Towards a Robust Top-Down Model for Valuation of Mining Assets

Blanchet, J., Dolan, C., Iyengar, G., and Lall, U.

Abstract

Our goal is to create a simple, yet robust, statistical model which can be used to quantify the risk present in a portfolio of mining assets. In pursuit of this goal we aim at explaining a systematic approach which takes as input a model which is constructed based on fundamental economic principles and simple statistical technniques (e.g. a mixed-effect linear model with explanatory variables chosen from economic reasoning). Additional enrichment is then imposed, based on input coming from a more detailed model (built, for instance, from bottom up). And finally, the robustification step is obtained by computing worst-case performance analysis among all models that are within some distance of our simple model. This step quantifies the error induced by using a simple-yet-tractable model, which might be incorrect.

1. Introduction

There is growing interest in financial risk associated with water scarcity in the mining sector. This concern emanates largely from significant recent investments in desalination plants associated with mines in Chile. Increasing capital and operating costs associated with water for mining and mineral processing are seen as a potential challenge for the profitability of the mines. Since this concern has been publicly discussed, one would anticipate that these risk factors are priced into the credit risk ratings and net asset values of mining companies. This note explores whether or not a top down approach that relates these valuation measures to publicly available financial and water scarcity indicators can reveal something which factors are robustly reflected in the valuations.

Our goal is to provide an interface between financial indicators, which are inherent in a top-down type analysis, and more specific bottom-up indicators which are built from specific characteristics such as geophysical and operational attributes of the mines in consideration.

We provide a model for the financial risk quantification of mining companies with the following characteristics:

- a) Simplicity: The model must be easy to grasp and explain from a fundamental perspective.
- **b)** Scalability: Despite its simplicity, the model must be able to incorporate additional covariates and risk factors specially those coming from a more specific bottom-up type approach.
 - c) Robustness: Provide a systematic approach to evaluate model misspecification.

We understand that risk is quantified relative to a portfolio of assets in the mining (say gold) industry. So, we are interested in measuring the risk associated with holding a certain portfolio which is a linear combination of assets for a specific time horizon, assuming that the

portfolio's holdings remain constant during this time horizon. The types of models that we are able to build are reliable for a medium time horizon (on the order of a few years, about 250 weeks). Beyond this time horizon, there are time in-homogeneities which are difficult to quantify uniformly across all risk factors, even though the statistical properties of some risk factors (e.g. climate related characteristics, incorporated in some of indices) might be reliably estimated for longer time horizons. We are considering simulation models to handle longer time horizons.

We have focused on the development of a top-down statistical model for financial valuation. The model is a mixed-effect linear model with fundamental economic covariates. The model is enriched with additional covariates which are obtained from a bottom-up methodology, but further enhancements will be provided in the next stages of the project when additional information is processed from specific asset characteristics.

We shall first concentrate on describing the model, emphasizing item a). It is worth emphasizing that one of the most interesting observations suggested by our analysis is that water scarcity (which we believe is a relevant indicator of environmental risk for the mining industry) is actually positively associated with market values. While this sounds counter-intuitive, given that suitable water supply is rather important for the mining process, we believe that such an association might result because operational mines which are "risky" from the water scarcity perspective are so profitable that they are worth exploiting despite the risk

It is important to keep in mind that likely there is a systematic bias explaining this positive association as follows. Developers likely recognize that water scarcity might pose a threat. So, they might decide not to develop or operate a mine unless suitable conditions are in place; however, if they decide to operate despite scarcity, the mine probably is substantially profitable. Our data universe only includes mines that are operational, we do not include mines that are not operational. In order to appropriately incorporate the impact of water scarcity we should also consider non-operational mines (because maybe water scarcity is a significant factor for not having a mine operating). Economic analysis of non-operational mines is under development using real option valuation and the outcome of this analysis will be reported in the future.

In our model we include over 50 companies, and we add specific financial information from databases such as Bloomberg and SNL. A water scarcity index is used to capture some of the environmental risk. We recognize that there are other environmental risk indicators and we will include them as soon as they are properly developed from a bottom-up approach. We expect additional explanatory variables arising from a parallel bottom-up construction to be eventually incorporated into the simple model described in Section 2.

In order to cope with high dimensionality issues (keeping in mind item b)), we plan to pursue a Bayesian hierarchical construction, which can be suitably scaled if we are in the presence of the right (conjugate) structures. We provide a short review of these types of Bayesian models in Section 8.1.

Finally, we describe a robustification methodology (corresponding to item c)). The methodology that we described has connections to robustness notions studied in Economics (see [7]) and Operations Research ([1]); our discussion here is based on [2].

