)$ are continuous random variables, there are no ties, almost surely. Next, we bound it from above using (3.4.5):

$$\begin{aligned}\left|\frac{\partial}{\partial\rho}M(1)M(2)\right| &\leq |M(1)|\sum_{i=1}^{N}|g(Z_n,\rho)| \\ &= |M(1)|\sum_{i=1}^{N}|f(\rho)^T\cdot Z_n|.\end{aligned}$$

Now, applying Cauchy-Schwarz, we get

$$\begin{aligned}|M(1)|\sum_{i=1}^{N}|f(p)^T\cdot Z_n| &\leq |M(1)|\sum_{i=1}^{N}\|f(\rho)\|_2\|Z_n\|_2 \\ &\leq |M(1)|\,\|f(1-\epsilon)\|_2\sum_{i=1}^{N}\|Z_n\|_2\,.\end{aligned}$$

Since N has finite moments of all orders [45, 6], it is elementary to see that $\sum_{i=1}^{N}\|Z_n\|_2$ likewise has finite moments of all orders. Therefore,

$$\left|\frac{\partial}{\partial\rho}M(1)M(2)\right|$$

is bounded above by the integrable function

$$|M(1)|\,\|f(1-\epsilon)\|_2\sum_{i=1}^{N}\|Z_n\|_2\,,$$

and therefore we can use the dominated convergence theorem to show:

$$\frac{\partial}{\partial\rho}E[M(1)M(2)] = E\left[\frac{\partial}{\partial\rho}M(1)M(2)\right].$$

□

COROLLARY 2. *For all $\epsilon > 0$, for $\rho \in [-1+\epsilon, 1-\epsilon]$, the function $c_M(\rho)$ is Lipschitz. Likewise, c_M^{-1} is Lipschitz for all $\rho \in [c_M(-1+\epsilon), c_M(1-\epsilon)]$.*

PROOF. Using Lemma 1 we have that

$$\begin{aligned}
\left|\frac{\partial c_M}{\partial \rho}\right| &= \left|\frac{\partial}{\partial \rho} E[M(1)M(2)]\right| \\
&= \left|E\left[\frac{\partial}{\partial \rho} M(1)M(2)\right]\right| \\
&\leq E[|M(1)| \, \|f(1-\epsilon)\|_2 \sum_{i=1}^{N} \|Z_n\|_2] \\
&\leq \|f(1-\epsilon)\|_2 \sqrt{E[M(1)^2] E[(\sum_{i=1}^{N} \|Z_n\|_2)^2]},
\end{aligned}$$

with the last inequality following from Cauchy-Schwarz. Once again, $E[M(1)^2]$ and $E[(\sum_{i=1}^{N} \|Z_n\|_2)^2]$ are finite, so the derivative of the function is bounded a compact set, and therefore Lipschitz. Likewise, since the function is invertible, and differentiable on the compact set $[-1+\epsilon, 1-\epsilon]$, the inverse will differentiable on a compact, and also Lipschitz. □

LEMMA 3. *For sample correlation matrix $\hat{\Sigma}$ with entries $\hat{\sigma}_{ij}$ of a max-stable vector, and true correlation matrix Σ (for the Gaussian generative process) with entries σ_{ij} we have the following:*

$$\begin{aligned}
(i) \max_i \sum_{j=1}^{p} &|\hat{\sigma}(i,j) - c_M(0)| \cdot 1(|\hat{\sigma}(i,j) - c_M(0)| \geq \lambda, |c_M(\sigma(i,j)) - c_M(0)| < \lambda) \\
&= O_P\left(c_0(p)\lambda^{-q}\left(\frac{p^{2/(1+\gamma)}}{n^{1/2}}\right) + c_0(p)\lambda^{1-q}\right) \\
(ii) \max_i \sum_{j=1}^{p} &|c_M(\sigma(i,j)) - c_M(0)| \cdot 1(|\hat{\sigma}(i,j) - c_M(0)| < \lambda, |c_M(\sigma(i,j)) - c_M(0)| \geq \lambda) \\
&= O_P\left(c_0(p)\lambda^{-q}\left(\frac{p^{2/(1+\gamma)}}{n^{1/2}}\right) + c_0(p)\lambda^{1-q}\right) \\
(iii) \max_i \sum_{j=1}^{p} &|\hat{\sigma}(i,j) - c_M(\sigma(i,j))| \cdot 1(|\hat{\sigma}(i,j) - c_M(0)| \geq \lambda, |c_M(\sigma(i,j)) - c_M(0)| \geq \lambda) \\
&= O_P\left(c_0(p)\lambda^{-q}\left(\frac{p^{2/(1+\gamma)}}{n^{1/2}}\right)\right)
\end{aligned}$$

$$(iv) P(\max_{i,j} |\hat{\sigma}(i,j) - c_M(\sigma(i,j))| > t) \leq p^2 K A(\gamma) \frac{n^{-(1+\gamma)/2}}{t^{1+\gamma}}.$$

PROOF. Since we have $E|M_i|^{2(1+\gamma)} \leq K$ for some $\gamma > 0$ for all i since all the M_i are Gumbel distributions, then by Markov's inequality, we have:

$$P(|\hat{\sigma}(i,j) - c_M(\sigma(i,j))| \geq t) \leq K A(\gamma) \frac{n^{-(1+\gamma)/2}}{t^{1+\gamma}}$$

$$P(\max_{i,j} |\hat{\sigma}(i,j) - c_M(\sigma(i,j))| \geq t) \leq p^2 K A(\gamma) \frac{n^{-(1+\gamma)/2}}{t^{1+\gamma}}$$

$$\max_{i,j} |\hat{\sigma}(i,j) - c_M(\sigma(i,j))| = O_P\left(\frac{p^{2/(1+\gamma)}}{n^{1/2}}\right)$$

when we choose $t_n = \frac{p^{2/(1+\gamma)}}{n^{1/2}}$. Using this result, the proofs of (i), (ii), and (iii) follow from substituting this inequality into the proofs of the analogous results of Gaussian random variables in [4], which we will reproduce here in. For the notational simplicity, we denote the bounded terms in (i), (ii), and (iii) as I, II, and III respectively.

First we consider (iii), and note that using (iv):

$$\begin{aligned} \text{III} &\leq \max_{i,j} |\hat{\sigma}(i,j) - c_M(\sigma(i,j))| \max_i \sum_{j=1}^{p} |c_M(\sigma(i,j)) - c_M(0)|^q \lambda^{-q} \\ &= O_P\left(c_0(p)\lambda^{-q}\left(\frac{p^{2/(1+\gamma)}}{n^{1/2}}\right)\right). \end{aligned}$$

To place a bound on I, we split it into two terms:

$$\begin{aligned} \text{I} &\leq \max_i \sum_{j=1}^{p} |\hat{\sigma}(i,j) - c_M(\sigma(i,j))| \cdot 1(|\hat{\sigma}_{ij} - c_M(0)| \geq \lambda, |\sigma_{ij} - c_M(0)| < \lambda) \\ &\quad + \max_i \sum_{j=1}^{p} |\sigma(i,j) - c_M(0)| \cdot 1(|\sigma(i,j) - c_M(0)| < \lambda) \\ &= \text{IV} + \text{V}. \end{aligned}$$

The second term V is bounded above:

$$V \leq \lambda^{1-q} c_0(p).$$

For IV, if we take $a \in (0,1)$:

$$\begin{aligned}
\text{IV} &\leq \max_i \sum |\hat{\sigma}(i,j) - c_M(\sigma(i,j))| \cdot 1(|\hat{\sigma}(i,j) - c_M(0)| \geq \lambda, |\hat{\sigma}(i,j) - c_M(0)| \leq a\lambda) \\
&\quad + \max_i \sum |\hat{\sigma}(i,j) - c_M(\sigma(i,j))| \cdot 1(|\hat{\sigma}(i,j) - c_M(0)| \geq \lambda, a\lambda < |\hat{\sigma}(i,j) - c_M(0)| \leq \lambda) \\
&\leq \max_{i,j} |\hat{\sigma}(i,j) - c_M(\sigma(i,j))| \max_i \sum_{j=1}^{p} 1(|\hat{\sigma}(i,j) - c_M(\sigma(i,j))| > (1-a)\lambda) \\
&\quad + c_0(p)(at)^{-q} \max_{i,j} |\hat{\sigma}(i,j) - c_M(\sigma(i,j)|.
\end{aligned}$$

We note then that:

$$\begin{aligned}
&P[\max_i \sum_{j=1}^{p} 1(|\hat{\sigma}(i,j) - c_M(\sigma(i,j))| > (1-a)\lambda) > 0] \\
&\quad = P(\max_{i,j} |\hat{\sigma}(i,j) - c_M(\sigma(i,j))| > (1-a)\lambda) \\
&\quad \leq p^2 K A(\gamma) \frac{n^{-(1+\gamma)/2}}{((1-a)\lambda)^{1+\gamma}}.
\end{aligned}$$

Then, if

$$2\log(p) - (1+\gamma)\log(\sqrt{n}\lambda) \to -\infty,$$

we have that

$$\text{IV} = O_P\left(c_0(p)\lambda^{-q}\left(\frac{p^{2/(1+\gamma)}}{n^{1/2}}\right)\right).$$

Combining IV and V, we are able to bound I:

$$\text{I} = O_P\left(c_0(p)\lambda^{-q}\left(\frac{p^{2/(1+\gamma)}}{n^{1/2}}\right) + c_0(p)\lambda^{1-q}\right),$$

Finally, from term (ii), we have that:

$$\begin{aligned}
\text{II} &\leq \max \sum_{j=1}^{p} (|\hat{\sigma}(i,j) - c_M(\sigma(i,j))| + |\hat{\sigma}(i,j) - c_M(0)|) \\
&\quad \cdot 1(|\hat{\sigma}(i,j) - c_M(0) < \lambda, |c_M(\sigma(i,j) - c_M(0) > \lambda) \\
&\leq \max_{i,j} |\hat{\sigma}(i,j) - c_m(\sigma(i,j)) \sum_{j=1}^{p} 1|\sigma(i,j) \geq \lambda)
\end{aligned}$$

$$+ \lambda \max_i \sum_{j=1}^{p} 1(|c_M(\sigma(i,j) - c_M(0)| \geq \lambda)$$
$$\leq O_P\left(c_0(p)\lambda^{-q}\left(\frac{p^{2/(1+\gamma)}}{n^{1/2}}\right) + c_0(p)\lambda^{1-q}\right).$$

□

THEOREM 4. *Suppose $M_1, ...M_n$ are a sample from a max-stable vector as above with mean vector 0 and correlation matrix Σ, where the correlation terms ρ_{ij} are bounded above by $(1-\epsilon)$. We define the estimators*

$$\hat{\Sigma}^{STO(M)} = argmin_{\Sigma_{jj=1}} \frac{1}{2} \|S_M - \Sigma\|_F^2 + \lambda \|\Sigma - C_M(I)\|_{1,off} \tag{3.4.6}$$
$$= C_M^{-1}(\hat{\Sigma}^{STO-MS}(S_M, \lambda)) \tag{3.4.7}$$

Then, uniformly on $\mathcal{U}_\tau(q, c_0(p), 1-\epsilon)$ for sufficiently small ϵ, if $\lambda_n = c\frac{p^{2/(1+\gamma)}}{n^{1/2}}$, then

$$\left\|\Sigma^{ST\hat{O}(M)}(S_M, \lambda_n) - C_M(\Sigma)\right\|_2 = O_p\left(c_o(p)\left(\frac{p^{2/(1+\gamma)}}{n^{1/2}}\right)^{1-q}\right)$$
$$\left\|\Sigma_C^{ST\hat{O}-MS}(S_M, \lambda_n) - \Sigma\right\|_2 = O_p\left(c_o(p)\left(\frac{p^{2/(1+\gamma)}}{n^{1/2}}\right)^{1-q}\right).$$

PROOF. This proof follows closely the approach of [56]. Denote the component-wise transformation of $\Sigma^{STO(M)}(S_M, \lambda)$ as $s_\lambda(z)$. Note that we have:

$$(i)|s_\lambda(z) - c_M(0)| \leq |z - c_M(0)|$$
$$(ii)s_\lambda(z) = 0 \text{ if } |z - c_M(0)| \leq \lambda$$
$$\text{(iii)}|s_\lambda(z) - z| \leq \lambda.$$

Also note that for the operator norm:

$$\|A\|_2 \leq \max_i \sum_j |a(i,j)|.$$

Consider the decomposition

$$\left\|\Sigma^{STO(M)}(S_M,\lambda)-C_M(\Sigma)\right\|_2 \le \left\|s_\lambda(C_M(\Sigma))-C_M(\Sigma)\right\|_2 + \left\|\Sigma^{STO(M)}(S_M,\lambda)-s_\lambda(C_M(\Sigma))\right\|_2. \tag{3.4.8}$$

For the first term in (3.4.8), we have that:

$$\begin{aligned}
\sum_{j=1}^{p}|s_\lambda(\sigma(i,j))-\sigma(i,j)| &\le \sum_{j=1}^{p}|\sigma(i,j)-c_M(0)|\cdot 1(|\sigma(i,j)-c_M(0)|\le\lambda)\\
&\quad+\lambda\sum_{j=1}^{p}1(|\sigma(i,j)-c_M(0)|>\lambda)\\
&=\sum_{j=1}^{p}|\sigma(i,j)-c_M(0)|^q|\sigma(i,j)-c_M(0)|^{1-q}\cdot 1(|\sigma(i,j)-c_M(0)|\le\lambda)\\
&\quad+\sum_{j=1}^{p}\lambda^q\lambda^{1-q}\cdot 1(|\sigma(i,j)-c_M(0)|>\lambda)\\
&\le \lambda^{1-q}\sum_{j=1}^{p}|\sigma(i,j)-c_M(0)|^q\\
&\le \lambda^{1-q}c_0(p).
\end{aligned}$$

