Robust risk measurement and model risk
PAUL GLASSERMAN and XINGBO XU
Proofs & Derivations

Prithwish Maiti Zhijiang Yang Zhao Qu

2. Overview of the approach

Main Idea : If the law of X is correctly specified, then the expectation E[V(X)] is the true value of the risk measure of interest.
The author introduces model uncertainty by acknowledging that the law of X may be mis-specified. We consider alternative probability laws that are not too far from the nominal law in a sense quantified by relative entropy. This procedure is what the author Defines to be **robust risk measurement**.
The author seeks to evaluate the , the bounds

$$\inf_{m \in P_\eta} E[m(X)V(X)] \;\;\&\;\; \sup_{m \in P_\eta} E[m(X)V(X)]$$

where $E[m(X)V(X)]$ is the alternative risk measure incorporating the Change of measure m(x).
Note: For a **maximization problem**, the dual can be formed by subtracting the constraint equation using a **positive Lagrange multiplier**. Conversely, for a **minimization problem**, the dual is constructed by subtracting the constraint equation using a **negative multiplier**.

Derivation : Solution of inner supremum of the dual problem.

We have to consider another constraint apart from $\theta > 0$, $E(m) = 1$, because

$$\int \frac{\tilde{f}(x)}{f(x)} f(x)dx = \int \tilde{f}(x)dx = 1$$

Proof. We take the Lagrangian of the objective equation in the dual of (1)

$$\implies E\left[mV(X) - \frac{1}{\theta}(m\log m - \eta)\right] - \lambda(E(m) - 1)$$

$$\implies E\left[mV(X) - \frac{1}{\theta}(m\log m) - \lambda m\right] + \frac{\eta}{\theta} + \lambda$$

Only the Expectation term contains m, Hence, taking the derivative of it w.r.t m

$$\implies \frac{\partial}{\partial m} E\left[mV(X) - \frac{1}{\theta}(m\log m) - \lambda m\right] = 0$$

$$\implies E\left[\theta V(X) - \log m - 1 - \theta\lambda\right] = 0$$

For this equation to be 0, we have to set the expression inside the Expectation to be zero as, we have taken an arbitrary nominal density f(x)

$$\implies \theta V(X) - \log m - 1 - \theta\lambda = 0$$

$$\implies m = \frac{\exp(\theta V(X))}{\exp(1+\theta\lambda)}$$

To get the value of λ, we use $E(m) = 1$

$$\implies E\left[\frac{\exp(\theta V(X))}{\exp(1+\theta\lambda)}\right] = 1 \implies \exp(1+\theta\lambda) = E[\exp(\theta V(X))]$$

$$\implies m_\theta^* = \frac{\exp(\theta V(X))}{E[\exp(\theta V(X))]} \tag{1}$$

In other words, the worst-case model error is characterized by an exponential change of measure defined through the function V and a parameter $\theta > 0$.
From the above note, The lower bound is solved the same way but with $theta < 0$. □

2.1. A first example: portfolio variance

Here, the author introduces an example of portfolio variance as the nominal risk measure, where there is a linear allotment of weights to n stocks.
Using the above derived solution of the likelihood ratio m_θ^* we try to find the alternative density in the worst case-

Derivation : Portfolio Variance - Worst case density

Proof. We have to use the fact that the 2 inner products appearing in the expression of m_θ^* are interchangeable ie, communicative.

$$\implies m_\theta^* \propto \exp\left(\theta\left[a^\top(X-\mu)(X-\mu)^\top a\right]\right) = \exp\left(\theta\left[(X-\mu)^\top aa^\top(X-\mu)\right]\right)$$

$$\implies \tilde{f}(X) \propto \exp\left(\theta\left[(X-\mu)^\top aa^\top(X-\mu)\right]\right) \times \exp\left(-\frac{1}{2}\left[(X-\mu)^\top\Sigma^{-1}(X-\mu)\right]\right)$$

$$\implies \tilde{f}(X) \propto \exp\left(-\frac{1}{2}\left[(X-\mu)^\top(\Sigma^{-1} - 2\theta aa^\top)(X-\mu)\right]\right)$$

This is the kernel of a multivariate normal distribution with covariance matrix $\tilde{\Sigma} = (\Sigma^{-1} - 2\theta aa^\top)^{-1}$. □

For small θ, $\tilde{\Sigma} = \Sigma + 2\theta\Sigma aa^\top\Sigma + o(\theta^2)$

Proof. The maclaurin series $y = 1/(b-x)$ is-

$$y = b^{-1} + xb^{-2} + x^{-2}b^{-3} + x^{-3}b^{-4}\ldots$$

similar expression is defined for a matrix notation, The maclaurin series for $Y = (B-X)^{-1}$ is-

$$Y = B^{-1} + B^{-1}X(B^{-1})^\top\ldots$$

Taking $B = \Sigma^{-1}$and$X = 2\theta aa^\top$, we get the result by substituting $B^{-1} = \Sigma$ □

Conclusion: The resulting worst-case variance of the portfolio is increased by approximately 2θ times the square of the original variance
The worst-case change in distribution we found in this example depends on the portfolio vector a.
The author tries to interpret model error as the work of a malicious adversary. The adversary perturbs our original model, but the error introduced by the adversary is not arbitrary—it is tailored to have the most severe impact possible, subject to a entropy constraint which he call entropy budget.

Derivation : Generalized Quadratic function - Worst case density

The portfolio variance example generalizes to any quadratic function $V(x) = x^\top Ax + B$

Proof. This proof can be done by comparing terms of equal degrees on both sides of this equations. The left hand side contains $m^*(\theta) \times f(X)$ and the right hand side contains the alternative density $\tilde{f}(X)$

$$\implies \exp(\theta x^\top Ax) \times \exp\left(-\frac{1}{2}\left[(X-\mu)^\top \Sigma^{-1}(X-\mu)\right]\right) = \exp\left(-\frac{1}{2}\left[(X-\tilde{\mu})^\top \tilde{\Sigma}^{-1}(X-\tilde{\mu})\right]\right)$$

$\implies$ This boils down to comparing -

$$\implies x^\top(\Sigma^{-1} - 2\theta A)x - x^\top\Sigma^{-1}\mu - \mu^\top\Sigma^{-1}x + C_1 = x^\top\tilde{\Sigma}^{-1}x - x^\top\tilde{\Sigma}^{-1}\tilde{\mu} - \tilde{\mu}^\top\tilde{\Sigma}^{-1}x + C_2$$

$$\implies \tilde{\Sigma}^{-1} = \Sigma^{-1} - 2\theta A \ \ \& \ \ \Sigma^{-1}\mu = \tilde{\Sigma}^{-1}\tilde{\mu}$$

$$\implies \tilde{\Sigma} = \left(\Sigma^{-1} - 2\theta A\right)^{-1} \ \ \& \ \ \tilde{\mu} = \tilde{\Sigma}\Sigma^{-1}\mu$$

Presuming that the end result will be Multi-variate normal gives the end result easily but is not full proof. □

2.2. Optimization problems and precise conditions

In the previous section we signified our model of portfolio variance by taking the vector of portfolio weight a. In this section, the author suggest a parametric function(taking parameters from a set) to evaluate the risk measure. Now our risk measure becomes -

$$\inf_{a\in A} E[V_a(X)] \tag{2}$$

This interprets to the operation of Choosing the model parameters such that we minimise the risk. The **robust version** of the optimization problem(upper bound) becomes-

$$\inf_{a\in A} \sup_{m\in P_\eta} E[mV_a(X)] \tag{3}$$

We go through the same process of hypothetical adversary imposing model error. We sought to find the worst case model risk in a **robust** fashion.

$$\inf_{a} \inf_{\theta>0} \sup_{m} E\left[mV_a(X) - \frac{1}{\theta}\left(m\log m - \eta\right)\right] \tag{4}$$

Derivation : Proof of Proposition 2.1

This discuss the flipping of the 2 inf operations. For that, we first need to go through Assumptions A.1 & A.2