The idea is to use our constructed models to quantify risk. In order to recognize that these models, while simple, might contain structural errors, we discuss an approach based on convex optimization which consists in finding the worst case risk measures among all models

which are within some distance of the simple model under consideration. The point, as we shall explain in Section 4, is that by choosing suitable discrepancies between models, we can solve the robustification problem in terms of the baseline model. This is remarkable because we just need to make sure that a more realistic model (or even the hypothetical true model) is inside the feasible region in order to obtain a valid (i.e. correct) bound for the estimated risk in terms of the baseline model.

One potential problem with this general robustness methodology is that the bounds might be too large. In order to deal with this problem, we plan to use the information in the bottom-up models to constrain the optimization problem. We explain in Section 4 how constraints can be added so that the robustification problem still remains tractable.

In particular, to clarify, we plan to integrate the findings using bottom-up models and the current top-down model in two ways. First, by introducing explanatory variables, specially in the setting of environmental risk (in addition to the current use of the water scarcity index). Second, we plan to inform the quantitative risk assessments using the robustification procedure explained in Section 4 by introducing constraints suggested by the bottom-up model, which might have higher fidelity for certain quantities.

We shall provide the theoretical foundations for c), and discuss some of the aspects regarding item b), such as the use of Bayesian models. In future months we plan to apply these methodologies to our model. Regarding the basic model, item a), currently our model is calibrated at a company level. The value of the individual (single) assets (mines) can be obtained assuming that every variable that is unknown at the single asset level can be obtained by applying a proportionality factor based on mine production (which is known). Unfortunately, this type of assumption does not consider valuable assets which are not currently producing. In order to address this problem, in the next update of our model we will incorporate an additional factor based on real option valuation methodology, which is briefly discussed in Section 8.2. We point out that real option valuation is yet another way in which we can further add constraints to the optimization problem which must be solved to robustify our solution as discussed in Section 4.

2. A SIMPLE TOP-DOWN MODEL

The universe of companies used were a subset of names taken from the NYSE Arca Gold Miners Index (GDMNRT), for which we were able to acquire (reasonably) complete information of their gold mining exposure (from the SNL database). Financial and other indicators were obtained from Bloomberg.

We recognize that the data selection is representative of a particular population of companies, namely, those who have reasonably complete reporting practices and therefore this selection might induce biases. We believe, however, that this is precisely the type of universe which might be relevant for investment analysis within a risk profile consistent with that of institutional investors seeking controlled variability and strong long term returns. Therefore, as long as the findings in this report are used for the purpose of aiding such investor, the potential bias is not a significant source of concern.

A large number of production and financial variables were considered. In the end, the model selected uses the following basic variables:

Basic Variables.

- Water Scarcity Index (the sum, over mines, of mill capacity per-mine × BWS ratio of mine's location). The BWS ratio is the ratio of mine water use to water available in the area in which the mine is located.
- Underground Capacity Value (tons per year) × Price of Gold (per ton)
- Open Pit Capacity Value (tons per year) × Price of Gold (per ton)
- Net Income (in millions of \$)
- CAPEX (in millions of \$)
- Net Debt (in millions of \$)

2.1. The Models Considered.

We fit two linear regression models. We aim at obtaining a model for $Y_i(t)$, the market value (assets and equity) of a company at time t. The $X_i(t)$'s (predictor variables) are given according to the basic variables described above.

1. A fixed effect model (no specific effect for each company)

$$Y_{i}(t) = \alpha + \sum_{k=1}^{d} \beta_{k} X_{i,k}(t) + \varepsilon_{i}(t),$$

where *i* represents the *i*-th company, α represents the intercept, and the $X_{i,k}(\cdot)$'s are the basic variables. The time scale considered is of the order of two to three months.

2. A mixed effects (adding a random effects per company) which takes the form,

$$Y_{i}(t) = \alpha + U_{i} + \sum_{k=1}^{d} \beta_{k} X_{i,k}(t) + \varepsilon_{i}(t),$$

where $\varepsilon_i(t)$ captures errors within a company, across time, and U_i is a random variable which incorporates covariance structure between companies, and one assumes that, for all t, i, k

$$Cov\left(U_{i}, X_{i,k}\left(t\right)\right) = 0. \tag{1}$$

Model 1 is the departing point in our construction. Everything that appears in the right hand side of the equation has a direct economic interpretation. The random effects model (Model 2), is parsimonious and a natural extension of Model 1, but it is useful to keep Model 1 as a guiding fundamental tool given that its predictive power and the analysis of variance is earlier to interpret there. For more information on mixed effects models, see [6].

3. Model Output

The Model 1 produced an R^2 value of 72.9% (the adjusted R^2 is 72.7%) with p-value for the F statistic on the order of 10^{-30} , so there is certainly strong evidence to reject the hypothesis of no relation between the predictors and the value of the companies.

The coefficients are given next, all the p-values are substantially smaller than 10^{-4} , indicating that the coefficients are significant (given the model). It is important to note that the signs are well aligned with the fundamental interpretation of the corresponding variables.