We can decompose the second term in (3.4.8) using (i) and (ii):

$$\begin{aligned}
|s_\lambda(\hat\sigma(i,j))-s_\lambda(\sigma(i,j))| &\le |\hat\sigma(i,j)-c_M(0)|\\
&\quad\cdot 1(|\hat\sigma(i,j)-c_M(0)|\ge\lambda,|\sigma(i,j)-c_M(0)|<\lambda)\\
&\quad+|\sigma(i,j)-c_M(0)|\cdot 1(|\hat\sigma(i,j)-c_M(0)|<\lambda,|\sigma(i,j)-c_M(0)|<\lambda)\\
&\quad+\Big(|\hat\sigma(i,j)-\sigma(i,j)|\\
&\qquad+|s_\lambda(\hat\sigma(i,j))-\hat\sigma(i,j)|+|s_\lambda(\sigma(i,j))-\sigma(i,j)|\Big)\\
&\quad\cdot 1(|\hat\sigma(i,j)-c_M(0)|\ge\lambda,|\sigma(i,j)-c_M(0)|\ge\lambda).
\end{aligned}\tag{3.4.9}$$

The first four three terms of (3.4.9) are bounded using Lemma 3. For the fourth term, we apply property (iii) of the thresholding operator:

$$\begin{aligned}
&\max_i \sum_{j=1}^{p} |s_\lambda(\hat{\sigma}(i,j)) - \hat{\sigma}(i,j)| \cdot 1(|\hat{\sigma}(i,j) - c_M(0)| \geq \lambda, |\sigma(i,j) - c_M(0)| \geq \lambda) \\
&\leq \max_i \sum_{j=1}^{p} \lambda^q \lambda^{1-q} | \cdot 1(|\hat{\sigma}(i,j) - c_M(0)| \geq \lambda, |\sigma(i,j) - c_M(0)| \geq \lambda) \\
&\leq \lambda^{1-q} \max_i \sum_{j=1}^{p} |\sigma(i,j) - c_M(0)|^q \cdot 1(|\sigma(i,j) - c_M(0) \geq \lambda) \\
&\leq \lambda^{1-q} c_0(p).
\end{aligned}$$

By a similar argument, for the fifth term of (3.4.9):

$$\begin{aligned}
&\max_i \sum_{j=1}^{p} |s_\lambda(\sigma(i,j)) - \sigma(i,j)| \cdot 1(|\hat{\sigma}(i,j) - c_M(0)| \geq \lambda, |\sigma(i,j) - c_M(0)| \geq \lambda) \\
&\leq \lambda^{1-q} c_0(p).
\end{aligned}$$

Combining all terms, can see that:

$$\left\| \Sigma^{STO(M)}(\lambda) - C_M(\Sigma) \right\|_2 = O_p \left(c_0(p)(\lambda^{1-q} + \lambda^{-q} \frac{p^{2/(1+\gamma)}}{n^{1/2}} \right).$$

The first part of the theorem follows by substituting $\lambda_n = \frac{p^{2/(1+\gamma)}}{n^{1/2}}$. The second part of the theorem follows from the fact that C_M^{-1} is Lipschitz in the operator norm on the space $\mathcal{U}_\tau(q, c_0(p), 1-\epsilon)$ as in Lemma 2. □

THEOREM 5. *Suppose $M_1, ...M_n$ are a sample from a max-stable vector as above with mean vector μ and correlation matrix Σ, where the correlation terms ρ_{ij} are bounded above by $(1-\epsilon)$. Let δ_{min} be the minimal eigenvalue of $C_M(\Sigma)$. Let $\Sigma^{EC2-MS}(S_M, \lambda, \tau)$ be the estimator defined in (3.3.2). Let N_d be the of nonzero elements in Σ. Then with $\lambda = c_3(\frac{p^{2/(1+\gamma)}}{n^{1/2}})^{1-q}$, and $\tau \leq \delta_{min}$, we have that:*

$$\left\| \hat{\Sigma}^{EC2-MS}(S_M, \lambda, \tau) - C_M(\Sigma) \right\|_F = O_P \left(c_0(p) \sqrt{N_d} \left(\frac{p^{2/(1+\gamma)}}{n^{1/2}} \right)^{1-q} \right)$$

PROOF. The proof follows closely to that of Theorem 4.3 in [44]. To simplify the notation, we refer to $\hat{\Sigma}^{EC2-MS}(S_M, \lambda, \tau)$ as $\hat{\Sigma}$ throughout the proof.

Since $\tau \leq \delta_{min}$, then $C_M(\Sigma)$ is a feasible solution to (3.3.2), so we have

$$\frac{1}{2}\left\|S - \hat{\Sigma}\right\|_F^2 + \lambda\left\|\hat{\Sigma} - C_M(I)\right\|_{1,\text{off}} \leq \frac{1}{2}\left\|S - C_M(\Sigma)\right\|_F^2 + \lambda\left\|C_M(\Sigma) - C_M(I)\right\|_{1,\text{off}}.$$

Letting $\Delta = \hat{\Sigma} - C_M(\Sigma)$, after simple manipulation we have:

$$\begin{aligned} 0 \geq{}& \frac{1}{2}\left\|\Delta\right\|_F^2 - \left\langle S - \Sigma, \Delta\right\rangle \\ &+ \lambda\left\|C_M(\Sigma) + \Delta - C_M(I)\right\|_{1,\text{off}} - \lambda\left\|C_M(\Sigma) - C_M(I)\right\|_{1,\text{off}} \end{aligned} \tag{3.4.10}$$

where $\left\langle \cdot, \cdot \right\rangle$ denotes the inner product associated with the Froebenius norm. Let $\mathcal{E}$ be the set of all matrices $\mathbb{R}^{dxd}$ that has zeroes on the same entries as the zeros of Σ. Let $\mathcal{E}_\perp = \mathbb{R}^{dxd} \backslash \mathcal{E}$. We denote $A_\mathcal{E}$ as the projection of matrix A onto $\mathcal{E}$, and $A_{\mathcal{E}^\perp} = A - A_\mathcal{E}$. Then we have:

$$\begin{aligned} \left\|C_M(\Sigma) + \Delta - C_M(I)\right\|_{1,\text{off}} &= \left\|C_M(\Sigma_\mathcal{E}) + C_M(\Sigma_{\mathcal{E}^\perp}) + \Delta_\mathcal{E} + \Delta_{\mathcal{E}^\perp} - C_M(I)\right\|_{1,\text{off}} \\ &\geq \left\|C_M(\Sigma_\mathcal{E}) - C_M(I)\right\|_{1,\text{off}} + \left\|\Delta_{\mathcal{E}^\perp}\right\|_{1,\text{off}} - \left\|\Delta_\mathcal{E}\right\|_{1,\text{off}} \\ &\quad - \left\|C_M(\Sigma_{\mathcal{E}^\perp}) - C_M(I)\right\|_{1,\text{off}} - \left\|C_M(I)\right\|_{1,\text{off}}. \end{aligned}$$

Subtracting from $\left\|C_M(\Sigma_\mathcal{E}) - C_M(I)\right\|_{1,\text{off}}$, we can see that:

$$\begin{aligned} &\left\|C_M(\Sigma) + \Delta - C_M(I)\right\|_{1,\text{off}} - \left\|C_M(\Sigma) - C_M(I)\right\|_{1,\text{off}} \\ &\quad \geq \left\|\Delta_{\mathcal{E}^\perp}\right\|_{1,\text{off}} - \left\|\Delta_\mathcal{E}\right\|_{1,\text{off}} - 2\left\|C_M(\Sigma_{\mathcal{E}^\perp}) - C_M(I)\right\|_{1,\text{off}} - \left\|C_M(I)\right\|_{1,\text{off}}. \end{aligned} \tag{3.4.11}$$

By the Cauchy-Schwarz Inequality, it follows that:

$$\left\langle S - C_M(\Sigma), \Delta\right\rangle \leq \left\|S - C_M(\Sigma)\right\|_{\infty,\text{off}}\left\|\Delta\right\|_{1,\text{off}}. \tag{3.4.12}$$

In the event that $\left\|S - C_M(\Sigma)\right\|_{\infty,\text{off}} \leq \lambda/2$, we get by combining (3.4.10), (3.4.11), and (3.4.12):

$$\begin{aligned} 0 \geq{}& \frac{1}{2}\left\|\Delta\right\|_F^2 - \frac{\lambda}{2}(\left\|\Delta_\mathcal{E}\right\|_{1,\text{off}} + \left\|\Delta_{\mathcal{E}^\perp}\right\|) + \lambda(\left\|\Delta_{\mathcal{E}^\perp}\right\|_{1,\text{off}} - \left\|\Delta_\mathcal{E}\right\|_{1,\text{off}} \\ &- 2\left\|C_M(\Sigma_{\mathcal{E}^\perp}) - C_M(I)\right\|_{1,\text{off}} - \left\|C_M(I)\right\|_{1,\text{off}}) \end{aligned}$$

$$= \frac{1}{2}\|\Delta\|_F^2 - \frac{3\lambda}{2}\|\Delta_{\mathcal{E}}\|_{1,\text{off}} + \frac{\lambda}{2}\|\Delta_{\mathcal{E}^\perp}\|_{1,\text{off}}$$
$$- 2\lambda(\|C_M(\Sigma_{\mathcal{E}^\perp}) - C_M(I)\|_{1,\text{off}}) - \lambda\|C_M(I)\|_{1,\text{off}}. \tag{3.4.13}$$

Since $\|\Delta\|_F^2 > 0$, we have

$$\|\Delta_{\mathcal{E}^\perp}\|_{1,\text{off}} - \|C_M(I)\|_{1,\text{off}} \leq 3\|\Delta_{\mathcal{E}}\|_{1,\text{off}} + 4\|C_M(\Sigma_{\mathcal{E}^\perp}) - C_M(I)\|_{1,\text{off}}.$$

Combining this with (3.4.13), we have that:

$$\|\Delta\|_F^2 \leq 3\lambda\|\Delta_{\mathcal{E}}\|_{1,\text{off}} + 4\lambda\|C_M(\Sigma_{\mathcal{E}^\perp}) - C_M(I)\|_{1,\text{off}} \leq 3\lambda\sqrt{N_D}\|\Delta_{\mathcal{E}}\|_F.$$

And therefore:

$$\|\Delta\|_F \leq \frac{\|\Delta\|_F^2}{\|\Delta_{\mathcal{E}}\|_F} \leq 3\lambda\sqrt{N_D}.$$

Since $\|S - C_M(\Sigma)\|_{\infty,off}$ is $O_P(c_o(p)(\frac{p^{2/(1+\gamma)}}{n^{1/2}})^{1-q})$, it follows that

$$\|\Delta\|_F = \left\|\hat{\Sigma} - C_M(\Sigma)\right\|_F = O_P\left(c_o(p)(\frac{p^{2/(1+\gamma)}}{n^{1/2}})^{1-q}\right).$$

. □

3.4.2. Results for EC2-MS Estimator. The results for follow from a basic modification of Lemma 3.

LEMMA 6. *For sample correlation matrix S_M with entries $\hat{\sigma}(i,j)$, and underlying correlation matrix Σ with entries σ_{ij} we have the following:*

$$(i)\max_i \sum_{j=1}^{p} |c_M^{-1}(\hat{\sigma}(i,j))|1(|c_M^{-1}(\hat{\sigma}(i,j))| \geq \lambda, |\sigma(i,j)| < \lambda)$$
$$= O_P(c_0(p)\lambda^{-q}(\frac{p^{2/(1+\gamma)}}{n^{1/2}}) + c_0(p)\lambda^{1-q})$$
$$(ii)\max_i \sum_{j=1}^{p} |c_M^{-1}(\hat{\sigma}(i,j))|1(|c_M^{-1}(\hat{\sigma}(i,j))| < \lambda, |\sigma(i,j)| \geq \lambda)$$
$$= O_P(c_0(p)\lambda^{-q}(\frac{p^{2/(1+\gamma)}}{n^{1/2}}) + c_0(p)\lambda^{1-q})$$

$$(iii) \max_i \sum_{j=1}^{p} |c_M^{-1}(\hat{\sigma}(i,j))| 1(|c_M^{-1}(\hat{\sigma}(i,j))| \geq \lambda, |\sigma(i,j)| \geq \lambda)$$
$$= O_P(c_0(p)\lambda^{-q}(\frac{p^{2/(1+\gamma)}}{n^{1/2}}))$$
$$(iv) P(\max_{i,j} |c_M^{-1}(\hat{\sigma}(i,j)) - \sigma(i,j)| > t) \leq p^2 KC(\gamma)\frac{n^{-(1+\gamma)/2}}{t^{1+\gamma}}.$$

PROOF. These all follow from applying the Lipschitz property of $c_M^{-1}(\cdot)$ from Corollary 2 to the results in Lemma 3.