Assumption A.1

(1) The decision parameter set A is compact, i.e, Va(x) is convex in a for any x. If it is compact then $-\infty < V_a(X) < \infty \, \forall \, a \implies -\infty < E[V_a(X)] < \infty \, \forall \, a$.
Hence $\inf_a E[V_a(X)]$ exists
(2) For all $a \in A$, the moment generating function $\hat{F}_a(\theta) = E[exp(\theta V_a(X))]$ exists for θ in some open set containing the origin.
We denote such an interval by $(\theta_{min}(a), \theta_{max}(a))$ the interval (possibly infinite) in which $\hat{F}_a(\theta)$ is finite and thus an exponential change of measure defined by $\exp(\theta V(X))$ is well defined.
For any $\theta > 0$ and decision parameter a, if $E[exp(\theta V_a(X))]$ exits then the optimal solution of the inner supremum in (4) is-

$$m^*_{\theta,a} = \frac{\exp(\theta V_a(X))}{E[\exp(\theta V_a(X))]} \quad \theta \in (0, \theta_{max}(a))$$

This is similar to the derivation of the inner supremum done in section 2. Substituting it back to (4), we get-

$\implies$ Let $E[\exp(\theta V_a(X))] = K$ for the ease of writing $\therefore m^*_{\theta,a} = \exp(\theta V_a(X))/K$

$$\implies E\left[\frac{\exp(\theta V_a(X))}{K} V_a(X) - \frac{1}{\theta}\left(\frac{\exp(\theta V_a(X))}{K} \log \frac{\exp(\theta V_a(X))}{K} - \eta\right)\right]$$

$$\implies E\left[\frac{\exp(\theta V_a(X))}{K} V_a(X) - \frac{1}{\theta}\left(\frac{\exp(\theta V_a(X))}{K} (\theta V_a(X) - \log K) - \eta\right)\right]$$

$$\implies E\left[\frac{\exp(\theta V_a(X))}{K} V_a(X) - \frac{\exp(\theta V_a(X))}{K} (V_a(X) - \frac{\log K}{\theta}) + \frac{\eta}{\theta}\right]$$

$$\implies E\left[\frac{\exp(\theta V_a(X))}{K} \frac{\log K}{\theta}\right] + \frac{\eta}{\theta}$$

Substituting the value of K back -

$$\implies \frac{\log E[\exp(\theta V_a(X))]}{E[\exp(\theta V_a(X))]\,\theta} E[\exp(\theta V_a(X))] + \frac{\eta}{\theta}$$

$$\implies \frac{1}{\theta} \log E[\exp(\theta V_a(X))] + \frac{\eta}{\theta} \tag{5}$$

The infimum over θ will automatically make the optimal θ smaller than $\theta_{max}(a)$, because as $\theta \to \theta_{max}(a)$, The objective function containing the m.g.f. will go to infinity.
Doubt: I didn't understood this statement Hence we consider $\theta < \infty$ instead of $\theta \in (0, \theta_{max}(a))$. This allows us to change the order of $\inf_a \inf_{\theta>0}$ in (4).

Assumption A.2

To swap the order, we need this assumption.
Did not understood
Assuming we can change the order of the infimum, We can solve the 'sup' over m as done in (5). We are left with the following problem, after solving for m -

$$\inf_{\theta>0} \inf_a \frac{1}{\theta} \log E[\exp(\theta V_a(X))] + \frac{\eta}{\theta} \tag{6}$$

Now we write optimal $a(\theta)$ for the objective function-

$$a^*(\theta) = \arg\inf_a \frac{1}{\theta} \log E[\exp(\theta V_a(X))] \tag{7}$$

Hence the inner inf sup becomes -

$$H(\theta) + \frac{\eta}{\theta} = \frac{1}{\theta} \log E[\exp(\theta V_{a^*(\theta)}(X))] + \frac{\eta}{\theta}$$

So essentially, we are solving the objective backwards reducing the variables 1 at a time. First we find m in terms of θ, hence eliminate m. Next we eliminate a by finding its optimum expression in terms of θ. Finally We have only an objective in only 1 variable θ.
The other emphasis on the value of the expression above one theta tends to zero. This is done to evaluate the objective function at $\theta = 0$

$$\lim_{\theta \to 0^+} \frac{1}{\theta} \log E[\exp(\theta V_a(X))] = E[V_a(X)]$$

Hence $H(0) = E[V_{a^*(0)}(X)]$
Finally, For given η > 0, we can find an optimal θ^*_η , then plug this in the expression for the optimal a to get $a^*(\theta^*_\eta)$. Lastly plugging both of these to the expression of m^* to get $m^*\left(\theta^*_\eta, a^*(\theta^*_\eta)\right)$.
The **uncertainty upper bound** is reached at the optimal perturbation.

$$\eta = E\left[m^*\left(\theta^*_\eta, a^*(\theta^*_\eta)\right) \log m^*\left(\theta^*_\eta, a^*(\theta^*_\eta)\right)\right]$$

The original constraint problem (2) has the optimal objective, i.e, **worst case error** is -

$$E\left[m^*\left(\theta^*_\eta, a^*(\theta^*_\eta)\right) V_{a^*(\theta^*_\eta)}(X)\right] \tag{8}$$

I don't understand how it differs from the objective function of the penalty version (6) through the constant term.
The author suggest studying the relation between the level of uncertainty and the worst-case error. We can work directly with multiple values of $\theta > 0$ and evaluate $\eta(\theta)$ with each θ, to explore this relationship.

2.3 Robustness with Heavy Tails: Extension to α -Divergence

In this section, we extend the concept of robustness in risk measurement to handle heavy-tailed distributions by using α-divergence, an extension of relative entropy.

Relative Entropy and Heavy Tails

Relative entropy is effective for describing model uncertainty when the tails of the distribution of $V(X)$ are exponentially bounded. For heavy-tailed distributions, where exponential bounding may not hold, α-divergence is introduced as an alternative measure. To understand why adjusting the value of α can control the emphasis on the tails of a distribution, we need to delve into the mathematical definition of α-divergence and its properties.
Definition of α-Divergence: α-divergence (D_α) is defined as:

$$D_\alpha(P\|Q) = \frac{1}{\alpha(1-\alpha)}\left(1 - \int \left(\frac{Q(x)}{P(x)}\right)^\alpha P(x)\,dx\right)$$

where $P(x)$ and $Q(x)$ are two probability density functions.
Effect of Different α Values
Let's see how different values of α influence the emphasis on the tails:
1. $\alpha > 1$: When $\alpha > 1$, α-divergence places more emphasis on the high-density regions of $Q(x)$. This is because $\left(\frac{Q(x)}{P(x)}\right)^\alpha$ will amplify ratios greater than 1, thus emphasizing the main mode (high-probability regions).

2. $\alpha < 1$: When $\alpha < 1$, α-divergence becomes more sensitive to low-density regions of $Q(x)$. For heavy-tailed distributions, this means reducing the influence of outliers (tails). Since $\alpha < 1$, $\left(\frac{Q(x)}{P(x)}\right)^{\alpha}$ will amplify ratios less than 1, diminishing the contribution of the tails to the divergence.
3. $\alpha = 1$: Specifically, when $\alpha \to 0$, α-divergence approximates the KL divergence:

$$D_1(P\|Q) = \lim_{\alpha \to 1} D_\alpha(P\|Q) = \int Q(x) \log \frac{Q(x)}{P(x)}\, dx$$

This focuses more on the density of Q rather than P.

Example

Consider a simple example where $Q(x)$ is a heavy-tailed distribution, and $P(x)$ is a light-tailed distribution. Using KL divergence, the extreme values in the tails (due to the logarithmic term) will heavily influence the divergence. However, using α-divergence with $\alpha < 1$, the influence of the tails is reduced because $\left(\frac{Q(x)}{P(x)}\right)^{\alpha}$ will amplify smaller ratios, thereby decreasing the tail's impact.

Conclusion

By adjusting the value of α, α-divergence can flexibly control the emphasis on the tails of a distribution. When α is large, it emphasizes the main mode (high-density regions); when α is small, it reduces the influence of the tails, making it more robust to heavy-tailed distributions. This flexibility gives α-divergence an advantage in handling heavy-tailed distributions.