Model 1 (ANOVA)	SumSq	DF	F	p-value
Water Scarcity In.	20.45	1	21.7	3.53x10^-6
Undergrnd Cap.	243.33	1	258.5	1.58x10^-52
Open Pit Cap.	226.82	1	240.9	2.13x10^-49
Net Income	32.03	1	34	7.1x10^-9
CAPEX	129.47	1	137.5	4.73x10^-30
Net Debt	94.38	1	100.3	1.15x10^-22
Error	1061.8	1128		

Model 1	Estimate	SE	tStat	p-value
(Intercept)	108.04	31.35	3.45	0.0006
Water Scarcity In	1.89	0.4	4.66	3.53x10^-6
Undergrnd Cap.	2.41	0.15	16.08	1.58x10^-52
Open Pit Cap.	2.05	0.13	15.52	2.13x10^-49
Net Income	1.89	0.33	5.83	7.1x10^-9
CAPEX	-8.27	0.71	-11.72	4.73x10^-30
Net Debt	-1.4	0.14	-10.01	1.15x10^-22

The Water Scarcity Index coefficient has a positive sign. We have provided a possible explanation for this finding in the Introduction. For the particular case of the Water Scarcity Index, we note that it has a direct R^2 equal to .437 (by direct we mean the R^2 that would result if we only included Water Scarcity Index as the sole variable in the regression model.) If the Water Scarcity Index is removed from the model, the R^2 is reduced slightly, to 72%. So, from this perspective, the predictive power of water scarcity, given the rest of the covariates, although significant, is low.

The ANOVA analysis given next provides an indication of the variability explained.

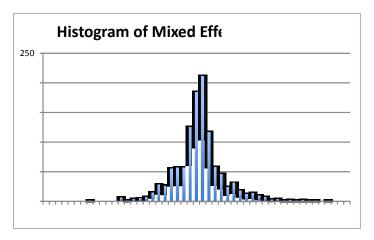
The output of Model 2, once again with p-values smaller than 10^{-6} is summarized in the next tables, first we show the coefficient estimates (once again note that the signs are in agreement with the fundamental interpretation of the variables),

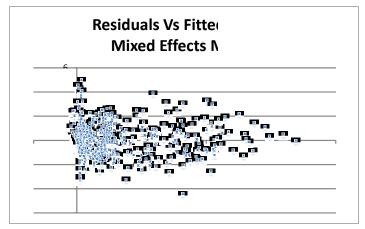
Model 2	Estimate	SE	tStat
(Intercept)	-424.5	382.4	-1.1
Water Scarcity In.	2.93	0.65	4.54
Undergrnd Cap.	2.31	0.25	9.17
Open Pit Cap.	2.35	0.17	13.94
Net Income	1.53	0.22	7.07
CAPEX	-5.54	0.55	-9.99
Net Debt	-1.86	0.11	-15.99

and the variance parameters of the random effects are given next:

Model 2	Estimate	Lower	Upper
Intercept	2335.8	1862.6	2929.2
Residuals	0.91	0.88	0.95

Finally, we show the behavior of the residuals, which show normal-like characteristics and relatively little structure, thus reassuring that normality can be used for prediction in the quantitative risk model.





4. Robustification

Ultimately, we will use the model for performance analysis, including risk measures, via Monte Carlo simulation. Our goal in this section is to explain how to estimate potential errors which arise from our model selection.

If we ultimately measure the risk associated to a portfolio of assets, then the underlying model is the probability distribution induced for the variable L, which represents the losses of a given portfolio of mines during a given time horizon. In turn, such a distribution is implicitly obtained by the large model described in the previous paragraphs. One might compute, for example, the Value-at-Risk (VaR) of the portfolio loss, L by running Monte Carlo simulations. If the associated distribution differs from reality, then the output of the simulation, namely the estimate of VaR, will generally be incorrect. The difference between the correct VaR and the incorrect VaR is, in this case, the impact of the incorrect model assumption.

Note that in our framework, any stochastic simulation comprises two types of errors:

1) Stochastic error due to the generation of only a finite number of replications or scenarios, which is only an approximation of any given probabilistic model (which may correspond

to an infinite number of scenarios),

2) Model error due to a misspecification of the probabilistic model away from reality, as described above.

In turn, the error of type 2) might occur because one assumes a "correct" parametric family of models (say, for example, multivariate Gaussian random variables) and such family is estimated from finitely many samples and therefore is subject to statistical error. Or, perhaps the parametric family is simply incorrect and therefore, inference based on such family will likely be incorrect. Even if these two situations are conceptually different, they both are treated in the same way in our setting, namely, for whatever reason there is a misspecification of the probabilistic model relative to reality. Moreover, even if one takes a non-parametric approach, still there will be error of type 2) in the sense that the probabilistic model will differ from reality.

We also note that an incorrect model assumption can arise for reasons other than the lack of data. One instance is non-stationarity.

Our focus is on the error of type 2); typically the error of type 1) can be addressed by increasing the number of scenarios generated.