We present analogous results for the $EC2-NS$ estimator: □

THEOREM 7. *Suppose $M_1, ...M_n$ are a sample from a max-stable vector as above with mean vector 0 and correlation matrix Σ, where the correlation terms $\Sigma(i,j)$ are bounded such that $|\Sigma(i,j)| < (1-\epsilon)$ for some $\epsilon > 0$. We consider the estimator $\hat{\Sigma}^{STO-NS}$ defined in (3.4.7). Then, uniformly on $\mathcal{U}_\tau(q, c_0(p), 1-\epsilon)$ for sufficiently small ϵ, if $\lambda_n = c\frac{p^{2/(1+\gamma)}}{n^{1/2}}$, then*

$$\left\|\hat{\Sigma}^{STO-NS}(\lambda_n) - \Sigma\right\|_2 = O_p\left(c_o(p)(\frac{p^{2/(1+\gamma)}}{n^{1/2}})^{1-q}\right).$$

THEOREM 8. *Suppose $M_1, ...M_n$ are a sample from a max-stable vector as above with mean vector 0 and correlation matrix Σ, where the correlation terms $\Sigma(i,j)$ are bounded bounded such that $|\Sigma(i,j)| < (1-\epsilon)$ for some $\epsilon > 0$. Let δ_{min} be the minimal eigenvalue of Σ. Let Σ_C^{EC2-NS} be the estimator defined in (3.3.1). Let N_d be the number of nonzero elements in Σ. Then with $\lambda = c_3(\frac{p^{2/(1+\gamma)}}{n^{1/2}})^{1-q}$, and $\tau \leq \delta_{min}$, we have that:*

$$\left\|\Sigma_C^{EC2-NS} - \Sigma\right\|_F = O_P\left(c_0(p)\sqrt{N_d}(\frac{p^{2/(1+\gamma)}}{n^{1/2}})^{1-q}\right).$$

The proofs of Theorems 7 and 8 are identical to the proofs of Theorems 4 and 5.

3.5. Numerical Examples

Before we demonstrate any theoretical results for our estimators, we conduct a series of numerical experiments (using simulation) to show the performance of our algorithms compared to STO operators and naive correlation matrices. Note that the estimators can be calculated using a

version of the Iterative Soft-Thresholding and Projection (ISP) algorithm designed by [44]. See the appendix for details. We compared the following 5 estimators:

- EC2 with Max-Stable Correlations in Normal Space (EC2-NS): $\hat{\Sigma}^{EC2-NS}(S_M, \lambda, \tau)$
- EC2 on Transformed Correlation in Max-Stable Space (EC2-SN): $\hat{\Sigma}^{EC2-MS}(S_M, \lambda, \tau)$
- EC2 with Separate Shrinkage and Normal Eigenvalue Constraint (EC2-SN): $\hat{\Sigma}^{EC2-SN}(S_M, \lambda_1, \lambda_2, \tau)$
- STO for Max-Stable Correlations: $\hat{\Sigma}^{STO-MS}(S_M, \lambda)$
- The naive correlation estimator: $C_M^{-1}(S_M)$

We used the following 3 models for the underlying correlation matrix:

- Toeplitz Matrix: $\Sigma(j,k) = 0.75^{|j-k|}$
- Block Matrix: We split the matrix evenly into 10 groups, where $\Sigma(j,k) = 0.8$ if they belong to the same group, and 0 otherwise.
- Banded Matrix: We split the matrix evenly into 10 groups, where the correlation was

$$\Sigma(j,k) = \max\left\{1 - \frac{|j-k|}{d/10}, 0\right\}.$$

Using these covariance models, for $d = 50, 100, 200, 400$ we generate a sample of size 110 from the corresponding max-stable vector using the exact sampling algorithm [45]. We re-divide this sample $K = 25$ times into a training set of size 82 and a test set of 28. For each soft-thresholding operator (corresponding to an EC2 estimator), we choose the optimal value of λ:

$$\hat{\lambda} = \text{argmin} \sum_{i=1}^{K} \left\| \hat{\Sigma}_\lambda^k - S^k \right\|_F^2$$

where $\hat{\Sigma}_\lambda^k$ is the STO estimator from the $k-th$ training set, using tuning parameter λ, and S^k is the correlation matrix from the $k-th$ training set. For each of the EC2 estimators, we use the estimate of λ for the corresponding STO operator, and select τ as the following:

$$\hat{\tau} = \text{argmin} \sum_{i=1}^{K} \left\| \hat{\Sigma}_\tau^k - S^k \right\|_F^2$$

where $\hat{\Sigma}_\tau^k$ is the Σ^{EC2} estimator from the kth training set using $\hat{\lambda}$ and τ.

	d	EC2-NS	EC2-MS	EC2-SN	STO	Naive
$\|\cdot\|_F$	50	7.43	7.84	5.06	8.38	12.8
	100	10.27	12.17	7.88	13.90	27.16
	200	18.32	19.15	12.86	21.85	55.71
	400	25.66	27.83	19.11	35.94	114.32
$\|\cdot\|_2$	50	4.66	4.91	2.81	5.61	10.09
	100	5.26	7.46	4.00	8.85	21.49
	200	5.73	6.11	3.52	9.39	44.06
	400	5.86	7.21	3.66	15.69	90.88

TABLE 3.1. Results with Toeplitz Matrix

	d	EC2-NS	EC2-MS	EC2-SN	STO	Naive
$\|\cdot\|_F$	50	8.19	8.18	4.24	8.71	13.39
	100	13.86	16.48	7.81	19.09	26.59
	200	32.95	41.39	20.64	44.56	52.84
	400	49.35	57.19	36.76	68.18	106.8
$\|\cdot\|_2$	50	3.51	3.35	1.65	4.11	10.48
	100	9.09	12.41	4.35	14.59	20.88
	200	21.93	32.28	8.79	34.77	41.28
	400	30.66	39.43	19.48	49.97	83.45

TABLE 3.2. Results with Banded Matrix

Using the tuning parameters we have selected, we now generate 1000 data sets from each of the underlying covariance models with sample size 110, in $d = 50, 100, 200, 400$ and calculate the corresponding estimators. We show the results in Tables 3.1, 3.2 and 3.3.

	d	EC2-NS	EC2-MS	EC2-SN	STO	Naive
$\|\cdot\|_F$	50	5.31	5.52	1.85	6.96	13.63
	100	17.99	21.09	5.3	25.59	27.12
	200	24.07	21.29	9.98	28.7	52.26
	400	44.26	48.85	21.82	66.15	108.59
$\|\cdot\|_2$	50	2.9	2.41	0.93	4.52	10.66
	100	12.01	16.43	2.33	20.09	21.28
	200	11.19	11.37	4.16	15.97	42.53
	400	27.49	29.09	10.95	48.46	84.79

TABLE 3.3. Results with Block Matrix

CHAPTER 4

Robust Risk Analysis

4.1. Introduction

The problem of properly evaluating the risk for a portfolio of mines is poorly understood. Market risks have been well understood in this area for many years - mines can be priced with real options models in the spirit of [10], and the approximate value of the mine can be evaluated given the structure of the mine (reserves, capacity, cost of extraction, local tax rates), and certain market inputs (price of the underlying mineral, volatility of the underlying, interest rates). In an idealized world, this would encapsulate all the risks of a mining portfolio, so standard techniques for calculating portfolio risks could be used.

A mining operation, though, contains many operational risks that have a very asymmetric payoff profile, and are infrequent and location-dependent. Consider the case of a tailings dam failure: the mechanism by which is happens is poorly understood, as failures happen much more frequently than one would expect, and we have sparse data from which to estimate such an occurrence. A tailings dam is typically built to withstand a certain rainfall event (say, one in fifty or one in hundred years) - however, estimating that threshold correctly is challenging and non-robust to model misspecification, as seen in Chapter 2. The risk of a tailings dam failure, though, is catastrophic - costs of repairs may exceed the value of the mine itself, and there is a great likelihood that the operator of the mine will lose even more money through lawsuits and penalties from local regulators. In addition to the risk of extreme events, the holder of a portfolio of mines must be concerned with the correlation between extreme events in a portfolio; one may be able to withstand certain losses as an individual event, but multiple simultaneous mine failures may require too much capital to remain solvent.

A more general concern (beyond risks inherent to just mines) is that our methods of estimating extremal risk metrics may not be adequate. We wish to use a technique that is robust to the model being used to price and calculate risk, and also to consider various stress scenarios.

In order to overcome these issues, we consider three recent advances that can be used together to overcome these issues:

(1) The Robust Real Options model developed in Chapter 2 builds on the tools of real options methodology to incorporate an estimation-robust disaster risk for a mine site using the techniques for distributionally robust extreme values analysis developed by [7]. This enables the user to better calculate the value of a mine given precipitation data from the mine site, and come up with a worst case arrival rate for the level of tolerance to which the mine was built.

(2) Multivariate Max Stable covariance estimation in Chapter 3 uses developments in the theory of high-dimensional covariance estimation and simulation of max-stable vectors to estimate a sparse covariance matrix for the generative Gaussian process of the maxima of rainfall at a variety of sites.

(3) The Sample-Out-of-Sample methodology [5] is a method of performing data-driven stress testing, in which one measures the impact of a plausible set of out-of-sample scenarios on a performance measure of interest. This way, we can incorporate simulated data using items (1) and (2) into a set of scenarios that include historical data, and calculate a variety of robust risk metrics

In this chapter, we will use these techniques on a portfolio of copper mines, first by valuing them with the robust real options model, to come up with the initial portfolio value, and calculating a set of Greeks (partial derivatives with respect to the underlying). A variety of stressed scenarios using the estimated covariance matrix from item (2) and the robust failure levels from item (1) will then be generated, along with market prices from the underlying asset. We can then combine this with the historical data to have two data sets that are used as the inputs to item (3), where we can then calculate a set of model robust stress tests and risk metrics.

4.2. Sample-Out-of-Sample Methodology

Sample-out-of-Sample (SOS) is a novel method of calculating robust non-parametric bounds on risk metrics that incorporate the effect of out-of-sample stress scenarios. Suppose one wishes to calculate a quantity of interest, $E[L(X)]$ where X is the risk factor of interest, and $L(X)$ is the corresponding loss. We let $X_1, ...X_n \in R^l$ be a historical i.i.d. sample of X. Typically, one would use these samples to estimate $E[L(X)]$. However, we also wish to incorporate a variety of plausible stress scenarios $Y_1, ..., Y_n$, which may be drawn from a different distribution than the X_i's. If we let $Z_k = X_k$ and $Z_{n+k} = Y_k$ for $k = 1, ..., n$ then the corresponding SOS profile function will be:

$$R_n^W(\theta) = \min \sum_{i,k} \|X_i - Z_k\|_2^2 \pi(i,k)$$
$$\text{s.t. } \sum_k \pi(i,k) = \frac{1}{n} \forall i,\ \pi(i,k) \geq 0 \ \forall i,k, \sum_{i=i,k} L(Z_k)\pi(i,k) = \theta.$$

For a number of cases, the asymptotic distribution of $R_n(\theta)$ is known. For example, in the one-dimensional case:

$$nR_n(\theta) \to vR$$

where R is a chi-squared distribution with one degree of freedom , and v has a semi-closed form solution. Then, we can consider an confidence interval of the form $\{\theta : R_n(\theta) \leq \frac{\delta}{n}\}$ where $\frac{\delta}{v}$ is the corresponding quantile of the χ^2 distribution.

We now formalize these definitions, and introduce the concepts of SOS for estimating equations, and implicit and explicit SOS, and summarize the asymptotic results.

4.2.1. Sample of Out Sample for Means.

Definition 9. The SOS function, $R_n^W(\cdot)$, to estimate $\theta_* = E(X)$ is defined as

$$R_n^W(\theta_*) = \inf\{\int\int \|x - z\|_2^2 \pi(dx, dz) :$$
$$\text{s.t. } \pi \in \mathcal{P}(\mathcal{X}_n \times \mathcal{Z}_{(n+m)}), \pi_X = \mu_n, \pi_Z = v_n, \int z v_n(dz) = \theta_*\}$$

where $\mathcal{X}_n = \{x_1, ...x_n\}$, $\mathcal{Y}_m = \{y_1, ..., y_m\}$, $\mathcal{Z}_{n+m} = \mathcal{X}_n \cup \mathcal{Y}_m$, $\mu_n = \frac{1}{n}\sum \delta_{X_i}(dx)$, $v_n = \frac{1}{n}\sum \delta_{Z_i}(dz)$. For any closed set $\mathcal{C}$, we denote by $\mathcal{P}(\mathcal{C})$ the set of probability measures supported on $\mathcal{C}$. For any

$\pi \in \mathcal{P}(\mathcal{X}_n \times \mathcal{Z}_{n+m})$, we write $\pi_X \in \mathcal{P}(\mathcal{X}_n)$ to denote the marginal distribution with respect to the first random variable X, and π_Z is defined similarly.

THEOREM 10. *(SOS Profile Function Asymptotics for Means) Assume* $E\|X_1\|_2^2 + E\|Y_1\|_2^2 < \infty$, *and that* X *and* Y *have positive densities* $f_X(\cdot)$ *and* $f_Y(\cdot)$. *Then*

$$nR_n^W(\theta_*) \Rightarrow \sigma^2 \chi_1^2$$

where $\sigma^2 = Var(X)$.

4.2.2. SOS Function for Estimating Equations. Suppose $h : R^d \times R \to R$ and that $\theta_* \in R^d$ satisfies:

$$E[h(\theta_*, X)] = 0$$

and

$$E\|h(\theta_*, X)\|_2^2 < \infty.$$

DEFINITION 11. (Implicit SOS Profile Function for Estimating Equations). Let

$$R_n^W(\theta_*) = \inf\{\int\int \|h(\theta_*, x) - h(\theta_*, z)\|_2^2 \, \pi(dx, dz) : \\ s.t.\ \pi \in \mathcal{P}(\mathcal{X}_n^h(\theta_*) \times Z_n^h(\theta_*)), \pi_X = \mu_n, \int h(\theta_*, z)\pi_Z(dz) = 0\}$$

where $\mathcal{X}_n^h(\theta_*) = \{h(\theta_*, X_i) : X_i \in \mathcal{X}_n\}$ and $\mathcal{Z}_n^h(\theta_*) = \{h(\theta_*, Z_i) : Z_i \in \mathcal{Z}_{n+m}\}$.