Definition of α -Divergence

α-divergence is defined as:

$$D_\alpha(m) = D_\alpha(f, \tilde{f}) = \frac{1 - \int \tilde{f}^{\alpha}(x) f^{1-\alpha}(x)\, dx}{\alpha(1-\alpha)} = \frac{1 - E[m^\alpha]}{\alpha(1-\alpha)},$$

where m is the likelihood ratio $\tilde{f}/f$, and the expectation is taken with respect to f. Relative entropy is a special case of α-divergence, as:

$$R(m) = E[m \log m] = \lim_{\alpha \to 1^+} D_\alpha(m).$$

To derive relative entropy from α-divergence, consider the limit of $D_\alpha(m)$ as α approaches 1 from the right:
1. **Take the Limit as $\alpha \to 1^+$:**

$$R(m) = \lim_{\alpha \to 1^+} D_\alpha(m)$$

2. **L'Hôpital's Rule:** Because the direct substitution of $\alpha = 1$ into $D_\alpha(m)$ results in an indeterminate form $\frac{0}{0}$, we use L'Hôpital's rule to resolve the limit. This requires differentiating the numerator and the denominator with respect to α.

$$\begin{aligned}
&\implies \lim_{\alpha \to 1^+} \frac{\frac{d}{d\alpha}(1 - E[m^\alpha])}{\frac{d}{d\alpha}(\alpha(1-\alpha))} \\
&\implies = \lim_{\alpha \to 1^+} \frac{-E[m^\alpha \log m]}{1 - 2\alpha} \\
&\implies E[m \log m]
\end{aligned}$$

Optimization Problem with α-Divergence

The constraint problem with α-divergence is formulated as:

$$\inf_a \sup_{m:D_\alpha(m)<\eta} E[mV_a(X)].$$

The corresponding penalty problem becomes:

$$\inf_a \inf_{\theta>0} \sup_m E\left[mV_a(X) - \frac{1}{\theta}(D_\alpha(m)-\eta)\right] = \inf_{\theta>0} \inf_a \sup_m E\left[mV_a(X) - \frac{1-m^\alpha}{\theta\alpha(1-\alpha)} + \frac{\eta}{\theta}\right] \tag{9}$$

The supremum is taken over valid likelihood ratios, which are non-negative random variables with mean 1.

Proposition 2.3 and Its Proof

Proposition 2.3 Suppose Assumption A.3 introduced in Appendix A holds. For any $a \in A$, $\theta > 0$ and $\alpha > 1$, the pair $(m^*(\theta,\alpha,a), c(\theta,\alpha,a))$ that solves the following equations with probability 1 is an optimal solution to (9).

$$m^*(\theta,\alpha,a) = (\theta(\alpha-1)V_a(X) + c(\theta,\alpha,a))^{\frac{1}{\alpha-1}}, \tag{10}$$

for some constant $c(\theta,\alpha,a)$, such that:

$$\theta(\alpha-1)V_a(X) + c(\theta,\alpha,a) \geq 0 \tag{11}$$

and

$$E\left[(\theta(\alpha-1)V_a(X) + c(\theta,\alpha,a))^{\frac{1}{\alpha-1}}\right] = 1. \tag{12}$$

Proof

To elaborate on the proof, consider the following steps:
1. Formulation of the Objective, collecting only m terms from (9):

$$J(m) = E\left[mV_a(X) + \frac{1}{\theta\alpha(1-\alpha)}m^\alpha\right] + \frac{\eta}{\theta}$$

2. Construct the Perturbed Likelihood Ratio: $\tilde{m} = (1-t)m^* + tm$. This is done to show that m^* is the optimal solution and we can use the fact that $\tilde{m}$ will only be the optimal solution if it coincides with m^*, i.e, $t = 0$
3. Evaluate the Objective at $\tilde{m}$:

$$K(t) := E\left[((1-t)m^* + tm)V_a + \frac{1}{\theta\alpha(1-\alpha)}((1-t)m^* + tm)^\alpha\right] + \frac{\eta}{\theta},$$

4. Differentiate with Respect to t at $t = 0$:

$$\begin{aligned}\frac{d}{dt}K(t)\bigg|_{t=0} &= E\left[(m-m^*)V_a + \frac{1}{\theta(1-\alpha)}(m^* + t(m-m^*))^{\alpha-1}(m-m^*)\right]\bigg|_{t=0} \\ &= E\left[\left(V_a + \frac{1}{\theta(1-\alpha)}(m^*)^{\alpha-1}\right)(m-m^*)\right]\end{aligned}$$

5. Optimality Condition: For the optimal m^*, this derivative must be zero for any m. For this to happen, we need the term inside braces to be constant.

$$V_a + \frac{1}{\theta(1-\alpha)}(m^*)^{\alpha-1} = \text{constant}$$

6. Solving for m^*: From the above condition, we get:

$$m^*(\theta, \alpha, a) = (\theta(\alpha-1)V_a(X) + c(\theta, \alpha, a))^{\frac{1}{\alpha-1}}$$

where $c(\theta, \alpha, a)$ is a constant. The constraint (11) comes from the fact that $m^*(\theta, \alpha, a)$ is a likelihood ratio. and (12) since $E[m^*] = 1$

Thus, we have shown that under the given assumptions, the pair $(m^*(\theta, \alpha, a), c(\theta, \alpha, a))$ provides the optimal solution for the objective function in (4), satisfying all required conditions.

Inequality (11) constraints: $\theta(\alpha-1)V_a(X) + c(\theta, \alpha, a) \geq 0$

This inequality requires that under all conditions, $\theta(\alpha-1)V_a(X) + c(\theta, \alpha, a)$ must be non-negative.

When $V_a(X)$ is not bounded from below: $V_a(X)$ can become extremely small or even approach negative infinity), then when $\theta > 0$ and $\alpha \geq 0$, this inequality may not be satisfied.

For the case in which the adversary seeks to minimize the objective function (that is, to get the lower bound of the error interval), we need $\alpha < 0$ to satisfy (11).

Feasible Likelihood Ratio Existence

First we show Feasibility at 0 then at near $\theta = 0$:

- Assume there exists a feasible likelihood ratio at $\theta = 0$. At this time, we can choose $c(0, \alpha, a) = 1$ because it satisfies (11) and (12), and:

$$m^*(0, \alpha, a) = c(0, \alpha, a)^{\frac{1}{\alpha-1}}$$

- Through continuity, a set $[0, \theta_0)$ can be found, ensuring that for any $\theta \in [0, \theta_0)$, there exists a $c(\theta, \alpha, a)$ satisfying (11) and (12).

Optimal decision

Once $c(\theta, \alpha, a)$ is found, equation (7) provides an optimal change of measure (not necessarily unique). The optimal decision changes to:

$$a^*(\theta) = \arg\min_a \left(\frac{\alpha-1}{\alpha} E\left[(\theta(\alpha-1)V_a(X) + c(\theta, \alpha, a))^{\frac{1}{\alpha-1}} V_a(X) \right] + \frac{c(\theta, \alpha, a)}{\theta\alpha(1-\alpha)} \right)$$