Our methodology aims to tackle a large class of uncertainties contributing to error of type 2) under a unified framework. It borrows ideas from the active research area of robust modeling. Techniques in this area have been developed by multiple communities, including the work of the Nobel laureates Hansen and Sargent, in the area of economics.

5. MATHEMATICAL DESCRIPTION

To describe our method mathematically, let us use X to denote a random variable (or random vector) governed by the probability model P. The outcome of X describes a given scenario in our Monte Carlo simulation. Think of X as a risk factor (or risk factors with a given joint distribution).

Let h(X) be a function of the risk factors, X, for example, the loss of an insurance portfolio at the end of a the year given the aggregated claims X. The target risk quantity of interest is E[h(X)], the expected value of h(X) under P.

The input model is typically specified to be some parametric model, say P_0 . In this case we write $E_0[h(X)]$ as the target quantity using the model P_0 . For instance, our current setting, P_0 denotes the simple model we fitted in earlier sections fit using the financial and water risk data; $h(\cdot)$ is composition of a portfolio, and X is a vector representing the values of the companies, and L = h(X). We can also handle quantiles, but we shall keep the discussion simple for the moment.

Model error arises when the true model, governed by the probability distribution P_{true} , differs from P_0 . To robustify our output, we proceed as follows. First, instead of specifying P_0 as our input, we list our available (partial) information on P_{true} via constraints.

Two examples of constraints that can be used to describe the uncertainty in the underlying model are:

1. Statistical discrepancy: suppose that the modeler has adopted a particular P_0 , but he/she acknowledges a non-zero distance of P_0 from the truth P_{true} . The difference between these two distributions can be represented via the notion of statistical discrepancy, i.e. a notion that differentiates between probability distributions and is not restricted to a fixed parametric class. One common example of statistical discrepancy

is the Kullback-Leibler (KL) divergence, also known as the relative entropy. This is denoted

$$D(P_{true}||P_0) = E_{true} \left[\log \frac{dP_{true}}{dP_0} \right] = E_0 \left[\frac{dP_{true}}{dP_0} \log \frac{dP_{true}}{dP_0} \right]$$

where dP_{true}/dP_0 denotes the likelihood ratio between the distributions P_{true} and P_0 (examples will be shown later to demonstrate, when densities exist dP_{true}/dP_0 is simply the ratio of the densities). The modeler might have a way to estimate an upper bound for $\eta_{true} := D(P_{true}||P_0)$. We shall discuss how to attempt to estimate such an upper bound η_* in the sequel. If $\eta_* > \eta_{true}$ one can, as we shall explain momentarily, systematically use η_* to find a P_* such that $E_*h(X) \ge E_{true}h(X)$, where P_* is the worst-case measure associated to robustifying using the relative entropy ball around P_0 with tolerance η_* .

2. Moment constraints: the modeler can also choose to specify only some bounds on moments for X. For instance, the modeler might know (or assume) that $\underline{\mu} \leq E_{true}[X] \leq \overline{\mu}$ and $\underline{\sigma} \leq E_{true}[X^2] \leq \overline{\sigma}$ for some $\mu, \overline{\mu}, \underline{\sigma}, \overline{\sigma}$.

In the sequel we use P_f as a generic probability model to be found and E_f is the associated expectation. We will solve optimization problems where the variable to optimize is a probability distribution – so these will be infinite dimensional optimization problems and P_f is precisely the variable to optimize.

Once the "uncertainty region" in the model is specified in a non-parametric way, the next step (Step 2 in Algorithm 0) is to set up an optimization problem. We wish to solve

$$\max_{P_f \in \mathcal{A}} E_f[h(X)] \quad \text{or} \quad \min_{P_f \in \mathcal{A}} E_f[h(X)] \tag{2}$$

The max or the min depends on the meaning of being "worst-case" or "conservative estimate". For Example, the max formulation is typically the relevant one in the risk analysis setting. As noted earlier, the optimization problem (2) has decision variable P_f , which is unknown but is believed to lie in the feasible region \mathcal{A} . Corresponding to the two examples of constraints above, \mathcal{A} can be equal to $\{P_f : D(P_f || P_0) \leq \eta\}$ as a statistical discrepancy based constraint, or $\{P_f : \underline{\mu} \leq E_f[X] \leq \overline{\mu}, \underline{\sigma} \leq E_f[X^2] \leq \overline{\sigma}\}$ as moment constraints. We emphasize that there are many other ways of encoding uncertainty regions \mathcal{A} .

Ultimately, since we do not wish to impose uncertainty regions \mathcal{A} that induce a specific class of parametric models – otherwise we do not allow for enough possible flexibility when quantifying for model uncertainty – we end up with very challenging optimization problems. So, our selection of \mathcal{A} ends up being infinite dimensional and it is important then to have model formulations of uncertainty quantification which are tractable.