The implicit formulation is useful as frequently the values given by the estimating equation are more informative than the underlying values of x, so the information is used in a more efficient manner.

THEOREM 12. *(Implicit SOS Profile Function Asymptotics) Suppose* $g_X(\cdot)$ *is the density for* $h(\theta_*, X_i)$ *and* $g_Y(\cdot)$ *is the density of* $h(\theta_*, Y_i) \in R^q$. *Then, the Wasserstein profile function satisfies:*

$$nR_n^W(\theta_*) \Rightarrow Var(h(\theta_*, X_1)\chi_1^2$$

The next SOS function type we consider is denoted the Explicit SOS because we use the explicit distances from the sample data.

DEFINITION 13. (Explicit SOS Profile Function for Estimating Equations)

$$R_n^W(\theta_*) = \inf\{\int\int \|x-z\|_2^2 \,\pi(dx,dz) : \\ s.t.\ \pi \in \mathcal{P}(\mathcal{X}_n \times Z_n), \pi_X = \mu_n, \int h(\theta_*, z)\pi_Z(dz) = 0\}.$$

THEOREM 14. *Suppose that the derivative of $h(\theta_*, x)$ with respect to x, $D_x h(\theta_*, \cdot)$ is a continuous function of x and the second derivative with respect to x is bounded for all x. Let $V_i = D_x h(\theta_*, X_i) \cdot D_x h(\theta_*, X_i)^T$ and assume that $\Upsilon = E(V_i)$ is strictly positive define. Then*

$$nR_n \Rightarrow \tilde{Z}^T \Upsilon^{-1} \tilde{Z} \tag{4.2.1}$$

where $\tilde{Z} \sim N(0, Var(h(\theta_*, X)))$

4.3. Application of the Techniques

Given a portfolio of mines, we can use the following procedure to calculate robust risk metrics, namely a 1-year $(1-\alpha)\%$ VaR. We used a dataset of $d = 23$ copper mines throughout the Americas in our example.

(1) Given a set of mine locations, calibrate the GEV processes for the individual time series of annual maxima. For this experiment, we use the NOAAA ERA20C dataset, which provides averaged data on a grid. We use the nearest grid-point to the mine's location.

(2) For the level of tolerance to which the mines were built, find the robust worst-case probability p_i for each individual time series. We assumed that every mine in our dataset was built to a 1-in-100 year specification; however, users can obviously enter a different quantile based on their research and beliefs.

(3) Price the mines using the robust real options model using the failure probabilities calculated in step 2. In doing this computation, also calculate the Greeks of interest - namely, Delta $\frac{\partial C}{\partial S}$ and Gamma $\frac{\partial^2 C}{\partial S^2}$, however, one can also calculate additional Greeks such as Vega $\frac{\partial C}{\partial \sigma}$ or

Rho $\frac{\partial C}{\partial r}$ if one is interested in investigating interesting rate or volatility risk as well. Using the current mineral price, we can calculate the value of the portfolio V_0.

(4) Using the methods of EC2-SN estimator from Chapter 3, estimate a sparse correlation matrix $\hat{\Sigma}$ for the underlying max-stable vector for the portfolio of mines. It is recommended that one estimate the tuning parameters λ_1 and λ_2 by cross-validation using the analogous soft-thresholding algorithm, and then estimating τ using cross-validation.

(5) For the sample-out-of-sample procedure, generate the $X_i \in \mathbb{R}^{d+1}$ where the first column is a set of copper prices generated from (or from a calibrated model in which we have a high degree of confidence), and the remaining d columns are indicators for mine failure - for the X_i we assume they are all zeros. Therefore, $L(X)$ can be calculated simply by using the change in copper prices and the Taylor expansion for each mine to re-value the portfolio.

(6) For the sample-out-of-sample procedure, generate the $Y_i \in \mathbb{R}^{d+1}$ as follows: For the copper prices, either use the same historical model or use a stressed model. Mine failure are simulated by generating values from a max-stable vector with the covariance matrix $\hat{\Sigma}$, and recording a failure for a particular mine site whenever the corresponding element in the random vector is greater than the $(1-p_i)$ quantile of the marginal. $L(Y)$ is calculated as follows: If a failure for a particular mine does not occur, we use the Delta-Gamma expansion as we did for X, and in the event of a mine failure, the value of the mine instead to 0. The new value of the portfolio is then calculated.

(7) With vectors $L(X)$ and $L(Y)$, we use the estimating equation for $(1-\alpha)\%$ VaR

$$h(\theta; L(X)) = 1\{L(X) > \theta\} - \alpha$$

which will have the limiting distribution (4.2.1). We can then calculate an appropriate SOS confidence interval.

(8) We can repeat steps 5-7 multiple times and average the results.

4.4. Results

For the purpose of strictly isolating the joint disaster risk in a portfolio, we do not consider the market/delta risk in the following examples, that is, we assume that the change in market price used in Section4.3 is zero for the underlying. Losses for a disaster were assumed to be 100% of the value of the mine, as priced with the robust real options model. We used the following sets of simulated data as inputs to the SOS framework:

(1) Disaster arrivals with the correlation structure estimated from the real data, with a 1% quantile threshold

(2) Disaster arrivals with the correlation structure as above, with the robust worst-case quantile estimated from the rainfall data

(3) Disaster arrivals with the 1% threshold combined as the base data, and disaster arrivals with the robust worst-case thresholds as the stressed data

For each of these data types, we calculated two 95% SOS-based confidence intervals for the following quantities of interest:

(1) Mean losses for a 1 year portfolio horizon

(2) 90% CVAR (Expected Shortfall)

We used an explicit formulation of SOS, where the underlying data was assumed to be 1-dimensional, i.e.:

$$X_i = \sum_{j=1}^{n} V_j 1_{\{D_i^{(j)}=1\}}$$

where V_j is the value of the jth mine at time 0, and $D_i^{(j)}$ is an indicator for the event that the $j-th$ mine has a disaster in the 1 year time period. We consider four settings, $n = 20, 50, 100$, and 200. For each setting, we repeated the experiment 1000 times, and noted the mean of upper and lower bounds, and the mean and standard deviation of the interval width for each method. The results are summarized in Table 4.1 for the mean losses and Table 4.2 for the expected shortfall.

n	Data Set	Mean Lower Bound	Mean Upper Bound	Mean Interval Length	SD of Length
20	1% quantile	18.2	67.8	49.6	12.85
	Robust Quantile	98.5	287.1	188.6	15.2
	Stress Test	15.8	70.7	54.9	9.3
50	1% quantile	21.8	69.2	47.4	3.6
	Robust Quantile	117.0	247.5	130.5	3.8
	Stress Test	21.4	70.2	48.8	3.2
100	1% quantile	12.2	32.3	20.1	2.3
	Robust Quantile	133.9	219.6	86.3	1.7
	Stress Test	12.0	32.9	20.9	1.5
200	1% quantile	17.7	38.9	21.1	0.7
	Robust Quantile	142.0	205.0	63.0	0.85
	Stress Test	17.5	39.1	21.5	0.6

TABLE 4.1. Mean Losses

n	Data Set	Mean Lower Bound	Mean Upper Bound	Mean Interval Length	SD of Length
20	1% quantile	277.6	734.1	456.5	280.1
	Robust Quantile	469.7	1414.4	944.7	444.6
	Stress Test	191.2	790.3	599.1	266.6
50	1% quantile	154.9	652.4	497.5	111.1
	Robust Quantile	551.7	1274.5	722.8	174.2
	Stress Test	148.9	682.4	533.4	117.8
100	1% quantile	74.2	571.9	497.7	50.2
	Robust Quantile	604.2	1078.4	474.2	73.5
	Stress Test	61.9	583.7	521.8	43.3
200	1% quantile	123.9	374.1	250.2	28.2
	Robust Quantile	608.4	1008.3	399.9	40.8
	Stress Test	118.1	381.4	263.2	30.7

TABLE 4.2. 90% Conditional Value at Risk

CHAPTER 5

Robust Performance Analysis with Independence Constraints

5.1. Introduction

So far, we have demonstrated several applications of robust performance analysis and related ideas to a number of problems in the mining sector. In this chapter, we will investigate several results in the field of robust performance analysis that are of independent theoretical interest and can be applied to a wide variety of other problems.

Robust performance analysis is concerned with the problem of evaluating the worst case performance measure of interest (typically described as an expectation) among all plausible probability models, such as those within certain tolerance of a baseline model which is believed to be reflective of reality. Taken literally, this problem formulation can be challenging because it gives rise to an infinite dimensional optimization problem (note that we mentioned "all models" within certain tolerance). When the tolerance region is described in terms of Kullback-Leibler divergence (and other related notions), this apparently daunting optimization problem is often tractable, and this tractability feature has been exploited in a range of literature in recent years, for example in control theory ([40, 50, 52]), distributionally robust optimization ([1]), finance ([30]), economics ([33]) and queueing ([41]).

Tolerance regions based on the Kullback-Leibler divergence, however, fail to incorporate information that is often quite natural to assume in common stochastic settings, and that should be added in terms of constraints in the underlying robust performance analysis formulation. One such natural and important constraint is the i.i.d. property, often arising in models involving random walk input. Failing to inform the i.i.d. property even in simple situations involving random walk models can have important consequences in terms of the accurate assessment of worst case performance measures of interest.

Unfortunately, however, a robust formulation in which the i.i.d. property is added as an extra constraint on top of the Kullback-Leibler imposed tolerance gives rise to an optimization problem

which is no longer easy to handle. However, in the context of performance analysis associated a class of large deviations events, such robust formulation gives rise to a problem for which asymptotically optimal solutions can be constructed; we illustrate this idea in the setting of i.i.d. random walks.

The rest of the chapter is organized as follows. In Section 2 we provide a precise mathematical formulation of the robust performance analysis problem with i.i.d. constraints and explain why the problem is very challenging. In Section 3 we provide a strategy that allows to solve this challenging problem asymptotically in a large deviations regime. In Section 4, we provide numerical examples which illustrate the performance of our proposed solution and the impact of adding i.i.d. constraints in the robust formulation. In Section 5, we provide an alternate algorithm that does not require the large deviations setting, has promising numerical results, and unknown theoretical properties.

5.2. Problem Formulation

Let $\{X_k : k \geq 0\}$ be a sequence of zero mean i.i.d. random variables. Define $S_0 = 0$ and put $S_n = X_1 + \cdots + X_n$. Let us use $F(\cdot)$ to denote the CDF (Cumulative Distribution Function) of X_i, that is, $P(X_i \leq x) = F(x)$ and we use $P_F(\cdot)$ to denote the product measure generated by $F(\cdot)$. We use $P_F^n(\cdot)$ to denote the projection of $P_F(\cdot)$ onto its n first coordinates. Simply put, P_F^n describes the joint distribution of the random variables $(X_1, \ldots, X_n)$. The expectation operator associated to $P_F(\cdot)$ and $P_F^n(\cdot)$ is denoted by $E_F(\cdot)$ and $E_F^n(\cdot)$, respectively. We define $\psi_F(\theta) = \log E_F^1 \exp(\theta X_1)$ and assume that $\psi_F(\theta) < \infty$ for θ in a neighborhood of the origin.

Now, define $A_n = \{S_n/n \in A\}$ for a closed set A which does not contain the mean of X_k. We are concerned with the problem of estimating $P_F(A_n)$. Observe that $P_F(A_n) \to 0$ as $n \to \infty$ because of the law of large numbers. Moreover, because $\psi_F(\cdot)$ is finite in a neighborhood of the origin we have that $P_F(A_n) \leq \exp(-\delta n)$ for some $\delta > 0$ for all n sufficiently large.

In contrast to standard rare event estimation problems, however, here we assume that $F(\cdot)$ is unknown. Nevertheless, based on some evidence (for example based on data or expert knowledge) let us assume that we have obtained a CDF $G(\cdot)$, which approximates $F(\cdot)$ in a suitable sense, for example in the Kullback-Leibler sense which we shall review momentarily. Let us write $P_G(\cdot)$ to denote the product measure associated to $G(\cdot)$ and we use $E_G(\cdot)$ for the expectation operator

corresponding to $P_G(\cdot)$. Similarly as before, $P_G^n(\cdot)$ is the projection of $P_G(\cdot)$ onto its n first coordinates and we use $E_G^n(\cdot)$ to denote the expectation operator associated to $P_G^n(\cdot)$.

We assume that the likelihood ratio dP_F^n/dP_G^n is well defined and therefore the Kullback-Leibler divergence of P_F^n with respect to P_G^n is defined via

$$R(P_F^n||P_G^n) = E_F^n \log\left(\frac{dP_F^n}{dP_G^n}\right) = nE_F^1 \log\left(\frac{dP_F^1}{dP_G^1}(X_1)\right) = n\int \log\left(\frac{dF}{dG}(x)\right) dF(x).$$

If dP_F^n/dP_G^n fails to exist (i.e. P_F^n is not absolutely continuous with respect to P_G^n), then the Kullback-Leibler divergence is defined as infinity. It is elementary to verify that $R(\cdot\,||P_G^n)$ is convex (actually $R(\cdot||\cdot)$ is convex in both of its arguments; ([24]). The associated robust performance analysis problem with Kullback-Leibler constraint consists in solving

$$\max_{Q^n}\{Q^n(A_n) : R(Q^n||P_G^n) \leq \eta_n\}, \tag{5.2.1}$$

where η_n should be chosen to satisfy

$$\eta_n \approx n\int \log\left(\frac{dF}{dG}(x)\right) dF(x).$$

One might select η_n by estimating $\int \log(dF/dG(x))\, dF(x)$ using available data.