Proof. The inner decision problem was-

$$a^*(\theta) = \arg\min_a E\left[m^* V_a(X) - \frac{1-(m^*)^\alpha}{\theta\alpha(1-\alpha)} \right] + \frac{\eta}{\theta}$$

$$\implies a^*(\theta) = \arg\min_a E\left[m^* V_a(X) - \frac{1 - (\theta(\alpha-1)V_a(X) + c(\theta, \alpha, a))^{\frac{\alpha}{\alpha-1}}}{\theta\alpha(1-\alpha)} \right] + \frac{\eta}{\theta}$$

$$\implies a^*(\theta) = \arg\min_a E\left[m^* V_a(X) + \frac{(\theta(\alpha-1)V_a(X) + c(\theta, \alpha, a))^{\frac{\alpha}{\alpha-1}}}{\theta\alpha(1-\alpha)} - \frac{1}{\theta(\alpha(1-\alpha))} \right] + \frac{\eta}{\theta}$$

$$\implies a^*(\theta) = \arg\min_a E\left[m^* V_a(X) + \frac{(\theta(\alpha-1)V_a(X) + c(\theta,\alpha,a))^{\frac{1}{\alpha-1}+1}}{\theta\alpha(1-\alpha)}\right]$$
$$\implies a^*(\theta) = \arg\min_a E\left[m^* V_a(X) + \frac{m^*(\theta(\alpha-1)V_a(X) + c(\theta,\alpha,a))}{\theta\alpha(1-\alpha)}\right]$$
$$\implies a^*(\theta) = \arg\min_a E\left[m^* (V_a(X) + \frac{(\theta(\alpha-1)V_a(X) + c(\theta,\alpha,a))}{\theta\alpha(1-\alpha)})\right]$$
$$\implies a^*(\theta) = \arg\min_a E\left[m^* (V_a(X) - \frac{V_a(X)}{\alpha})\right] + E\left[m^* \frac{c(\theta,\alpha,a)}{\theta\alpha(1-\alpha)}\right]$$

The second term has all constants, which are independent of X. Hence the constant come out and $E[m^*] = 1$

$$\implies a^*(\theta) = \arg\min_a \left(\frac{\alpha-1}{\alpha} E\left[m^* V_a(X)\right] + \frac{c(\theta,\alpha,a)}{\theta\alpha(1-\alpha)}\right)$$

Substituting the value of m^* will result in the above expression. $\square$

Relative Entropy Case:

- In contrast of relative entropy, it is not clear whether the objective function is convex in a.

Measuring Model Error through α-Divergence:

- Using α-divergence to measure potential model error will concentrate the uncertainty on the tail decay of the nominal probability density. For example, in the simple scalar case where $V_a(x) = x^k$, choosing $\alpha > 1$ leads to the worst-case density function:

$$\tilde{f}_X(x) \approx c x^{k/(\alpha-1)} f_X(x)$$

  for $x \gg 0$, where f_X is the density function of X under the nominal measure. Incorporating model uncertainty makes the tail heavier asymptotically by a factor of $x^{k/(\alpha-1)}$.

Generalized Measures or Pre-metrics:

- If we can find a measure or pre-metric that makes it easy to derive the feasible likelihood ratio under the worst-case scenario, then most analyses can proceed without much difficulty.
- A possible form of measure is $E[\phi(m)]$, where $\phi(m) \geq 0$ holds for any likelihood ratio m. Relative entropy and α-divergence are special cases of these measures.

Summary

This section explains how α-divergence can be used to extend robustness analysis to heavy-tailed distributions, offering a more flexible approach than relative entropy. The mathematical formulation and practical example demonstrate the application of this method in risk measurement for financial portfolios.

3 Robust Monte carlo

The following 3 sections will -

- Present methods for estimating the model error bounds.
- Present ways of examining the worst-case model perturbation.
- Show how to constrain the possible sources of model error.

3.1 Estimating the bounds on model error

We use Monte Carlo estimation techniques to evaluate the worst case model error(i.e, its bounds). For eg-standard Monte Carlo estimator of $E[V(X)] = \frac{1}{N}\sum_{i=1}^{N} V(X_i)$. where $X_1, ..., X_N$ are sampled stochastic elements X. This section contains the Monte Carlo Approximation of model error bound calculation using the non robust version of risk measurement and the Relative Entropy bounded budget. For any fixed θ and likelihood ratio $m_\theta \propto \exp(\theta V(X))$, then $E(m_\theta V(X)) =$

$$\begin{aligned}
\Longrightarrow &= E\left[\frac{\exp(\theta V(X))}{E[\exp(\theta V(X))]} V(X)\right] \\
\Longrightarrow &= \frac{E[V(X)\exp(\theta V(X))]}{E[\exp(\theta V(X))]} \\
\Longrightarrow &= \frac{\sum_{i=1}^{N} V(X_i)\exp(\theta V(X_i))}{\sum_{i=1}^{N}\exp(\theta V(X_i))}
\end{aligned} \tag{13}$$

The following conclusion can be derived-

- $E[V(X)] \leq E[m_\theta V(X)]$ if $\theta > 0$. This is a conclusion of the Jensen's Inequality. $f(Ey) \leq E(fY)$, with $f(y) = e^{\theta y} y$ .It can also be observed analytically that $\exp(\theta V(X_i))$ puts more density to regions of high$V(x)$, hence increasing the Expectation.
- $E[V(X)] \geq E[m_\theta V(X)]$ if $\theta < 0$. This is a conclusion of the Jensen's Inequality for concave function.It can also be observed analytically that $\exp(\theta V(X_i))$ puts less density to regions of high$V(x)$ as $\theta < 0$.

From the same replications $X_1, ..., X_N$, we can estimate the likelihood ratio by setting the denominator (1) to its monte-carlo estimate using all the samples. Hence now, we can create a sample of $\hat{m}_{\theta,i}$. This in turn allows us to estimate the relative entropy at θ as

$$\hat{\eta}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \hat{m}_{\theta,i} \log \hat{m}_{\theta,i}$$

Getting samples of $\hat{m}_{\theta,i}$ is important because it helps in estimating expectations $E[m_\theta h(X)]$ of auxiliary functions h(X)

In some case, we may want to sample from the worst-case law, and not evaluate expectations under the change of measure. If V is bounded, we can achieve this through acceptance–rejection-

- We generate candidates X from the original nominal law.
- Accept them with probability $exp(\theta V(X))/M$, with M chosen so that this ratio is between 0 and 1.

These techniques extend to problems of optimization over a decision parameter a, introduced in Section 2.2.

- First we get the montecarlo equivalent of the optimal $a^*(\theta)$- (7)

$$\arg\inf_a \frac{1}{\theta}\log\left(\frac{1}{N}\sum_{i=1}^{N}\exp(\theta V_a(X_i))\right)$$

.

- Then substituting the value of $\hat{a}^*$ which minimizes this, in the parameterized function $V_a(X)$ of (3) to get the **worst-case objective function**-

$$\frac{\sum_{i=1}^{N} V_{\hat{a}^*}(X_i)\exp(\theta V_{\hat{a}^*}(X_i))}{\sum_{i=1}^{N}\exp(\theta V_{\hat{a}^*}(X_i))}$$

- Continuous mapping theorem states that stochastic convergence is preserved by continuous functions, i.e, functions of consistent estimators are consistent . For given $\hat{a}_N^*$ and any $\theta \in [0, \theta_{max}]$, the averages of both numerator and denominator of this expression are consistent estimators. Hence, (12) is a consistent estimator for (8) with $\hat{a}^*$.

Similar estimators are available in the α-divergence framework. For given $\theta > 0, \alpha > 1, a$, we estimate the worst-case likelihood ratio as-

$$m^*_{\theta,\alpha,a,i} = (\theta(\alpha-1)V_a(X_i) + \hat{c}(\theta,\alpha,a))^{\frac{1}{\alpha-1}} \tag{14}$$

with the constraints- (11) being satisfied for all X_i and (12) being satisfied in the average version over all i.
For given $\theta > 0, \alpha > 1$, we solve for an optimal a as-

$$\hat{a}^*(\theta) = \arg\min_a \left(\frac{\alpha-1}{\alpha} E\left[(\theta(\alpha-1)V_a(X) + \hat{c}(\theta,\alpha,a))^{\frac{1}{\alpha-1}} V_a(X) \right] + \frac{\hat{c}(\theta,\alpha,a)}{\theta\alpha(1-\alpha)} \right)$$

The robust estimator for the objective becomes-

$$\frac{1}{N}\sum_{i=1}^{N} V_a(X) m^*_{\theta,\alpha,\hat{a}^*(\theta),i}$$

3.2. Incorporating expectation constraints

This section talks about using additional information available about the 'true'model. Those additional information take the form of the following constraints-

$$\mathbb{E}[mh_i(X)] \leq \eta_i \quad \text{or} \quad \mathbb{E}[mh_i(X)] = \eta_i \quad \text{for some function } h_i \text{ and scalars } \eta_i.$$