Consequently, there are two challenges regarding (2): 1) Solving (2) (technical challenge), and 2) calibrating the necessary parameters in the constraints, e.g. $\eta, \underline{\mu}, \overline{\mu}, \underline{\sigma}, \overline{\sigma}$ – which we understand more as a practical challenge.

6. Solving the Worst-Case Optimization

6.1. KL Constraint.

We consider the worst-case optimization under KL constraint:

$$\max E_f[h(X)]$$
 subject to $D(P_f||P_0) \le \eta$ (3)

(To avoid redundancy, we focus on the maximization problem here).

For simplicity, we shall assume here that model P_0 has a density $f_0(x)$, so that

$$E_0 h(X) = \int h(x) f_0(x) dx.$$

Under this assumption, the solution to the optimization (3) is given by the following: Let $h^* = \max_x h(x)$, and assume that

$$p^* = P_0(h(X) = h^*) > 0.$$

If $-\log p^* < \eta$, then the optimal solution to (3) is a probability model P_* with density $f_*(x)$ such that

$$f_*(x) = \frac{f_0(x)}{p^*} I(h(x) = h^*),$$

where $I(\cdot)$ denotes the indicator function, and the optimal value is h^* . In other words, $f_*(x) = f_0(x|h(X) = h^*)$, is the conditional density of X given that $h(X) = h^*$.

Now, if $-\log p^* \ge \eta$, then the optimal solution is given by the density $f_*(x)$ such that

$$f_*(x) = \frac{e^{\beta h(x)} f_0(x)}{E_0[e^{\beta h(X)}]},\tag{4}$$

where $\beta > 0$ is the root of the equation

$$\frac{\beta E_0[h(X)e^{\beta h(X)}]}{E_0[e^{\beta h(X)}]} - \log E_0[e^{\beta h(X)}] = \eta$$
 (5)

In many cases, the expressions (4) and (5) cannot be written in closed-form. However, we can replace the above solutions with Monte Carlo simulation estimators. It is very important to note that the optimal solution of the optimization problem, that is, the worst-case density f_* , is given fully in terms of f_0 , the assumed / baseline model. So, in order to obtain β and p_* one can actually use Monte Carlo simulation.

The procedure in which we replace the theoretical distributions by Monte Carlo samples is shown in Algorithm 6.1, which follows next.

ALGORITHM

Procedure for solving (3)

Initialization: the baseline input density f_0 , and threshold level η .

Procedure:

- 1. Generate m samples from f_0 , say they are Y_1, \ldots, Y_m .
- **2.** Let $\operatorname{argmax}\{j:h(Y_j)\}\$ be the set of indices in $\{1,\ldots,m\}$ that has the maximum $h(Y_j)$. Let $|\operatorname{argmax}\{j:h(Y_j)\}|$ be the cardinality, i.e. number of elements, of $\operatorname{argmax}\{j:h(Y_j)\}$. Then

If
$$-\log \frac{|\operatorname{argmax}\{j:h(Y_j)\}|}{m} \le \eta$$
, then

$$w_i = \begin{cases} \frac{1}{|\operatorname{argmax}\{j:h(Y_j)\}|} & \text{for } i \in \operatorname{argmax}\{j:h(Y_j)\}\\ 0 & \text{otherwise} \end{cases}$$
 (6)

If $-\log \frac{|\operatorname{argmax}\{j:h(Y_j)\}|}{m} > \eta$, then

$$w_i = \frac{e^{\beta h(Y_i)}}{\sum_{i=1}^n e^{\beta h(Y_i)}} \tag{7}$$

for all i, where $\beta > 0$ satisfies

$$\frac{\beta \sum_{i} h(Y_i) e^{\beta h(Y_i)}}{\sum_{i} e^{\beta h(Y_i)}} - \log \left(\frac{1}{m} \sum_{i} e^{\beta h(Y_i)} \right) = \eta \tag{8}$$

3. Output

$$\sum_{i=1}^{m} h(Y_i)w_i$$

- 1. Interpretation: The output of the procedure is an approximation to the optimal value of (3). The vector $(w_i)_{i=1,...,m}$ approximates the worst-case probability distribution on the support $(Y_i)_{i=1,...,m}$ that achieves this optimal value. When m is large, the procedural output becomes very close to the actual optimal value.
- 2. Comparison with standard procedure: The initial step involves a sampling of Y_1, \ldots, Y_m from P_0 . The difference is the weight put on each of these samples. In standard procedure, the weight is 1/m for each sample. In the worst-case procedure, the weights are computed using Step 2 in Algorithm 6.1 that automatically leans towards the more risky scenario.
- 3. Computational efficiency: the most involved step in our robustification algorithm is only a one-dimensional line search for the β , thus this procedures is easy to implement.
- 4. Minimization counterpart: For minimization problem, the only changes in the procedures is that $\operatorname{argmax}\{j:h(Y_j)\}$ is replaced by $\operatorname{argmin}\{j:h(Y_j)\}$ and $\beta<0$ is replaced by $\beta>0$.