The optimization problem (5.2.1) is a concave mathematical program; the objective function to maximize is linear (in particular concave) in the variable Q^n and, as mentioned earlier, the constraint is convex. Moreover, as we shall see in the body of the section (see Equations (5.2.5) and (5.2.6)), the optimal solution to (5.2.1), $Q_*^n(\cdot)$, can be characterized as a suitable mixture between $P_G^n(\cdot|A_n)$ and $P_G^n(\cdot|A_n^c)$. As the next result shows, it turns out that $Q_*^n(A_n)$ might differ substantially from $P_F^n(A_n)$ even if F is close to G in the Kullback-Leibler sense, that is, even in cases in which $\eta_n = o(n)$ as $n \to \infty$. In more detail, typically we will have $P_F^n(A_n) = \exp\left(-n\bar{I}_F(A) + o(n)\right)$, for some positive constant $\bar{I}_F(A)$, whereas the next result indicates that typically $Q_*^n(A_n) \geq \delta\eta_n/n$ for some $\delta > 0$ and large enough n. So, for example, if one builds an approximation G to F from data, one would need an exponentially large sample size (in n) in order to obtain an accurate estimate of the probability of interest using only the relative entropy constraint without recognizing that the data might have come from an i.i.d. model.

THEOREM 15. *Suppose that $\eta_n = o(n)$ and that $\eta_n > \delta > 0$ for some $\delta > 0$ uniformly over n. Assume also that $P_G^n(A_n) \in (\exp(-n/\delta'), \exp(-\delta' n))$ for some $\delta' > 0$ and all n sufficiently large. Then the optimal value of (5.2.1), $Q_*^n(A_n)$, satisfies*

$$Q_*^n(A_n) = \frac{\eta_n}{-\log P_G^n(A_n)}(1 + o(1))$$

as $n \to \infty$.

One of the main reasons for such a disparity, as we shall establish in the next section, is that the feasible region (i.e. $\{Q^n : R(Q^n||P_G^n) \leq \eta_n\}$) fails to recognize that we are interested only in models for which the i.i.d. property of the X_i's is preserved. So, introducing the i.i.d. constraint transforms problem (5.2.1) into the alternative form

$$\max_H \Bigg\{ P_H^n(A_n) = \int \cdots \int I\left(\frac{x_1 + \cdots + x_n}{n} \in A\right) dH(x_1) \cdots dH(x_n) : n \int \log\left(\frac{dH}{dG}(x)\right) dH(x) \leq \eta_n \Bigg\}. \tag{5.2.2}$$

Observe that the previous problem is not a concave program because the objective function to maximize is no longer concave. Unfortunately, in general (5.2.2) is very challenging to solve. In the next section we explain how to use large deviations theory to solve problem (5.2.2) in an asymptotic sense. We finish this section with a proof of our first theorem.

PROOF. To solve (5.2.1), we rewrite it in terms of the likelihood ratio between Q^n and G^n, namely $L = dQ^n/dP_G^n$, as

$$\begin{aligned} &\max && E_G^n[L; A_n] \\ &\text{subject to} && E_G^n[L \log L] \leq \eta_n, \end{aligned} \tag{5.2.3}$$

where the maximization is over $L \in \mathcal{L} = \{L \geq 0 : E_G^n L = 1\}$ and consider the Lagrangian relaxation

$$\max_{L \in \mathcal{L}} E_G^n[L; A_n] - \alpha(E_G^n[L \log L] - \eta_n). \tag{5.2.4}$$

Our goal is to find $\alpha^* \geq 0$ such that there is an L^* that solves (5.2.1) and moreover that $E_G^n[L^* \log L^*] = \eta_n$. Then this L^* will be optimal for (5.2.1) (c.f. [47], Theorem 1, p. 220).

First, note that when $\alpha = 0$, the optimal solution to (5.2.4) is clearly $L^* = I(A_n)/P_G^n(A_n)$, where $I(A_n)$ denotes the indicator function of the set A_n, which yields the optimal value 1. But then $E_G^n[L^* \log L^*] = -\log P_G^n(A_n) = \Omega(n)$ by our assumption in Theorem 15, and since we assume $\eta_n = o(n)$ we cannot have $E_G^n[L^* \log L^*] = \eta_n$ as n increases. Therefore the case $\alpha^* = 0$ is discriminated.

Now, given any fixed $\alpha > 0$, it can be verified by a convexity argument that the solution to the maximization (5.2.1) is given by

$$L^* \propto e^{I(A_n)/\alpha} = e^{1/\alpha} I(A_n); \tag{5.2.5}$$

see [33]. Now we write α_n for α to highlight the role of n, and introduce $\beta_n = 1/\alpha_n$ for convenience. We also write $p_n = P_G^n(A_n)$ and put $q_n = P_G^n\left(\bar{A}_n\right) = 1 - p_n$. Then (5.2.5) can be written as

$$L^* = \begin{cases} \frac{e^{\beta_n}}{p_n e^{\beta_n} + q_n} \text{ on } A_n \\ \frac{1}{p_n e^{\beta_n} + q_n} \text{ on } A_n^c \end{cases}. \tag{5.2.6}$$

We now proceed to find $\alpha_n^* > 0$, or $\beta_n^* = 1/\alpha_n^*$, such that

$$E_G^n[L^* \log L^*] = \eta_n. \tag{5.2.7}$$

Using the form of (5.2.6), (5.2.7) becomes

$$\beta_n \frac{p_n e^{\beta_n}}{p_n e^{\beta_n} + q_n} - \log(p_n e^{\beta_n} + q_n) = \eta_n. \tag{5.2.8}$$

Since $\eta_n > \delta > 0$ and $p_n \to 0$ as $n \to \infty$, we must have that all β_n satisfying (5.2.8) must also satisfy $\beta_n \to \infty$ as $n \to \infty$. Otherwise the left hand side converges to zero on some subsequence while the right hand side stays positively bounded away from zero. Now, we claim that $\limsup p_n e^{\beta_n} = 0$. Let us proceed assuming this claim for the moment and come back to this issue at the end of our proof. Then, by a Taylor series expansion applied to the left hand side of (5.2.8), we have that

$$\beta_n p_n e^{\beta_n}(1 + o(1)) = \eta_n, \tag{5.2.9}$$

which gives

$$\log \beta_n + \log p_n + \beta_n + o(1) = \log \eta_n.$$

Heuristically, we must have

$$\beta_n = \log \eta_n - \log p_n - \log \beta_n + o(1) = (\log \eta_n - \log p_n)(1 + o(1)). \tag{5.2.10}$$

To verify (5.2.10) rigorously, note that when we choose $\beta_n = \log(\eta_n/p_n)$, the left hand side of (5.2.9) becomes $\eta_n(\log(\eta_n/p_n))(1 + o(1))$ which is much more larger than η_n for n large enough. On the other hand, setting $\beta_n = 0$ gives the left hand side $o(1)$. Therefore, by continuity there must be a solution to (5.2.9) in the range $[0, \log(\eta_n/p_n)]$. Consequently, we have that

$$\beta_n = \log \eta_n - \log p_n - \log \beta_n + o(1) = \log \eta_n - \log p_n + r_n, \tag{5.2.11}$$

where the remainder term r_n satisfies $|r_n| \leq \log(\log(\eta_n/p_n)) + o(1)$, or equivalently we obtain that $r_n = o(\log(\eta_n/p_n))$, and hence (5.2.10).

Iterating the first equality in (5.2.10) using (5.2.11), we get further that

$$\beta_n = \log \eta_n - \log p_n - \log(\log \eta_n - \log p_n + r_n) + o(1) = \log \eta_n - \log p_n - \log(\log \eta_n - \log p_n) + o(1).$$

Finally, the optimal value is

$$E_G^n[L; A_n] = \frac{p_n e^{\beta_n}}{p_n e^{\beta_n} + q_n} \sim p_n e^{\beta_n} = \frac{\eta_n}{\log \eta_n - \log p_n}(1 + o(1)) \sim \frac{\eta_n}{-\log p_n}.$$

Now we must verify that indeed $\limsup p_n e^{\beta_n} = 0$. Assuming that $\limsup p_n e^{\beta_n} > 0$, then $\beta_{n_k} \geq \delta n_k$ along a subsequence $n_k \to \infty$ for $\delta > 0$. But then we must have from (5.2.8) that $\eta_{n_k} \geq \delta' n_k$ for some $\delta' > 0$, contradicting our assumption that $\eta_n = o(n)$. We therefore conclude the statement of our theorem. □

5.3. Our Main Result

5.3.1. A Large Deviations Rate Characterization. In order to prove our main result we shall impose additional technical conditions. We assume that $\psi_G(\cdot)$ is steep in the sense that for all $a \in (-\infty, \infty)$ there is θ_a such that $\psi_G'(\theta_a) = a$. Under this assumption we have that

$$P_G^n(A_n) = \exp\left(-n\bar{I}_G(A) + o(n)\right),$$

where

$$\bar{I}_G(A) = \inf_{x \in A} I_G(x) = \inf_{x \in A} \sup_{\theta} (\theta x - \psi_G(\theta)).$$

Similarly, for any CDF H, we define $\psi_H(\theta) = \log E_H^1 \exp(\theta X)$ and $I_H(x) = \sup_\theta (\theta x - \psi_H(\theta))$. Similarly as for the definition of $\bar{I}_G(A)$, we write $\bar{I}_H(A) = \inf_{x \in A} I_H(x)$. Now we are ready to state our main result.

THEOREM 16. *Let $int(A)$ denote the interior of the closed set A. Suppose that $\bar{I}_H(A) = \bar{I}_H(int(A)) \in (0, \infty)$ for any $H \in \mathcal{P}$ for some feasible set $\mathcal{P}$. We have*

$$\lim_{n \to \infty} \frac{1}{n} \log \max_{H \in \mathcal{P}, X_i \overset{i.i.d.}{\sim} H} P_H^n(A) = - \min_{H \in \mathcal{P}} \bar{I}_H(A).$$

PROOF. The proof follows from a large deviations argument. First,

$$\begin{aligned}
\liminf_{n \to \infty} \frac{1}{n} \log \max_{H \in \mathcal{P}, X_i \overset{i.i.d.}{\sim} H} P_H^n(A_n) &= \liminf_{n \to \infty} \max_{H \in \mathcal{P}, X_i \overset{i.i.d.}{\sim} H} \frac{1}{n} \log P_H^n(A_n) \\
&\geq \max_{H \in \mathcal{P}, X_i \overset{i.i.d.}{\sim} H} \liminf_{n \to \infty} \frac{1}{n} \log P_H^n(A_n) \\
&\geq - \min_{H \in \mathcal{P}} \bar{I}_H(\text{int}(A)) = - \min_{H \in \mathcal{P}} \bar{I}_H(A).
\end{aligned}$$

Next, since A is closed, using Chebycheff inequality (as in the proof of Cramer's theorem; ,see p. 27, Remark (c) in [21]) gives

$$P_H(A_n) \leq 2 \exp\left(-n \bar{I}_H(A)\right)$$

and hence

$$\limsup_{n \to \infty} \frac{1}{n} \log \max_{H \in \mathcal{P}, X_i \overset{i.i.d.}{\sim} H} P_H^n(A_n) \leq - \min_{H \in \mathcal{P}} \bar{I}_H(A).$$

Combining the upper and lower bounds we get our conclusion. □

The significance of the previous result is that the optimization problem that must be solved can now be cast as a concave program and thus it is more tractable. In order to have a concrete class of examples, let us focus on the case in which $A = [a, \infty)$. Really the key property that holds using this specific selection is that we can identify a specific element $a \in A$ such that $I_H(a) = \bar{I}_H(A)$. We consider the general problem of finding the minimum rate function over a class of distributions, namely

$$\inf_{H\in\mathcal{P}} I_H(a) = \inf_{H\in\mathcal{P}} \sup_{\theta}\{\theta a - \psi_H(\theta)\}. \tag{5.3.1}$$

LEMMA 17. *If $\mathcal{P}$ is a convex set, then the optimization program (5.3.1) is convex.*

PROOF. The inner objective function in (5.3.1) is concave as a function of θ, since $\psi_H(\cdot)$ is convex. For the outer objective function, note that $\psi_H(\theta)$ is concave in H, and so $\theta a - \psi_H(\theta)$ is convex in H. As the maximum over a set of convex functions (indexed by θ), the outer objective function $\sup_\theta\{\theta a - \psi_H(\theta)\}$ is also convex as a function of H. Therefore both the inner and outer optimizations in (5.3.1) are convex programs. □

5.3.2. Numerical Procedure. For our numerical procedure, in order to avoid the issue of optimization over infinite-dimensional variables, we concentrate on the case of discrete H. Also we focus on $A = [a, \infty)$. For convenience, we write $\mathbf{p} = (p(1), \ldots, p(m))$ as the weights over support points $\{x(1), \ldots, x(m)\}$. Moreover, we write

$$Z(\mathbf{p}) = \max_{\theta}\left\{\theta a - \log\sum_{i=1}^{m} p(i)e^{\theta x(i)}\right\} \tag{5.3.2}$$

as the outer objective function in (5.3.1). We concentrate on the case when a lies between $\min_{i=1,\ldots,m} x(i)$ and $\max_{i=1,\ldots,m} x(i)$; if $a > \max_{i=1,\ldots,m} x(i)$, then the worst-case probability of interest $\max_{P_H\in\mathcal{P}, X_i \overset{i.i.d.}{\sim} P_H} P_H^n(A_n)$ is trivially 0, whereas if $a < \min_{i=1,\ldots,m} x(i)$ then one can replace a by $\min_{i=1,\ldots,m} x(i)$ without changing the probability of interest. In the case $a \in [\min_{i=1,\ldots,m} x(i), \max_{i=1,\ldots,m} x(i)]$, the optimal solution for θ in (5.3.2) can be solved simply by finding the root of

$$\frac{\sum_{i=1}^m p(i)x(i)e^{\theta x(i)}}{\sum_{i=1}^m p(i)e^{\theta x(i)}} = a.$$