The Lagrangian dual becomes-

$$\inf_{\theta>0,\lambda_i>0} \sup_m E\left[mV(X) - \frac{1}{\theta}(m\log m - \eta) + \sum_{i=1}^{n_M} \lambda_i (mh_i(X) - \eta_i) \right]$$

Proposition 3.1: For fixed $\theta > 0$ and $\lambda_i > 0,\ i = 1,\ldots,n_M$, such that

$$\mathbb{E}\left[\exp\left(\theta\left(V(X) - \sum_{i=1}^{n_M} \lambda_i h_i(X) \right) \right) \right] < \infty.$$

The worst change of measure is-

$$m^*_\theta \propto \exp\left(\theta\left(V(X) - \sum_{i=1}^{n_M} \lambda_i h_i(X) \right) \right) \tag{15}$$

The optimization over (θ, λ_i) becomes

$$\inf_{\theta>0,\lambda_i>0} \frac{1}{\theta} \log \mathbb{E}\left[\exp\left(\theta\left(V(X) - \sum_{i=1}^{n} \lambda_i h_i(X) \right) \right) \right] + \frac{\eta}{\theta} + \sum_{i=1}^{n_M} \eta_i \lambda_i \tag{16}$$

Proof. We take the Lagrangian of the objective equation in the dual of (1) with respect to the constraint done $E[m] = 1$ as done in section 2.

$$\implies E\left[mV(X) - \frac{1}{\theta}\left(m\log m - \eta\right) + \sum_{i=1}^{n_M}\lambda_i(mh_i(X) - \eta_i)\right] - \kappa(E(m) - 1)$$

$$\implies E\left[m\left(V(X) - \sum_{i=1}^{n_M}\lambda_i h_i(X)\right) - \frac{1}{\theta}\left(m\log m\right) - \kappa m\right] + \frac{\eta}{\theta} + \sum_{i=1}^{n_M}\lambda_i\eta_i + \kappa$$

$\implies$ Doing a first order derivative w.r.t m, we get

$$\implies E\left[V(X) - \sum_{i=1}^{n_M}\lambda_i h_i(X) - \frac{1}{\theta}(1 + \log m) - \kappa\right]$$

$\implies$ Similarly, we set the expression inside the expectation to be zero

$$\implies \theta\left(V(X) - \sum_{i=1}^{n}\lambda_i h_i(X)\right) - 1 - \log m - \theta\kappa = 0$$

$$\implies m^* = \exp\left(\theta\left(V(X) - \sum_{i=1}^{n_M}\lambda_i h_i(X)\right)\right)\exp\left(-1 - \theta\kappa\right)$$

To find the value of κ, we take the expectation of m^* and set it to 1 , we can see that the second exponent is anyways a constant. Hence m^* is proportional to the expression given in (15)
The optimal objective can be found by the same derivation as done in (5). The expression to be solved has the change $V(x) \rightarrow V(x) - \sum_{i=1}^{n_M}\lambda_i h_i(X)$ everywhere after substituting m^*, apart from the extra terms $\sum_{i=1}^{n_M}\eta_i\lambda_i$. Hence we make the same changes in (5) to get the solution (16). □

For an optimization problem as in (2), adding constraints entails solving another layer of optimization. Hence we will have an additional $\inf_a$ in the objective. So given θ, we solve for $m^*_(a(\theta))$ and then for $a(\theta)$ and get the functional forms in reverse order.

$$a^*(\theta, \lambda_i) = arg\inf_a \frac{1}{\theta}\log\mathbb{E}\left[\exp\left(\theta\left(V_a(X) - \sum_{i=1}^{n}\lambda_i h_i(a, X)\right)\right)\right] + \sum_{i=1}^{n_M}\eta_i\lambda_i$$

3.3 Restricting Sources of Model Uncertainty

In some cases, we want to go beyond imposing constraints on expectations to leave entire distributions unchanged by concerns about model error. We can use this device to focus robustness on parts of the model of particular concern.
Suppose, then, that the stochastic input has a representation as (X, Y), for a pair of random variables or vectors X and Y. We want to introduce robustness to model error in the law of X, but we have no uncertainty about the law of Y. For a given $\theta > 0$, we require that $\mathbb{E}[\exp(\theta V_a(X, Y)) \mid Y = y] < \infty$ for any y, and formulate the penalty problem

$$\inf_a\sup_m\mathbb{E}\left[m(X,Y)V_a(X,Y) - \frac{1}{\theta}\left(m(X,Y)\log m(X,Y) - \eta\right)\right]$$

subject to

$$\mathbb{E}[m(X,Y) \mid Y=y] = 1, \quad \forall y \tag{17}$$

$$m(x,y) \geq 0, \quad \forall x, y.$$

We have written $m(X,Y)$ to emphasize that the likelihood ratio may be a function of both inputs even if we want to leave the law of Y unchanged.

Proof of Proposition 3.2

(1) The condition $\mathbb{E}[m(X,Y) \mid Y=y] = 1$ ensures that the marginal distribution of Y remains unchanged. Consider the joint distribution $P_{X,Y}$ of random variables (X,Y). The new distribution $Q_{X,Y}$ is defined as:

$$Q_{X,Y}(x,y) = m(x,y)P_{X,Y}(x,y).$$

For the new distribution $Q_{X,Y}$, the marginal distribution Q_Y of Y is:

$$Q_Y(y) = \int Q_{X,Y}(x,y)\,dx = \int m(x,y)P_{X,Y}(x,y)\,dx.$$

According to the condition $E[m(X,Y)|Y=y] = 1, \forall y$, we have:

$$\int m(x,y)P_{X|Y=y}(x|y)\,dx = 1.$$

Multiplying by $P_Y(y)$, we obtain:

$$\int m(x,y)P_{X,Y}(x,y)\,dx = P_Y(y).$$

This shows that $Q_Y(y) = P_Y(y)$.

Convexity of Likelihood Ratios: For the likelihood ratio m, defined at each point x and y, the likelihood ratio function $m(X,Y)$ is non-negative and satisfies:

$$\mathbb{E}[m(X,Y) \mid Y=y] = 1, \quad \forall y.$$

Now, if m_1 and m_2 are likelihood ratios satisfying the above condition, then for any $t \in [0,1]$, the mixed likelihood ratio $tm_1 + (1-t)m_2$ also satisfies these conditions:

Proof. Non-negativity: Since m_1 and m_2 are non-negative, and t and $1-t$ are non-negative, $tm_1 + (1-t)m_2$ is also non-negative.

Condition: $\mathbb{E}[tm_1(X,Y) + (1-t)m_2(X,Y) \mid Y=y]$
$= t\mathbb{E}[m_1(X,Y) \mid Y=y] + (1-t)\mathbb{E}[m_2(X,Y) \mid Y=y] = t \cdot 1 + (1-t) \cdot 1 = 1.$

Therefore, the set of likelihood ratios m is convex. □

Concavity of Objective Function: The objective function is concave in m. **{Proof?}**

(2) Optimality of $m^*(x,y) \;\; = \frac{\exp(\theta V_a(X,Y))}{\mathbb{E}[\exp(\theta V_a(X,Y))|Y]}$ Consider the following equation-

$$\bar{K}(t) = \mathbb{E}\left[(tm^* + (1-t)m)\, V_a(X,Y) - \frac{1}{\theta}\left((tm^* + (1-t)m)\log\left(tm^* + (1-t)m\right) - \eta\right)\right],$$

where m is an arbitrary likelihood ratio satisfying (17). Obviously, m^* satisfies (17). Taking the derivative of $\bar{K}$ at zero and substituting for m^*, we get

$$\begin{aligned}\bar{K}'(0) &= \mathbb{E}\left[\left(V_a(X,Y) - \frac{1}{\theta}\log m^* - \frac{1}{\theta}\right)(m^* - m)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\left(V_a(X,Y) - \frac{1}{\theta}\log m^* - \frac{1}{\theta}\right)(m^* - m) \mid Y\right]\right]\end{aligned}$$

Substituting m^* as defined above (Similar to the derivation done in Assumption A.1 (5)), we get -

$$= \mathbb{E}\left[\frac{1}{\theta}\left(\log \mathbb{E}[\exp(\theta V_a(X,Y)) \mid Y] - 1\right)\mathbb{E}[(m^* - m) \mid Y]\right]. \qquad (18)$$

By constraint (17), for any $Y = y$, the conditional expectation $\mathbb{E}[(m^* - m) \mid Y]$ in (18) equals zero, so $\bar{K}'(0) = 0$. Hence m^* is an optimal solution satisfying constraint (17).
To implement M^*, we need to generate multiple copies $X_1, \ldots, X_N$ for each outcome of y and then form the Monte Carlo counterpart of M^*:

$$\hat{m}^*(x,y) = \frac{\exp(\theta V_a(x,y))}{\sum_{i=1}^N \exp(\theta V_a(X_i,y))/N}.$$

Robust Monte Carlo Recap:

This section give a good summary of the theory learned in section 3.1, 3.2 and 3.3. It gives the conclusion of the mathematical derivation in these sections.