6.2. Moment Constraints.

We now consider the worst-case optimization under moment constraints:

$$\max_{\substack{E_f[b(X)]\\ \text{subject to}}} E_f[v_i(X)] \leq \alpha_i, \ i = 1, \dots, s\\ E_f[v_i(X)] = \alpha_i, \ i = s + 1, \dots, M$$

$$(9)$$

where again we focus on the maximization problem here. This is a general formulation that has M moment constraints, and $v_i(\cdot)$ can represent any function. For instance, we can put $v_1(x) = x$ and $v_2(x) = -x$, $v_3(x) = x^2$, $v_4(x) = -x^2$, and $\alpha_1 = \overline{\mu}$, $\alpha_2 = -\underline{\mu}$, $\alpha_3 = \overline{\sigma}$, $\alpha_4 = -\underline{\sigma}$, and all constraints are inequalities. Some of the information on the moment constraints might come from a more detailed (bottom-up) model or from models which are more detailed for some portions of the overall model, for instance, based on real option valuation – see Section 8.2.

The procedure for solving (9) is shown in Algorithm 6.2, which follows next.

ALGORITHM

Generalized linear programming procedure for solving (9)

Initialization: an arbitrary probability distribution on the support $\{x_1, \ldots, x_L\}$, where $L \leq M+1$, that lies in the feasible region in (9). Set $\tau = L$.

Procedure: For each iteration k = 1, 2, ..., given $\{x_1, ..., x_{\tau}\}$:

1. (Master problem solution) Solve

max
$$\sum_{j=1}^{\tau} h(x_j) p_j$$
subject to
$$\sum_{j=1}^{\tau} v_i(x_j) p_j \leq \alpha_i, \ i = 1, \dots, s$$
$$\sum_{j=1}^{\tau} v_i(x_j) p_j = \alpha_i, \ i = s+1, \dots, M$$
$$\sum_{j=1}^{\tau} p_j = 1$$
$$p_j \geq 0, \ j = 1, \dots, \tau$$

Let $\{p_1^k, \dots, p_{\tau}^k\}$ be the optimal solution. Find the dual multipliers $\{\theta^k, \pi_1^k, \dots, \pi_M^k\}$ that satisfy

$$\theta^{k} + \sum_{i=1}^{M} \pi_{i}^{k} v_{i}(x_{j}) = v_{0}(x_{j}), \text{ if } p_{j} > 0, j = 1, \dots, \tau$$

$$\theta^{k} + \sum_{i=1}^{M} \pi_{i}^{k} v_{i}(x_{j}) \geq v_{0}(x_{j}), \text{ if } p_{j} = 0, j = 1, \dots, \tau$$

$$\pi_{i}^{k} \geq 0, i = 1, \dots, s$$

2. (Subproblem solution) Find $x_{\tau+1}$ that maximizes

$$\rho(x; \theta^k, \pi_1^k, \dots, \pi_M^k) = h(x) - \theta^k - \sum_{i=1}^M \pi_i^k v_i(x)$$

If $\rho(x_{\tau+1}; \theta^k, \pi_1^k, \dots, \pi_M^k) > 0$, then let $\tau = \tau + 1$; otherwise, stop the procedure and $\{x_1, \dots, x_\tau\}$ are the optimal support points, with $\{p_1^k, \dots, p_\tau^k\}$ the associated weights.

After the last iteration, output

$$\sum_{j=1}^{\tau} h(x_j) p_j^k$$

- 1. Interpretation: The output of the procedure is an exact optimal value of (9). The worst-case probability distribution is a finite-support discrete distribution on $\{x_1, \ldots, x_{\tau}\}$ with weights $\{p_1^k, \ldots, p_{\tau}^k\}$ obtained in the last iteration.
- 2. Comparison with standard and KL-constrained procedure: Unlike the previous procedures, the formulation (9) does not have a baseline input distribution to begin with.
- 3. Computational efficiency: Step 1 in each iteration of Algorithm 6.2 can be carried out by standard linear programming solver, which can output both the optimal $\{p_j\}$ and the dual multipliers $\{\theta^k, \pi_1^k, \ldots, \pi_M^k\}$. Step 2 is a one-dimensional line search if X is one-dimensional.
- 4. Minimization counterpart: For minimization problem, simply replace h by -h in the whole procedure of Algorithm 6.2 except in the last output $\sum_{i=1}^{\tau} h(x_i) p_i^k$.