We now focus on $\min_{\mathbf{p}\in\mathcal{P}} Z(\mathbf{p})$. Suppose that $\mathcal{P}$ is a feasible region dictated by Kullback-Leibler divergence constraint, i.e. $\mathcal{P} = \{\mathbf{p} \geq 0 : \sum_{i=1}^m p(i)\log(p(i)/p^0(i)) \leq \eta, \sum_{i=1}^m p(i = 1\}$ for some baseline distribution $\mathbf{p}^0 = (p^0(1), \ldots, p^0(m))$. A particularly convenient procedure to approximate the optimal solution is to use the conditional gradient (or Frank-Wolfe) method [27]. This lies on the stepwise optimization, given the current solution $\mathbf{p}^k = (p^k(1), \ldots, p^k(m))$ at step k,

$$\min_{\mathbf{p}\in\mathcal{P}} \nabla Z(\mathbf{p}^k)(\mathbf{p} - \mathbf{p}^k). \tag{5.3.3}$$

This subroutine can be easily solved. In fact, we have

$$\nabla Z(\mathbf{p}^k) = \left(- \frac{d}{dp(i)} \left(\log \sum_{j=1}^{m} p(j) e^{\theta^k x(j)} \right) \Bigg|_{p(i)=p^k(i)} \right)_i = \left(- \frac{e^{\theta^k x(i)}}{\sum_{j=1}^{m} p(j) e^{\theta^k x(j)}} \right)_i$$

by simple arithmetic or by the use of the envelope theorem, where θ^k is the solution to

$$\frac{\sum_{i=1}^{m} p^k(i) x(i) e^{\theta x(i)}}{\sum_{i=1}^{m} p^k(i) e^{\theta x(i)}} = a.$$

For convenience, we let

$$\xi(i)(\mathbf{p}^k) = - \frac{e^{\theta^k x(i)}}{\sum_{j=1}^{m} p_j^k e^{\theta^k x(i)}}$$

be the i-th coordinate of $\nabla Z(\mathbf{p}^k)$. The solution to (5.3.3) is given by $\mathbf{q}^{k+1} = (q^{k+1}(1), \ldots, q^{k+1}(m))$, where

$$q^{k+1}(i) = \frac{p^0(i) e^{\beta \xi(i)(\mathbf{p}^k)}}{\sum_{j=1}^{m} p^0(j) e^{\beta \xi(j)(\mathbf{p}^k)}}$$

and $\beta < 0$ satisfies the equation

$$\frac{\sum_{i=1}^{m} \beta p^0(i) \xi(i)(\mathbf{p}^k) e^{\beta \xi(i)(\mathbf{p}^k)}}{\sum_{j=1}^{m} p^0(j) e^{\beta \xi(j)(\mathbf{p}^k)}} - \log \sum_{j=1}^{m} p^0(j) e^{\beta \xi(j)(\mathbf{p}^k)} = \eta.$$

If there is no negative root to this equation, then $\mathbf{q}^{k+1}$ is plainly a degenerate mass on $\operatorname{argmin}\{\nabla Z(\mathbf{p}^k)\}$.

Therefore, the iterative procedure is the following:

Iterative Procedure: Start from the baseline distribution $\mathbf{p}^0$ (or any other distribution). At each iteration k, given $\mathbf{p}^k$, do the following:

(1) Compute the root θ that solves

$$\frac{\sum_{i=1}^{m} p^k(i) x(i) e^{\theta x(i)}}{\sum_{i=1}^{m} p^k(i) e^{\theta x(i)}} = a.$$

(2) Compute

$$\xi(i) = - \frac{e^{\theta x(i)}}{\sum_{j=1}^{m} p^k(j) e^{\theta x(j)}} \quad \text{for } i = 1, \ldots, m.$$

(3) Compute, if any, the negative root of

$$\frac{\sum_{i=1}^{m} \beta p^0(i)\xi_i e^{\beta\xi(i)}}{\sum_{j=1}^{m} p^0(j)e^{\beta\xi(j)}} - \log \sum_{j=1}^{m} p^0(j)e^{\beta\xi(j)} = \eta.$$

(4) If there is a negative root β, then

$$q^{k+1}(i) = \frac{p^0(i)e^{\beta\xi(i)}}{\sum_{j=1}^{m} p^0(j)e^{\beta\xi(j)}} \quad \text{for } i = 1, \ldots, m.$$

Otherwise $q^{k+1}(i) = 1$ if $i = \text{argmin}\{\xi(i)\}$, and 0 for all other i's.

(5) Update $\mathbf{p}^{k+1} = (1 - \epsilon^{k+1})\mathbf{p}^k + \epsilon^{k+1}\mathbf{q}^{k+1}$ for some step size ϵ^{k+1}.

There are several choices for the step size ϵ^{k+1} in the above procedure. It can be a constant, or one can use the so-called limited minimization rule or the Armijo rule (see [3], p. 217). The latter two choices guarantee convergence to the optimal solution, in the sense that every limit point of the sequence $\mathbf{p}^{k+1}$, as computed by the procedure above with the chosen rule, will be optimal for minimizing $Z(\mathbf{p})$ ([3], Proposition 2.2.1 and Section 2.2.2).

5.4. Numerical Experiments

We will apply our algorithm to the case of two standard discrete distributions with finite support, namely the binomial distribution and a discrete distribution with random weights. We compare the outcome of robust performance analysis with i.i.d. constraints and without i.i.d. constraints, respectively.

Figure 5.1 shows the log-probabilities of the event $A_n = \{S_n/n > a\}$ with $a = 8$ associated with a binomial model with parameters $m = 10, p = 0.5$ as n increases. The true model is assumed to be binomial with $m = 10, p = 0.55$. This gives us $\eta = .05$, which is relatively low and chosen for illustrative purposes only. In both optimizations, we simply used step size $\epsilon^k = k^{-\frac{2}{3}}$ which resulted in empirical convergence of our procedure.

Figure 5.2 shows the log-probabilities of the event $A_n = \{S_n/n > a\}$ with $a = 8$ associated with a discrete distribution with on the integer support of $\{1, 2, \ldots, 10\}$ with the vector of weights

$$(.05, .12, .08, .13, .06, .04, .14, .13, .13, .12)$$

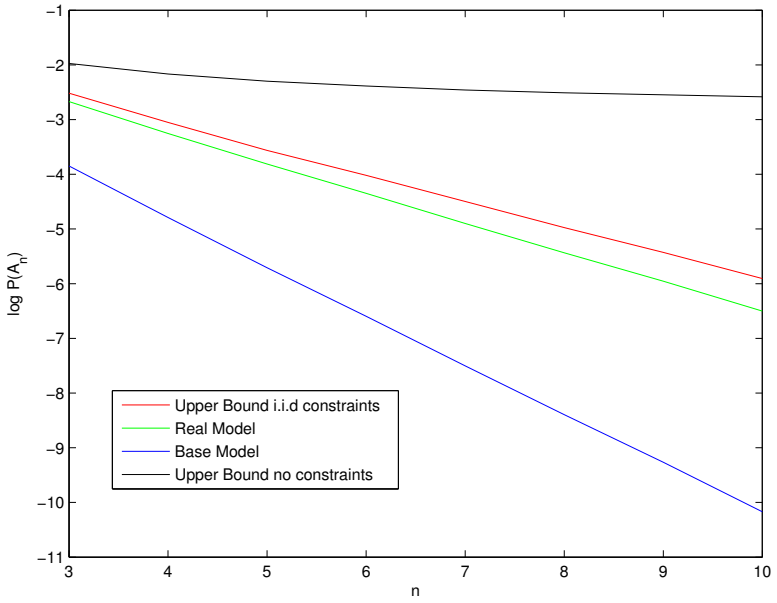

FIGURE 5.1. Binomial

obtained by random assignment (truncated to two decimal places here). We simulated $N = 300$ i.i.d. replications from the model. We took the standpoint of a modeler who does not have access to the true model, but instead uses maximum likelihood estimation (MLE) to estimate the weights and thus obtain a baseline distribution. This gives us $\eta \approx .02$, which is consistent for data-driven selection. As it can be seen, in both cases the upper bound with i.i.d constraint provides a much tighter bound to the real model than otherwise.

5.5. Preserving Independence Constraints in a Non-Asymptotic Setting

5.5.1. Motivation and Proposed Solution. Consider a set of i.i.d discrete random variables $X_1, ..., X_T$ with the state space: $(x(1), ..., x(m))$ and the corresponding probability vector $p = (p(x(1)), ..., p(x(m)))$. We will denote the baseline as $p^0 = (\frac{1}{m}, ..., \frac{1}{m})$, i.e., the uniform case. Suppose we wish to solve the following:

$$\begin{aligned} \max \quad & E^p[f(X(1), ..., X(T))] \\ s/t \quad & \sum_{i=1}^{m} p(i) \log(mp(i)) \leq \delta_0 \end{aligned}$$

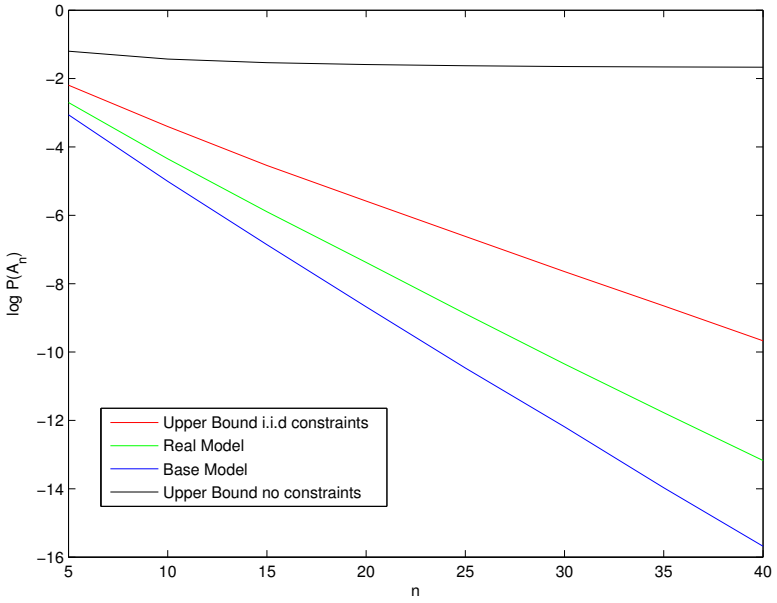

FIGURE 5.2. Random Weights

$$\sum_{i=1}^{m} p(i) = 1$$
$$p(i) \geq 0, \forall i \tag{5.5.1}$$

As discussed in the previous sections, while the i.i.d. assumption is natural and clearly provides a superior bound to the general worst-case bound with the Kullback-Leibler divergence constraint (at least in the large deviations case), the optimization problem is non-convex and therefore not amenable to the standard methods for such a problem. We present here an attempt to circumvent this issue with a bespoke algorithm.

The motivation for this approach is as follows: if we were to assume that the underlying distribution of the components of the random variables $X_1, ..., X_T$ were indeed i.i.d., that is $P(X_1, ..., X_T) = \prod_{i=1}^{n} p(X_1)$, then if we have two vectors (generated in this manner) $Y_1 = (Y_1(1) = y_1, ..., Y_1(T-1) = y_{T-1}, Y_1(T) = z_1)$, and $Y_2 = (Y_2(1) = y_1, ..., Y_2(T-1) = y_{T-1}, Y_2(T) = z_2)$ that differ by only one entry (in this case, the T-th coordinate), the ratio of their probabilities of occurrence would simply

be

$$\frac{P(Y_1)}{P(Y_2)} = \frac{p(z_1)\prod_{i=1}^{T-1} p(y_i)}{p(z_2)\prod_{i=1}^{T-1} p(y_i)}$$
$$= \frac{p(z_1)}{p(z_2)}$$

If one were to randomly generate using p^0 a vector Y with dimension, $(T-1)$, we can the generate a series of vectors,

$$Y_i = (Y_i(1), ..., Y_i(T-1), x(i))$$

by appending the ith coordinate of x, and then the solution to the optimization:

$$\max_{w(i), i=1,...m} \sum_{i=1}^{m} w(i) f(Y_i(1), ..., Y_i(T))$$
$$\text{s.t.} \sum_{i=1}^{m} w(i) \log(w(i)m) \leq \delta_1$$
$$\sum_{i=1}^{m} w(i) = 1$$
$$w_i \geq 0 \qquad \forall i$$

can be used to find a set of candidate solutions $\hat{p} = (\hat{p}(1), ..., \hat{p}(m))$ by solving the system of equations:

$$\frac{\hat{p}(i)}{\hat{p}(1)} = \frac{w(i)}{w(1)}, \qquad i = 2, ..., m$$
$$\sum_{i=1}^{m} \hat{p}(i) = 1$$

that will approximate the true solution to (5.5.1).