4. Portfolio variance

From here on, the author switches to application of the above mention theories on Robust Model risk on different Risk measurements - $V(x)$. In this section he is taking $V(x)$ as the portfolio variance.

4.1. Mean-variance optimal portfolio

The optimization problem takes the shape of -

$$\sup_a E\left[a^T X - \frac{\gamma}{2} a(X - E[X])(X - E[X])a^T\right] \qquad (19)$$

This is mean-variance objective where we are trying to maximize the mean return constrained to the absolute variance. Hence the higher the variance , the more penalty we are incurring, Here $\gamma > 0$ be a risk-aversion parameter.
To solve this problem, we use the method in Section 3.2. We constrain the mean vector and limit uncertainty to the covariance matrix. Hence the added constraint is $E(mX) = \mu$ and here

$$V(X) = -\left(a^T X - \frac{\gamma}{2} a(X - E[X])(X - E[X])a^T\right)$$

. The -ve sign is for changing the supremum problem to an infremum problem.
Following the solution of (15), we get the worst-case likelihood ratio-

$$m_\theta^* \propto \exp\left(\theta\left(V(X) - \lambda^T(X-\mu)\right)\right)$$

Correspondingly, the expression(solution) for λ , given by (16) is-

$$\arg\min_{\lambda>0} \frac{1}{\theta} \log \mathbb{E}\left[\exp\left(\theta\left(V(X) - \lambda^T X\right)\right)\right] + \lambda^T \mu \tag{20}$$

The term with λ is linear in X and, therefore, affects only the mean of X. what this means is-

$$\exp\left(-\frac{1}{2}\left[(X-\mu)^\top \Sigma^{-1}(X-\mu)\right]\right) \times \exp(C^T X)$$

$$= \exp\left(-\frac{1}{2}\left[(X-\tilde{\mu})^\top \Sigma^{-1}(X-\tilde{\mu})\right]\right)$$

Where $\tilde{\mu}$ depends on C and whether its constant or not.
What I understood: We find λ by taking an arbitrary value of θ in (20). So λ is variable i,e depends on θ. So we get a variable mean $\tilde{\mu}(\theta)$. Hence by using the mean constraint,we can find $\tilde{\mu}(\theta) = \mu$. Finally we conclude-

$$m^* \propto \exp\left(\frac{\theta\gamma}{2} a^T (X-\mu)(X-\mu)^T a\right)$$

Matching the worst-case likelihood ratio with the desired form, we can rewrite:

$$\begin{aligned}
&\implies m^* \propto \exp\left(\theta\left(V_a(X) - \lambda^T(X-\mu)\right)\right) \\
&\implies = \exp\left(-\theta\left(a^T X - \frac{\gamma}{2} a^T (X-\mu)(X-\mu)^T a - \lambda^T (X-\mu)\right)\right) \\
&\implies = \exp\left(-\theta a^T X + \frac{\theta\gamma}{2} a^T (X-\mu)(X-\mu)^T a + \theta\lambda^T (X-\mu)\right) \\
&\implies \propto \exp\left(\frac{\theta\gamma}{2} a^T (X-\mu)(X-\mu)^T a\right) \cdot \exp\left(\theta(\lambda - a)^T (X-\mu)\right)
\end{aligned}$$

We find that λ = a.
What is the logic behind this? Are we saying that we don't need to find the optimal portfolio separately?
Now We can use the derivation of section 2.1 to get the variance of the alternate distribution. Here, the interesting fact is that the unchanging mean which we found in section 2.1 is pre-stated here.
Then, for given (a, θ) such that $\theta > 0$ and a s.t, $\Sigma^{-1} - \theta\gamma a a^\top > 0$, the worst-case change of measure has $X \sim N(\mu, \tilde{\Sigma})$, where $\tilde{\Sigma}^{-1} = \Sigma^{-1} - \theta\gamma a a^\top$.
The optimal portfolio a can be found by numerically solving:

$$\implies a^*(\theta) = \arg \inf_{a\in\mathcal{A}(\theta)} \frac{1}{\theta} \log \mathbb{E}\left[\exp\left(\theta\left[V(X) - \lambda^T X\right]\right)\right] + \lambda^T \mu$$

$$\implies \arg \inf_{a\in\mathcal{A}(\theta)} \frac{1}{\theta} \log \mathbb{E}\left[\exp\left(\frac{\theta\gamma}{2} a^T (X-\mu)(X-\mu)^T a\right)\right] + a^T \mu$$

Using change of Variance for multivariate normal distribution,

$$\implies \arg \inf_{a\in\mathcal{A}(\theta)} \frac{1}{\theta} \log\left(\det(\tilde{\Sigma})^{1/2} / \det(\Sigma)^{1/2}\right) + a^T \mu$$

$$\implies \arg \inf_{a\in\mathcal{A}(\theta)} \frac{1}{\theta}\log\left(\det((\Sigma^{-1}-\theta\gamma aa^T)\Sigma)^{-1/2}\right)+a^T\mu$$

$$\implies \arg \inf_{a\in\mathcal{A}(\theta)} \frac{1}{\theta}\log\det(I-\theta\gamma aa^T\Sigma)^{-1/2}+a^T\mu$$

$$\implies \arg \inf_{a\in\mathcal{A}(\theta)} \frac{1}{\theta}\log\frac{1}{\sqrt{\det(I-\theta\gamma aa^T\Sigma)}}+a^T\mu$$

The corresponding relative entropy is:

$$\eta(\theta)=\mathbb{E}\left[m^*\log m^*\right]$$

$$=\mathbb{E}\left[m^*\left(\frac{\theta\gamma}{2}a^T(X-\mu)(X-\mu)^Ta-\log\mathbb{E}\left[\exp\left(\frac{\theta\gamma}{2}a^T(X-\mu)(X-\mu)^Ta\right)\right]\right)\right]$$

$$=\frac{\theta\gamma}{2}\mathbb{E}\left[m^*(X-\mu)^Taa^T(X-\mu)\right]-\log\mathbb{E}\left[\exp\left(\frac{\theta\gamma}{2}a^T(X-\mu)(X-\mu)^Ta\right)\right]$$

Using the change of measure- $\mathbb{E}\left[m^*(X-\mu)^Taa^T(X-\mu)\right]=\tilde{\mathbb{E}}\left[(X-\mu)^Taa^T(X-\mu)\right]$

Next, we can use the standard expression , $\tilde{\mathbb{E}}\left[(X-\mu)^TA(X-\mu)\right]=\text{tr}(A\tilde{\Sigma})$ from the matrix cookbook

$$=\frac{\theta\gamma}{2}\text{tr}(aa^T\tilde{\Sigma})-\log\mathbb{E}\left[\exp\left(\frac{\theta\gamma}{2}a^T(X-\mu)(X-\mu)^Ta\right)\right]$$

Using change of Variance for multivariate normal distribution for the second term

$$=\frac{1}{2}\text{tr}(\theta\gamma aa^T\tilde{\Sigma})-\log\det(\tilde{\Sigma})^{1/2}/\det(\Sigma)^{1/2}$$