7. Calibrating the Constraints

For KL constraint, the quantity η captures the distance between the baseline model and the truth. When data is available, η can be estimated by an empirical version of the KL divergence. There are several methods available for estimation of KL divergence. Here, we present a direct approach based on first principles as follows. First, organize the data into a histogram with bins say $(a_0, a_1], (a_1, a_2], \ldots, (a_{N-1}, a_N]$, where a_0 and a_N are the end points of the distribution support (potentially $-\infty$ or ∞). The relative frequency of the data at

each bin is given by $\hat{P}(a_i < X \leq a_{i+1})$. Moreover, calculate $P_0(a_i < X \leq a_{i+1})$ for the baseline distribution P_0 . Then a calibrated choice of η is

$$\eta = \sum_{i=1}^{N-1} \hat{P}(a_i < X \le a_{i+1}) \log \frac{\hat{P}(a_i < X \le a_{i+1})}{P_0(a_i < X \le a_{i+1})}$$
(10)

The bins can be taken to be equal length except the lowest and the highest ones.

For moment constraints, the quantities such as $\underline{\mu}, \overline{\mu}, \underline{\sigma}, \overline{\sigma}$ can be easily calibrated by using the confidence intervals (CIs) for the respective moments. For instance $\underline{\mu}, \overline{\mu}$ can be the $1-\alpha$ CI for the mean of X, and $\underline{\sigma}, \overline{\sigma}$ can be the CI for the second moment. Note that when 2 pairs of moment constraints are placed, Bonferroni correction needs to be made. For instance, if one is interested in calibrating a moment set such that with 95% chance it contains the truth, then the individual α should be taken as 0.05/2.

8. Ongoing Research: Bayesian Framework and Real Options for Model Enrichment

8.1. Bayesian Modelling Framework.

We propose using Bayesian Hierarchical models as an alternative to current Hierarchical modelling structure. Given the high dimensional nature of the covariates we wish to observe, and the relatively sparse data set, Bayesian modelling provides a natural method of shrinkage and incorporating prior beliefs into our modelling.

8.1.1. Bayesian Models.

In essence, a Bayesian model is a statistical model in which the parameters are assumed to random variables with a distribution specified by the modeler. Once we observe data, we are able to specify a new distribution for the parameters, conditional on the data (using Bayes' Theorem), called a posterior distribution.

A simple example would be the estimation of a sample mean. Say we some data, $X_1, ..., X_n$, and sample mean μ , with the following distributional assumptions:

$$X_i \sim N(\mu, 1)$$

 $\mu \sim N(0, 1)$

The likelihood of the data given the parameter μ can be written as:

$$P(X_1, ..., X_n | \mu) = \frac{1}{(2\pi)^{n/2}} \exp(-\frac{(X_i - \mu)^2}{2})$$

While the distribution of μ can be written:

$$P(\mu) = \frac{1}{(2\pi)} \exp(-\frac{\mu^2}{2})$$

Using Bayes' theorem, we have:

$$P(\mu|X_1,...,X_n) = \frac{P(X_1,...,X_n|\mu)P(\mu)}{\int P(X_1,...,X_n|\mu)P(\mu)d\mu}$$

Which is the posterior distribution for μ that we are interested in. Omitting the calculations, we can find that:

 $\mu|X_1,...,X_n \sim N(\frac{\sum X_i}{n+1},(n+1)^{-1})$

We now have a way of estimating the parameter of interest using, typically, the posterior expectation, and a way of quantifying our uncertainty using the posterior distribution. For further discussion on Bayesian statistics, see the discussion in Chapter 2 of [5].

8.1.2. Bayesian Hierarchical Models.

The Bayesian framework turns out to be highly flexible. We do not have to specify particular quantities for the hyperparameters (namely, the parameters of the distribution of the model parameters), rather, we can specify distributions for those as well. Using precisely the same Bayesian framework, we can then find a posterior distribution for the parameters of interest. We modify our original model, like so:

$$X_1, ..., X_n \sim N(\mu, 1)$$

 $\mu \sim N(\gamma, 1)$
 $\gamma \sim N(0, 1)$

So we can write our posterior for μ and γ as:

$$P(\mu, \gamma | X_1, ..., X_n) = \frac{P(X_1, ..., X_n | \mu, \gamma) P(\mu | \gamma) P(\gamma)}{\int \int P(X_1, ..., X_n | \mu, \gamma) P(\mu | \gamma) P(\gamma) d\mu d\gamma}$$

Typically, calculating the integral in the denominator is very challenging, and we are not able to come up with closed-form solutions for the posterior. For further discussion, see Chapter 5 of [5]. However, since, we know the posterior up to a constant of proportionality (the value of the integral), we are able to utilize Markov Chain Monte Carlo (MCMC) algorithms to sample from the posterior distribution, as in [4] and [11]. The STAN programming language provides a very flexible and stable environment for writing Bayesian Hierarchical models and then sampling efficiently from the posterior distribution.