Likewise, we can generate a set of vectors $Y_1, ...Y_n$ with dimension $(T-1)$ using p^0, and then create $n \times m$ vectors:

$$Y_{i,j} = (Y_j(1), ..., Y_j(T-1), x(i))$$

by appending each member of x to each of the $Y_1, .., Y_n$. Now we can solve the optimization:

$$\max_{w(i,j),i=1,...m,j=1,...n} \sum_{i=1}^{m}\sum_{j=1}^{n} w(i,j) f(Y_{i,j}(1), ..., Y_{i,j}(T)) \tag{5.5.2}$$

$$\text{s.t.} \sum_{i=1}^{m}\sum_{j=1}^{n} w(i,j)\log(w(i,j)mn) \leq \delta_1$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} w(i,j) = 1$$

$$w_{i,j} \geq 0, \forall i, j$$

and we can find a candidate solution by solving the system of equations:

$$\frac{\hat{p}(i)}{\hat{p}(1)} = \frac{\sum_{j=1}^{n} w(i,j)}{\sum_{j=1}^{n} w(1,j)}, \forall i \tag{5.5.3}$$

$$\sum \hat{p}(i) = 1$$

$E^{\hat{p}}[f(X_1, ..., X_T)]$ can then be estimated using Monte Carlo simulation using $\hat{p}$.

There are two issues that this approach in and of itself does not address:

(1) $\hat{p}$ will depend on the "paths" that have been sampled, as will the Kullback-Liebler divergence $\delta = D(\hat{p}|p^0)$ and the expectation $E^{\hat{p}}[f(X_1, ..., X_T)]$.

(2) It is not immediately obvious what the relationship between δ_1 and δ is (this will be further obfuscated by the the variability in δ as a function).

We propose the following approach to remedy this:

(1) Fix a set of paths, and use a line-search method to find δ_1 such that the corresponding solution $\hat{p}$ has $D(\hat{p}|p^0) = \delta_0$

(2) Repeatedly generate a new set of samples, solve the optimization (5.5.2) using a fixed value of δ_1, and record $\delta = D(\hat{p}|p^0)$ and $E^{\hat{p}}[f(X_1, .., X_T)]$.

(3) Fit a linear regression model for δ and $E^{\hat{p}}[f(X_1, ..., X_T)]$ and use the expected value of $E^{\hat{p}}[f(X_1, ..., X_T)]$ given δ_0 as the estimated upper bound. Other alternatives might be K-Nearest Neighbors, etc.

See Algorithm 5 for a more complete summary of this algorithm.

Algorithm 5 To estimate $\max E^P[F(X_1,...,X_T)]$ where $D_1(P|P_0) \leq \delta_0$ and $X_1,...,X_T$ are i.i.d.
Given: State vector $(x_1,...,x_m)$, divergence bound δ_0

Step 1: Generate a set of vectors $Y_1,..,Y_n$ of length $(T-1)$ using p_0
Step 2: Construct a new set of $n \times m$ vectors by appending each member of x to each member of $Y_1,...Y_n$:

$$Y_{i,j} = (Y_j(1),...,Y_j(T-1),x(i))$$

Step 3: Now we solve the following optimization problem:

$$\begin{aligned} \max_{w(i,j),i=1,...m,j=1,...n} & \sum_{i=1}^{m}\sum_{j=1}^{n} w(i,j) f(Y_{i,j}(1),...,Y_{i,j}(T)) \\ \text{s.t.} & \sum_{i=1}^{m}\sum_{j=1}^{n} w(i,j)\log(w(i,j)mn) \leq \delta_1 \\ & \sum_{i=1}^{m}\sum_{j=1}^{n} w(i,j) = 1 \\ & w_{i,j} \geq 0, \forall i,j \end{aligned}$$

A reasonable initial guess for δ_1 is $T * \delta_0$.
The solution to this optimization will be an exponential tilting of p_0, that is:

$$w(i,j) \propto e^{\theta f(Y_{i,j}(1),...,Y_{i,j}(T))}$$

Step 4: We propose a candidate solution for to (5.5.1)using the solution to (5.5.2) to solve the system of equations:

$$\begin{aligned} \frac{\hat{p}(i)}{\hat{p}(1)} &= \frac{\sum_{j=1}^{n} w(i,j)}{\sum_{j=1}^{n} w(1,j)}, \text{for } i = 2,...n \\ \sum_{i=1}^{n} \hat{p}(i) &= 1 \end{aligned}$$

We can then simulate a series of vectors $X_1,...,X_k$ from $\hat{p}$, and estimate $\max_{p:D(p|p_0)\leq D(\hat{p}|p_0)} E^p[f(X_1,...,X_T)]$ as $Z = \sum_{i=1}^{n} f(X_i(1),...,X_i(T))$.
Step 5: Calculate

$$\delta = \sum_{i=1}^{m} \hat{p}(i)\log(\hat{p}(i)m)$$

Step 6: Repeat steps $3-5$, modifying δ_1 until $\delta \approx \delta_0$.
Step 7: Using the δ_1 found in step 5, repeat steps 1-4 N times, recording $\delta(i)$ and $Z(i)$.
Step 8: Fit a linear regression model $Z(i) = \alpha + \beta\delta(i) + \epsilon_i$, and estimate

$$\max_{p:D(p|p_0)<\delta_0} E^p[f(X_1,...,X_T)] = \alpha + \beta\delta_0$$

5.6. Numerical Results

In order to test the general performance of this algorithm, we fix a level of $\delta_1 = 0.1 \cdot T$, generate a set of $m \times n$ paths as described above, using $x = [-\frac{(m-1)}{2},...,\frac{(m-1)}{2}]$ and given some

$m=5, T=5$			$m=5, T=10$		
n	Algorithm 5	Local Maxima	n	Algorithm 5	Local Maxima
50	3.35	3.28	50	6.16	5.81
100	3.29	3.27	100	6.06	5.86
200	3.27	3.26	200	6.01	5.91
400	3.26	3.25	400	5.97	5.93
$m=9, T=5$			$m=9, T=10$		
n	Algorithm 5	Local Maxima	n	Algorithm 5	Local Maxima
50	6.90	6.4	50	12.62	10.54
100	6.60	6.28	100	12.01	10.56
200	4.67	6.28	200	11.8	10.78
400	6.37	6.28	400	11.66	11.07

TABLE 5.1. Comparison of Performance of Algorithm 5 versus local maxima for Vanilla Call Option

function of the paths f, calculate the solution $\hat{p}$ using (5.5.3) calculate $\delta = \sum_{i=1}^{m} \hat{p}(i) \log(m\hat{p}(i))$. Then, on the same set of paths, we calculate a probability measure $\bar{p}$ such that it maximizes the conditional expectation of $f(X_1, ..., X_T)$ assuming the random variables $X_1, .., X_T$ are i.i.d. and $\sum_{i=1}^{m} \bar{p}(i) \log(m\bar{p}(i)) \leq \delta$ using a conventional optimization tool (fmincon in MATLAB). Note that this is a non-convex problem, so we are likely to converge to a local minima. For both $\hat{p}$ and $\bar{p}$, we generate 5000 paths, and record their respective (functional) expectations. This procedure is repeated 1000 times, and we take the mean for each level of T and m. As we can see, this modified version of Algorithm 5 consistently outperforms the regular optimization, especially as m and T grow. For Table 5.1 we use $f(X_1, ..., X_T) = \max(\sum_{i=1}^{T} X_i, 0)$ (call option), and for Table 5.2 we use $f(X_1, ..., X_T) = \max(\sum_{i=1}^{T} X_i, 0) \cdot 1(\min_{j=1,...,m} \sum_{i=1}^{j} X_i < -5)$ (down-and-in-call).

$m = 5, T = 5$			$m = 5, T = 10$		
n	Algorithm 5	Local Maxima	n	Algorithm 5	Local Maxima
50	0.21	0.21	50	0.76	0.73
100	0.20	0.21	100	0.75	0.74
200	0.20	0.22	200	0.75	0.74
400	0.19	0.22	400	0.74	0.77
$m = 9, T = 5$			$m = 9, T = 10$		
n	Algorithm 5	Local Maxima	n	Algorithm 5	Local Maxima
50	0.77	0.71	50	2.36	2.00
100	0.75	0.71	100	2.34	2.04
200	0.74	0.73	200	2.32	2.08
400	0.73	0.76	400	2.32	2.14

TABLE 5.2. Comparison of Performance of Algorithm 5 versus local maxima for Down-and-in Call Option

Bibliography

[1] Aharon Ben-Tal, Dick den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

[2] Antonio E. Bernardo and Bhagwan Chowdhry. Resources, real options, and corporate strategy. *Journal of Financial Economics*, 63(2):211–234, 2002.

[3] Dimitri P Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.

[4] Peter J. Bickel and Elizaveta Levina. Covariance regularization by thresholding. *Ann. Statist.*, 36(6):2577–2604, 12 2008.

[5] Jose Blanchet and Yang Kang. Sample out-of-sample inference based on Wasserstein distance. *arXiv preprint arXiv:1605.01340*, 2016.

[6] Jose Blanchet and Zhipeng Liu. Malliavin-based multilevel Monte Carlo estimators for densities of max-stable processes. *arXiv preprint arXiv:1702.00428*, 2016.

[7] Jose Blanchet and Karthyek R. A. Murthy. On distributionally robust extreme value analysis. *arXiv preprint arXiv:1601.06858*, 2016.

[8] Adam Borison. Real options analysis: where are the emperor's clothes? *Journal of Applied Corporate Finance*, 17(2):17–31, 2005.

[9] Bruno Bouchard and Xavier Warin. *Numerical Methods in Finance*, chapter Monte-Carlo Valuation of American Options: Facts and New Algorithms to Improve Existing Methods, pages 215–255. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[10] Michael Brennan and Eduardo S Schwartz. Evaluating natural resource investments. *The Journal of Business*, 58(2):135–57, 1985.

[11] Axel Bücher and Johan Segers. On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes*, Mar 2017.

[12] René Carmona and Michael Ludkovski. Pricing asset scheduling flexibility using optimal switching. *Applied Mathematical Finance*, 15(5-6):405–447, 2008.

[13] Stefano Castruccio, Raphaël Huser, and Marc G. Genton. High-order composite likelihood inference for max-stable distributions and processes. *Journal of Computational and Graphical Statistics*, 25(4):1212–1229, 2016.

[14] Nicola Chiara, Michael J. Garvin, and Jan Vecer. Valuing simple multiple-exercise real options in infrastructure projects. *Journal of Infrastructure Systems*, 13(2), 2007.

[15] Andrzej Cichocki, Hyekyoung Lee, Yong-Deok Kim, and Seungjin Choi. Non-negative matrix factorization with alpha-divergence. *Pattern Recognition Letters*, 29(9):1433 – 1440, 2008.

[16] Emmanuelle Clément, Damien Lamberton, and Philip Protter. An analysis of a least squares regression method for American option pricing. *Finance and Stochastics*, 6(4):449–471, Oct 2002.

[17] Gonzalo Cortazar, Miguel Gravet, and Jorge Urzua. The valuation of multidimensional American real options using the LSM simulation method. *Comput. Oper. Res.*, 35(1):113–129, January 2008.

[18] John C. Cox, Stephen A. Ross, and Mark Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229 – 263, 1979.

[19] L. De Haan. A spectral representation for max-stable processes. *Ann. Probab.*, 12(4):1194–1204, 11 1984.

[20] Freddy Delbaen and Walter Schachermayer. *The Mathematics of Arbitrage*, chapter A General Version of the Fundamental Theorem of Asset Pricing (1994), pages 149–205. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[21] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. Springer Berlin Heidelberg, 2009.

[22] Clément Dombry, Frédéric Eyi-Minko, and Mathieu Ribatet. Conditional simulation of max-stable processes. *Biometrika*, 100(1):111–124, 2013.

[23] Clément Dombry, Marc G. Genton, Raphaël Huser, and Mathieu Ribatet. Full likelihood inference for max-stable data. *arXiv preprint arXiv:1703.08665*, 2017.

[24] Paul Dupuis and Richard S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons Inc, 1997.

[25] Ana F. Ferreira and Laurens de Haan. *Extreme Value Theory*. Springer, 2006.

[26] R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190, 1928.

[27] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, 1956.

[28] Andrea Gamba. Real options valuation: A Monte Carlo approach. *Faculty of Management, University of Calgary WP*, (2002/3), 2003.

[29] Marc G. Genton, Yanyuan Ma, and Huiyan Sang. On the likelihood function of Gaussian max-stable processes. *Biometrika*, 98(2):481–488, 2011.

[30] Paul Glasserman and Xingbo Xu. Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58, 2014.

[31] Boris V. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44(3):423–453, 1943.

[32] Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. *Robust Statistics*. John Wiley & Sons, 2011.

[33] Lars Peter Hansen and Thomas J. Sargent. *Robustness*. Princeton University Press, 2016.

[34] Lars Peter Hansen, Thomas J. Sargent, Gauhar Turmuhambetova, and Noah Williams. Robust control and model misspecification. *Journal of Economic Theory*, 128(1):45 – 90, 2006.

[35] J.Michael Harrison and Stanley R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3):215 – 260, 1981.

[36] Peter J. Huber and Elvezio M. Ronchetti. *Robust Statistics*. John Wiley & Sons Inc, 2009.

[37] John C. Hull. *Options, Futures, and Other Derivatives (8th Edition)*. Prentice Hall, 2011.

[38] Raphaël Huser and Anthony C. Davison. Composite likelihood estimation for the Brown-Resnick process. *Biometrika*, 100(2):511–518, 2013.

[39] Raphaël Huser and Marc G. Genton. Non-stationary dependence structures for spatial extremes. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(3):470–491, Sep 2016.

[40] Garud N. Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280, 2005.

[41] Ankit Jain, Andrew E. B. Lim, and J. George Shanthikumar. On the optimality of threshold control in queues with model uncertainty. *Queueing Systems*, 65(2):157–174, Jun 2010.

[42] Solomon Kullback and Richard A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 03 1951.