$= \text{Substituting } \theta\gamma aa^T=\Sigma^{-1}-\tilde{\Sigma}^{-1}$

$$=\frac{1}{2}\text{tr}\left((\Sigma^{-1}-\tilde{\Sigma}^{-1})\tilde{\Sigma}\right)-\log\det(\tilde{\Sigma}^{1/2}/\Sigma^{1/2})$$

$$=\frac{1}{2}\left(\text{tr}(\Sigma^{-1}\tilde{\Sigma}-I)+\log\det(\Sigma\tilde{\Sigma}^{-1})\right)$$

The author than states an example using synthetic data. He uses ten stocks with predetermined mean(μ) returns and variance covariance matrix(Σ). Description of optimal decision- Nominal portfolio (NP), Robust portfolio (RP)

- - Find NP=$\hat{a}$, by optimizing (19).
- - Find RP=$a^*(\theta)=\arg\inf_a \frac{1}{\theta}\log E[\exp(\theta V_a(X))]$ and $m^*_\theta \propto \exp\left(\theta V_{a^*(\theta)}(X)\right)$
- we compute the mean variance objective of both NP and RP, in both cases
- In nominal model , compute NP's as $E[V_{\hat{a}}(X)]$. Compute RP's $E[V_{a^*(\theta)}(X)]$
- In worst-case model, NP's $E[m^*_\theta V_{\hat{a}}(X)]$ and RP's $E[m^*_\theta V_{a^*(\theta)}(X)]$

At each value of θ, they plot the performance of the two portfolios (as measured by the mean-variance objective— recall that we are minimizing) against relative entropy. Observations-

- The performance of the RP portfolio under the nominal model is always inferior, as it must be since NP is optimal in the nominal model.
- However, under the worst-case model, the RP values are better than the NP values, as indicated by the upper portion of the figure 1.
- In the lower portion of the figure 1, we see the performance of the nominal portfolio under the best-case model perturbation (i.e we solve for $\inf_m$) possible at each level of relative entropy.

One more conclusion of this section is that, Worst-case change of measure results in significantly worse performance than any of these parameter perturbations. Effectively, model error as gauged by relative entropy does not necessarily correspond to a straight-forward error in parameters.

To show this , In the right panel they examine the performance of the nominal portfolio under specific parameter perturbations.

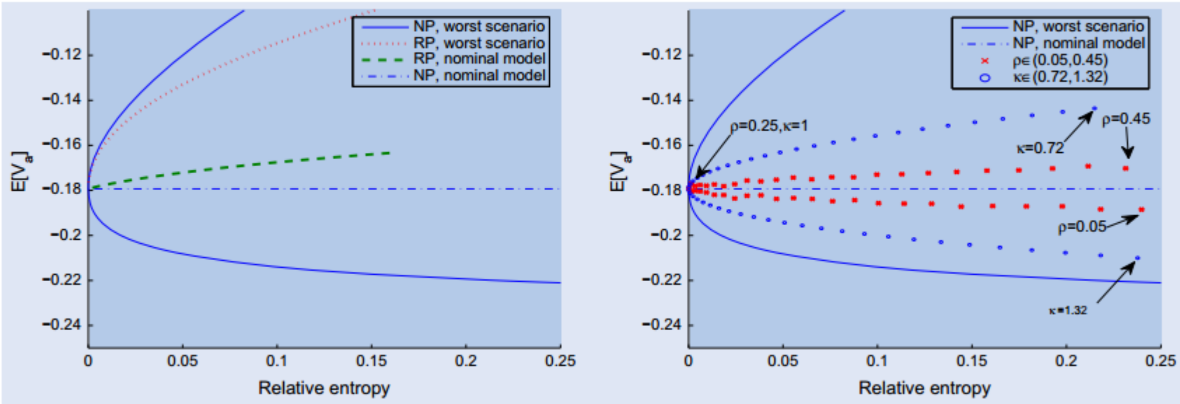

Figure 1: Expected performance vs. relative entropy.

4.2. Empirical example

To apply these ideas to data, we use daily returns from the CRSP database on the 126 stocks that were members of the S&P500 index. They calculate the the mean μ and covariance Σ of daily return using the first 12 years of data, through the end of 2001. They then construct the mean-variance optimal portfolio (the formula of which we have studied in FIM528)-

$$a = (\gamma\Sigma)^{-1}(\mu - \lambda I)$$

where

$$\lambda = \frac{I^\top(\gamma\Sigma)^{-1}\mu - 1}{I^\top(\gamma\Sigma)^{-1}I}$$

Forecasted variance = portfolio variance of this portfolio = 2.1×10^{-3} They then compare this forecast with the realized variance in 2002, when the dot-com bubble burst and 2008 when the financial crisis due to mortgage loans happened.

Realization Confidence intervals equal to two times the standard error of the realized variance and forecast have no overlap. The sampling variability in the initial period is not large enough to explain the realized variance.

Remedy Error intervals based on relative entropy. We use the portfolio variance as the objective and obtain the worst-case variance at different levels of θ. Let Model Error = |nominal variance-worst variance|. Now, we can form a new interval by combining both standard error and model error.

The forecast with both standard error and model error forms a pretty wide interval, which has a slight overlap with the confidence interval of the realized variance in 2008.

4.3. The heavy-tailed case

To demonstrate the heavy-tailed case, the vector of asset returns is given by $X \sim \mu + Z$, where $\mathbf{Z} \sim t_\nu(\mathbf{\Sigma}, \nu)$ has a multivariate t distribution with $\nu > 2$ degrees of freedom and covariance matrix $\nu\mathbf{\Sigma}/\nu - 2$.

In this example there is no implementation of optimal decision a, because they used a randomly generated portfolio weight vector. Table 2 shows the portfolio variance across various values of θ and α, with θ = 0 corresponding to the baseline nominal model. A smaller α in table 2 yields a heavier tail, but this does not necessarily imply a larger portfolio variance. We can think of choosing α based on an assessment of how heavy the tail might be and then varying θ to get a range of levels of uncertainty. If the application requires accommodating extreme values or outliers, a larger α might be chosen. Conversely, if the data is expected to be more concentrated around the mean, a smaller α might be appropriate.

- The log of the density $f_{|r|}(x)$ is:

$$\log f_{|r|}(x) = \log\left(2\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\right) - \frac{\nu+1}{2}\log\left(1+\frac{x^2}{\nu}\right)$$

- As $x \to \infty$, the dominant term in $\log\left(1+\frac{x^2}{\nu}\right)$ is $\log\left(\frac{x^2}{\nu}\right)$, which simplifies to:

$$\log\left(1+\frac{x^2}{\nu}\right) \approx \log\left(\frac{x^2}{\nu}\right) = \log x^2 - \log \nu = 2\log x - \log \nu$$

- Substituting this into the log density gives:

$$\log f_{|r|}(x) \approx \log\left(2\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\right) - \frac{\nu+1}{2}(2\log x - \log \nu)$$

- This simplifies to:

$$\log f_{|r|}(x) \approx \log\left(2\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\right) - (\nu+1)\log x + \frac{\nu+1}{2}\log \nu$$

- For large x, the term $-(\nu+1)\log x$ dominates, showing that $\log f_{|r|}(x)$ is asymptotically linear in $\log x$.

One thing to note is that, the above expression also gives a good idea of the degrees of freedom.
Taking the log of the solution of alpha-divergence likelihood ratio- as we find in (10),

$$\log m^*(\theta,\alpha) = \frac{1}{\alpha-1}\log\left(\theta(\alpha-1)r^2 + c(\theta,\alpha)\right) \approx \frac{2}{\alpha-1}\log|r|,$$

we separate the alternative distribution and the nominal distribution to get

$$\log \tilde{f}_{|r|}(x) - \log f_{|r|}(x) \approx \frac{2}{\alpha-1}\log x,$$

- The statement indicates that under the worst-case scenario, the tail of the distribution behaves similarly to a t-distribution but with modified degrees of freedom.
- Specifically, the degrees of freedom are adjusted to $\nu - \frac{2}{\alpha-1}$, where ν is the original degrees of freedom of the t-distribution. This adjustment reflects the increased heaviness of the tail under the worst-case scenario.