8.1.3. Example Model for Mining Data.

We are currently in the process of implementing a model for the market prices of mining companies given a set of covariates. In this framework, we can increase the number of parameters of interest to reflect certain economic realities, while more efficiently managing the downsides of model complexity. For example, we could model $Y_i(t)$, the market capitalization of company i at time t as a linear combination the reserves of its holdings (multiplied by the price of gold), broken down by country and mine type, and the value of the company's debt.

$$Y_i(t) \sim N(\sum_{j=1}^{N_{Country}} (\beta_j, \text{Underground}UG_{i,j}, \text{Underground} + \beta_j \text{Open Pit}OP_{i,j} + \gamma_i D_i(t))$$

 β will consist of $N_{Country}$ independent samples from a 2-dimensional normal distribution with correlation ρ , and γ will be $N_{Company}$ samples from a normal distribution. We assume that the mean of γ will be 1 for economic reasons. Since we do not have much of an understanding of the hyperparameters, we generally give them non-informative (very flat and

wide) priors. Ultimately, we plan to select more precise hyperparameters using information provided from a set of financial analysis reports.

$$\beta \sim N(\mu, \sigma_1^2(\frac{1}{\rho}, \frac{\rho}{1}))$$

 $\mu \sim N(0, 100)$

 $\sigma_1^2 \sim \text{InvGamma}(.001, .001)$

 $\rho \sim \text{Unif}(0,1)$

 $\gamma \sim N(1, \sigma_2^2)$

 $\sigma_2^2 \sim N(0, \sigma_2^2)$

8.2. Real Option Methodology.

For a given time period (with length dt), the owner of a mine (with production capacity q per unit time, with Q total amount of mineral, and cost C(Q,t)), has the choice to operate the mine, earning the value of the mineral extracted S less the costs, q(S(t) - C(Q,t)dt), keep the mine closed, paying a maintenance cost for the facility -Mdt, or abandon the facility altogether, incurring no future costs. At each time period, based on the amount of mineral left in the mine, and the amount the level of mineral price, the mine owner will make his decision in order to maximize the expected value of the future discounted cash flows, much like an American Option. Properties that have not yet been developed can be valued used very similar methodologies. Traditionally, this problem was formulated for the Black-Scholes model of asset prices, namely that log-returns are normally distributed with constant variance, and the price could be found by solving a partial differential equation in S and Q. We proposed to use the simulation methods developed as a generalization to [9] in [12].

Many implementations of these models (such as [8]) include a flat cost curve. We hope to build on the costs models developed in [10] and [3], to incorporate a random structure to includes the well-known factors (mine technology, mine resources, resource quantile, mine capacity), as well as environmental factors that can be included in our simulation. The user should be able to specify a water risk regime and then value the mine under that set of environmental scenarios, or look at the average cost across all the scenarios once probabilities have been estimated. In addition to having a stochastic cost regime, interest rates, convenience yields, and so forth can be easily added with little to no extra computational effort. Moreover, we can work with richer models of asset prices, to account for the fatter tails that can be found in actual time series of returns.

Once all the mines of interest have been priced, their values can be placed into the Bayesian regression framework and appropriate discount factors can be estimated for each country (to control for external risks not accounted for in the option model).

References

[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] Jose Blanchet, Henry Lam, Qihe Tang, and Zhongyi Yuan. Robust stochastic performance analysis with insurance applications. *In preparation*, 2014.
- [3] H.F. Campbell and D.L. Wrean. Deriving the long-run supply curve for a competitive mining industry: The case of saskatchewan uranium. In A. Scott, editor, *Progress in Natural Resource Economics*. Clarendon Press, Oxford, 1985.
- [4] Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Roweth Duncan. Hybrid monte carlo. *Physics Letters B*, 195 (2):216–222, 1987.
- [5] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL, 2013.
- [6] Andrew Gelman and Jennifer Hill. Data Analysis Using Regression Multilevel/Hierarchical Modelling. Cambridge University Press, New York, 2006.
- [7] Lars Peter Hansen and Thomas J. Sargent. Robust control and model uncertainty. *The American Economic Review*, 91(2):pp. 60–66, 2001.
- [8] T. Henker, DB Cowell, and KY Fong. Real options valuation of australian gold mines and mining companies. *Journal of Alternative Investments*, 6:23–38, 2003.
- [9] Francis Longstaff and Eduardo Schwartz. Valuing american options by simulation: A simple least-squares approach. *The Review of Financial Studies*, 14 (1):113–147, 2001.
- [10] Alberto Moel and Peter Tufano. When are real options exercised? an empirical study of mine closings. The Review of Financial Studies, 15 (1):35–64, 2002.
- [11] Radford Neal. MCMC using hamiltonian dynamics. In Galin L. Jones Andrew Gleman and Xiao-Li Meng, editors, *Handbook of Markov Chain Monte Carlo*, pages 113–161. Chapman and Hall/CRC, Boca Raton, FL, 2011.
- [12] Sabry Abdel Sabour and Richard Poulin. Valuing real capital investments using the least-squares monte carlo method. *The Engineering Economist*, 51:141–160, 2006.