[43] M. Ross Leadbetter, Georg Lindgren, and Holger Rootzen. *Extremes and Related Properties of Random Sequences and Processes*. Springer, 2011.

[44] Han Liu, Lie Wang, and Tuo Zhao. Sparse covariance matrix estimation with eigenvalue constraints. *Journal of Computational and Graphical Statistics*, 23(2):439–459, 2014. PMID: 25620866.

[45] Zhipeng Liu, Jose H. Blanchet, A. B. Dieker, and Thomas Mikosch. Optimal exact simulation of max-stable and related random fields. *arXiv preprint arXiv:1609.06001*.

[46] Francis A. Longstaff and Eduardo S. Schwartz. Valuing American options by simulation: A simple least-squares approach. *The Review of Financial Studies*, 14(1):113–147, 2001.

[47] David G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons Inc, 1997.

[48] David G. Luenberger. *Investment Science*. Oxford University Press, 2013.

[49] N. Meinshausen and B. M. Hambly. Monte Carlo methods for the valuation of multiple-exercise options. *Mathematical Finance*, 14(4):557–583, 2004.

[50] Arnab Nilim and Laurent El Ghaoui. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.

[51] Simone A. Padoan, Mathieu Ribatet, and Scott A. Sisson. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277, 2010.

[52] Ian R. Petersen, Matthew R. James, and Paul Dupuis. Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *IEEE Transactions on Automatic Control*, 45(3):398–412, Mar 2000.

[53] Barbnabas Poczos and Jeff Schneider. On the estimation of alpha-divergences. In *International Conference on AI and Statistics (AISTATS)*, volume 15 of *JMLR Workshop and Conference Proceedings*, pages 609–617, 2011.

[54] Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 547–561, Berkeley, Calif., 1961. University of California Press.

[55] Mathieu Ribatet. Spatial extremes: Max-stable processes at work. *Journal de la Société Française de Statistique*, 154(2):156–177, 2013.

[56] Adam J. Rothman, Elizaveta Levina, and Ji Zhu. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104(485):177–186, 2009.

[57] Sabry A. Abdel Sabour and Richard Poulin. Valuing real capital investments using the least-squares monte carlo method. *The Engineering Economist*, 51(2):141–160, 2006.

[58] Francesco Serinaldi, András Bárdossy, and Chris G. Kilsby. Upper tail dependence in rainfall extremes: would we know it if we saw it? *Stochastic Environmental Research and Risk Assessment*, 29(4):1211–1233, May 2015.

[59] James E. Smith and Kevin F. McCardle. Valuing oil properties: Integrating option pricing and decision analysis approaches. *Operations Research*, 46(2):198–217, 1998.

[60] James E. Smith and Robert F. Nau. Valuing risky projects: Option pricing theory and decision analysis. *Manage. Sci.*, 41(5):795–816, May 1995.

[61] Lenos Trigeorgis. *Real Options: Managerial Flexibility and Strategy in Resource Allocation*. MIT Press, 1996.

[62] Lingzhou Xue, Shiqian Ma, and Hui Zou. Positive definite ℓ-1 penalized estimation of large covariance matrices. *Journal of the American Statistical Association*, 107(5):1480–1491, 10 2012.

APPENDIX A

Review of Sparse Covariance Estimation Technique

A.1. Hard Thresholding

Perhaps the simplest estimator that produces sparse covariance matrices is hard thresholding [4], which sets elements of the sample covariance to zero if they are below some pre-determined threshold. For sample covariance matrix S estimated from data from distribution F, the hard thresholding estimator $\hat{\Sigma}^{HTO}(\lambda)$ is defined element-wise:

$$\hat{\Sigma}^{HTO}(i,j)(\lambda) = \begin{cases} S(i,j), & \text{if } i = j \\ S(i,j)1\{|S(i,j)| > \lambda, & \text{if } i \neq j \end{cases}$$

.Intuitively, when the matrix being estimated is sparse, the hard-thresholding operator has the advantage of not estimating small elements, so noise does not accumulate in the estimation error. If we let

$$\mathcal{U}_\tau(q, c_0(p), M) = \left\{ \Sigma : \sigma(i,i) \leq M, \sum_{j=1}^{p} |\sigma(i,j)|^q \leq c_0(p), \text{for all } i \right\}$$

then we can summarize the following convergence properties of the hard thresholding operator:

THEOREM 18. *(Theorem 1, [4]) Suppose F is Gaussian with covariance matrix* $\Sigma \in \mathcal{U}_\tau(q, c_0(p), M)$. *Then, uniformly on* $\mathcal{U}_\tau(q, c_0(p), M)$, *for sufficiently large M', if*

$$\lambda_n = M' \sqrt{\frac{\log p}{n}}$$

and $\frac{\log p}{n} = o(1)$,*then*

$$\left\| \hat{\Sigma}^{HTO}(\lambda) - \Sigma \right\|_2 = O_P\left(c_0(p) \left(\frac{\log p}{n} \right)^{(1-q)/2} \right).$$

THEOREM 19. *(Theorem 2, [4]) Suppose F is Gaussian with covariance matrix $\Sigma \in \mathcal{U}_\tau(q, c_0(p), M)$. Then, uniformly on $\mathcal{U}_\tau(q, c_0(p), M)$, for sufficiently large M', if*

$$\lambda_n = M' \sqrt{\frac{\log p}{n}}$$

and $\frac{\log p}{n} = o(1)$,then

$$\left\| \hat{\Sigma}^{HTO}(\lambda) - \Sigma \right\|_F^2 = O_P\left(c_0(p) \left(\frac{\log p}{n} \right)^{(1-q)/2} \right).$$

A.2. Soft Thresholding

The hard thresholding estimator can be modified to provide additional shrinkage to the non-zero coefficients. The soft-thresholding estimator is a specific case of the generalized thresholding estimator developed by [56]. We define an element-wise operator s_λ on a matrix A that satisfies the following properties

$$(i) |s_\lambda(z)| \leq z$$

$$(ii) s_\lambda(z) = 0 \text{ for } |z| \leq \lambda$$

$$(iii) |s_\lambda(z) - z| \leq z$$

In the case of soft thresholding, we chose $s_\lambda(z) = \text{sign}(z) s_\lambda(|z| - \lambda)_+$, giving us the soft thresholding estimator $\hat{\Sigma}^{STO}(\lambda)$, defined element-wise:

$$\hat{\Sigma}^{STO}(i,j)(\lambda) = \text{sign}(S(i,j))(S(i,j) - \lambda)_+$$

The soft thresholding estimator has the same advantages as the hard thresholding operator, and has the added benefit of variance reduction through shrinkage.

THEOREM 20. *(Theorem 1, [56]) Suppose s_λ satisfies conditions (i) through (iii) and the marginals of F satisfy*

$$E[\exp(\lambda t^2)] < \infty$$

for $0 < |\lambda| < \lambda_0$. *Then, uniformly on* $\mathcal{U}_\tau(q, c_0(p), M)$ *for sufficiently large* M', *if* $\lambda_n = M'\sqrt{\log(p/n)} = o(1)$, *then*

$$\|s_{\lambda_n}(S) - \Sigma\|_2 = O_P\left(c_0(p)\left(\frac{\log p}{n}\right)^{(1-q)/2}\right).$$

THEOREM 21. *(Theorem 2, [56]) Suppose* s_λ *satisfies conditions* (i) *through* (iii) *and the marginals of* F *satisfy*

$$E[\exp(\lambda t^2)] < \infty$$

for $0 < |\lambda| < \lambda_0$, *and* $\sigma(i,i) \leq M$ *for all* i. *Then, for all sufficiently large* M', *if* $\lambda = M'\sqrt{\frac{\log(p)}{n}} = o(1)$, *then*

$$s_{\lambda_n}(\hat{\sigma}(i,j)) = 0$$

for all (i,j) such that $\sigma(i,j) = 0$ *with probability tending to 1. If we additionally assume that all nonzero elements of* Σ *satisfy* $|\sigma(i,j)| > \tau$ *where* $\sqrt{n}(\tau - \lambda) \to \infty$, *we also have, with probability tending to 1,*

$$sign(s_{\lambda_n}(\hat{\sigma}(i,j))\sigma(i,j)) = 1$$

for all (i,j) such that $\sigma(i,j) \neq 0$.

A.3. EC2 Estimator

The EC2 Estimator developed in [44] is a method of estimating correlation and covariance matrices that is related to the soft-thresholding operator. We consider this for the case of correlation matrices only for the purposes of this dissertation. We begin by defining the soft thresholding operator for correlation matrices as the solution to the following optimization:

$$\hat{\Sigma}^{STO}(\lambda) = \text{argmin}_{\Sigma_{ii}=1} \|\Sigma - S\|_F^2 + \lambda \|\Sigma\|_{1,off}$$

where $\|A\|_{1,\text{off}} = \sum_{i \neq j} |a(i,j)|$. This has the closed form solution:

$$\hat{\Sigma}^{STO}(\lambda)(i,j) = \begin{cases} \text{sign}(S(i,j))(S(i,j) - \lambda)_+, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases}.$$

The EC2 estimator is a modification of the STO operator that ensures the estimator is positive definite.

$$\hat{\Sigma}^{EC2}(\lambda,\tau) = \text{argmin}_{\Sigma_{ii}=1} \|\Sigma - S\|_F^2 + \lambda \|\Sigma\|_{1,off}$$
$$\text{s.t. } \tau \leq \Lambda_{min}(\Sigma). \tag{A.3.1}$$

There is no closed form solution to this, but it can be solved with an iterative algorithm that has closed form solutions at every step. We will detail this algorithm in the next section.

If we let

$$\mathcal{U}_\tau(q, c_0(p), M, \delta_{min}) = \left\{ \Sigma : \sigma(i,i) = 1, \sum_{j=1}^{p} |\sigma(i,j)|^q \leq c_0(p) + 1, \text{for all } i, \delta_{min} \leq \Lambda_{min}(\Sigma) \right\}$$

then we can summarize the following convergence properties of the hard thresholding operator:

THEOREM 22. *(Theorem 4.3, [44]) Suppose the marginals of F satisfy*

$$E[\exp(\lambda t^2)] < \infty$$

for $0 < |\lambda| < \lambda_0$ and $\Sigma \in \mathcal{U}_\tau(0, c_0(p), M, \delta_{min})$. Let N_d be the number of non-zero off-diagonal elements in Σ. Then if we let $\lambda = M'\sqrt{\frac{\log d}{n}} = o(1)$ and $\tau \leq \delta_{min}$, then we have

$$\left\|\hat{\Sigma}^{EC2}(\lambda,\tau) - \Sigma\right\|_F = O_P\left(c_0(p)\sqrt{\frac{N_d \log d}{n}}\right).$$

A.4. Iterative Soft-Thresholding and Projection Algorithm

The Iterative Soft-thresholding and projection algorithm of [44] allows us to solve the optimization (A.3.1). It can first be reparametrized by introducing an auxiliary variable Γ:

$$(\hat{\Sigma}, \hat{\Gamma}) = \text{argmin}_{\Sigma(i,i)=1, \tau \leq \Lambda_{min}(\Gamma)} \frac{1}{2} \|S - \Gamma\|_F^2 + \lambda \|\Sigma\|_{1,\text{off}}$$
$$\text{s.t. } \Gamma = \Sigma$$

which can then be rewritten as the equivalent saddle point problem (with use of the polarization identity):

$$(\hat{\Sigma}, \hat{\Gamma}, \hat{U}) = \text{argmin}_{\Sigma(i,i)=1, \tau \leq \Lambda_{min}(\Gamma)} \max_{U} \frac{1}{2} \|S - \Gamma\|_F^2 + \lambda \|\Sigma\|_{1,\text{off}} + \langle U, \Gamma - \Sigma \rangle + \frac{\rho}{2} \|\Gamma - \Sigma\|_F^2$$

where U is the Lagrange multiplier matrix. We proceed iteratively, where we have the solution $\Sigma^{(t)}, \Gamma^{(t)}, U^{(t)}$ at the tth step.

Step 1. Update Σ with the following soft thresholding problem:

$$\Sigma^{(t+1)} = \text{argmin}_{\Sigma(i,i)=1} \lambda \|\Sigma\|_{1,\text{off}} + \frac{\rho}{2} \left\| \frac{1}{\rho} U^{(t)} + \Gamma^{(t)} - \Sigma \right\|_F^2$$

which admits the closed form solution:

$$\Sigma^{(t+1)}(i,j) = \begin{cases} \text{sign}(\Gamma^{(t)}(i,j) + \frac{1}{\rho} U^{(t)}(i,j))(\Gamma^{(t)}(i,j) + \frac{1}{\rho} U^{(t)}(i,j) - \frac{\lambda}{\rho})_+, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases}.$$

Step 2. Given $\Sigma^{(t+1)}$,we then update Γ with

$$\Gamma^{(t+1)} = \text{argmin}_{\tau \leq \Lambda_{min}(\Gamma)} \left\| \Gamma - \frac{S + \rho \Sigma^{(t+1)} - U^{(t)}}{(1+\rho)} \right\|_F^2$$

which has the solution

$$\Gamma^{(t+1)} = \mathcal{P}_+ \left(\frac{S + \rho \Sigma^{(t+1)} - U^{(t)}}{(1+\rho)}, \tau \right)$$

where, for a matrix A, with spectral decomposition

$$A = \sum_{j=1}^{d} \delta_j v_j v_j^T$$

$$\mathcal{P}_+(A, \tau) = \sum_{j=1}^{d} \max(\delta_j, \tau) v_j v_j^T.$$

.

Step 3: Update $U^{(t+1)} = U^{(t)} + \rho(\Gamma^{(t+1)} - \Sigma^{(t+1)})$.