Definitions: k_{no} :slopes of nominal and $k_{\theta,\alpha}$: slopes of worst scenario. They lists the differences $k_{\theta,\alpha} - k_{no}$; as they increase α, the difference of slopes gets closer to the limit $2/(\alpha-1)$.
They estimate the degrees of freedom parameter by subtracting the slope differential $k_{\theta,\alpha} - k_{no}$. A second method to estimate this parameter is a maximum likelihood estimate using $m^*_{\theta,\alpha}$ to weight the nominal samples the worst covariance using monte-carlo estimate. The variance results under the parameters estimated at $\alpha = 2.5$ are very close to those estimated under the worst-case model at $\alpha = 2.5$.

5. Conditional value at risk

We use the traditional CVaR which is
CVaR also equals to the optimal value of the minimization problem-

$$\min_a \left(\frac{1}{1-\beta} E[(X-a)^+] + a \right)$$

for which the optimal a is VaR_β.

Proof. **Complete this** □

Hence we can set

$$V_a(X) = (1-\beta)^{-1}(X-a)^+ + a.$$

The main source of model error in measuring CVaR is the distribution of X. As in previous sections, we can introduce robustness to model uncertainty by considering a hypothetical adversary who changes the distribution of X.

5.1. Relative entropy uncertainty

Suppose X follows a double exponential distribution $DE(\mu, b)$ with location parameter μ and scale parameter b.
Its density function is

$$f(x) \propto \exp\left(-\frac{|x-\mu|}{b} \right).$$

Then, for given a and $\theta > 0$, the density function of X under the worst-case change of measure becomes

$$\tilde{f}(x) = m^*_{\theta,a}(x) f(x) \propto \exp\left(-\frac{|x-\mu|}{b} + \frac{\theta}{1-\beta}(x-a)^+ \right).$$

The coefficient of X in the exponent when X > a is $-1/b + \theta/(1-\beta)$, hence for the alternative density to be well defined , $1/b > \theta/(1-\beta)$ needs to be the condition.
$E[\exp(\theta V_a(X))]$ =
Case $a > \mu$

$$\implies \int \exp\left(\theta \frac{(x-a)^+}{1-\beta} + a\right) \frac{1}{2b} \exp - \frac{|x-\mu|}{b} dx$$

$$\implies \exp(\theta a) \left[\int_{-\inf}^{\mu} \frac{1}{2b} \exp(\frac{-(\mu-x)}{b}) dx + \int_{\mu}^{a} \frac{1}{2b} \exp(\frac{-(x-\mu)}{b}) dx + \int_{a}^{\inf} \frac{1}{2b} \exp(\frac{\theta(x-a)}{1-\beta} - \frac{(x-\mu)}{b}) dx \right]$$

$$\implies \exp(\theta a) \left[\frac{1}{2} + \frac{1}{2} - \frac{1}{2} \exp(\frac{\mu-a}{b}) + \frac{1}{2b} \left(\frac{-\exp(\frac{-(a-\mu)}{b})}{\theta/(1-\beta) - 1/b} \right) \right]$$

$$\implies \exp(\theta a) \left[1 + \frac{1}{2} \left(\frac{1}{1 - \frac{\theta b}{1-\beta}} - 1 \right) \exp(\frac{\mu-a}{b}) \right] \quad (21)$$

Similarly we can prove for the case $a < \mu$, where the derivation is a bit lengthy, but not complicated.

To find $a(\theta)$ we use eq 7. We observe that the function $E[\exp(\theta V_a(X))]$ is convex in a and its derivative at $a = \mu$ is positive hence its increasing at this point and the minima must lye at $a > \mu$. (This requires a little bit of arithmetic to be shown).
Solving for a^* by taking derivative of eqn (21)

$$a^*(\theta) = \mu - b \log \left(\frac{2(1 - \beta - \theta b)}{1 - \theta b} \right),$$

Remember that the parameter a was the VaR in the CVaR calculation. To find the nominal we can compute the value inside (21) and take the first order derivative.

cell1	Nominal	worst case
VaR	$\mu - b\log(2(1-\beta))$	$\mu - b\log\left(\frac{2(1-\beta-\theta b)}{1-\theta b}\right)$
CVaR	$\mu - b\log(2(1-\beta)) + b$	$a^*(\theta) + \frac{1}{\frac{1}{b}+\frac{\theta}{1-\beta}}$

Observe that the worst-case CVaR increases compared to the nominal CVaR. For a particular test case, we see that the worst-case model error (for CVaR) at θ = 0.03 shifts more mass to the right tail(hence increasing the uncertainty) and increases the VaR. The increase in VaR and the corresponding increase in CVaR reflect the magnitude of underestimation of risk consistent with this level of the uncertainty parameter θ.

5.2. The heavy-tailed case

If the nominal distribution of the loss random variable X is heavy-tailed, then $E[\exp(\theta V_a(X))]$ is infinite.
Recap of Heavy tail under α-divergence:- If the density function of X under the nominal distribution is regularly varying with index ρ, $(\rho < 0)$ then under the worst-case change of measure it is regularly varying with index $\rho + 1/(\alpha - 1)$.
This section takes the example of two heavy tail distribution, the generalised Pareto distribution and The generalised extreme value distribution and tries to find the form of the worst case alternative distribution. It then finds the increase in VaR and CVaR from the nominal case just like as done in the previous section. Derivation of the VaR and CVaR follows from the change in the shape parameter of both the distributions. A standard min–max argument produces the worst-case change of measure given by (14).

$$\implies \therefore Nominal : \rho = (1 + 1/\xi)$$
$$\implies \rho^* = -(1 + 1/\xi) + 1/(\alpha - 1) = -(1 + 1/\xi^*)$$
$$\implies \xi^* = \frac{\xi(\alpha - 1)}{\alpha - 1 - \xi}$$

1) Generalised Pareto distribution

$$\mathrm{VaR}_\beta = \frac{b}{\xi}\Big[(1-\beta)^{-\xi} - 1\Big], \qquad \mathrm{CVaR}_\beta = \frac{\mathrm{VaR}_\beta + b}{1 - \xi}$$

2) Generalised extreme value distribution

$$\mathrm{VaR}_\alpha = -\frac{1}{\xi_{\mathrm{GEV}}}\Big[(-\ln\alpha)^{-\xi_{\mathrm{GEV}}} - 1\Big], \qquad \mathrm{CVaR}_\alpha = \frac{\gamma\big(1-\xi_{\mathrm{GEV}}, -\ln\alpha\big) \;-\; (1-\alpha)}{\xi_{\mathrm{GEV}}\,(1-\alpha)}$$

where γ is the lower incomplete gamma function.
The results are discusses for the uncertainty level $\theta = 0.01$ and $\alpha = 4$.
We don't get a standard form of the alternative distribution just like the case of double exponential worst case density in the previous section. The graphical conclusions shown here are on the tails of the worst case scenarios.

Portfolio credit risk- 6.1 Gaussian Copula model

The Gaussian copula model for measuring credit capital is defined as follows. The below expression is for single-factor homogeneous model, where each obligator's loss profile follows -

$$X_i = \rho Z + \sqrt{1-\rho^2}\epsilon_i$$

i = 1,2, .., n. These are the risky entities in a loan portfolio. $Z, \epsilon_1, ..., \epsilon_n$ are independent standard normal random variables. Hence X_i is also standard normal random variables. (Just find the mean and variance of X_i)

$$L = \sum_{i=1}^{n} c_i Y_i$$

is the total loss encountered , given that c_i is the loss given default of entity i and Y_i is the binary indicator of Loss, i.e, $Y_i = I(X_i > x_i)$

6.2. Robustness and model error

We are interested in finding which perturbations of the model (in the sense of relative entropy) produce the greatest error in measuring tail loss probabilities $P(L > x) = E[I_{L>x}]$. Hence we are not interested in optimizing parameters in this case. We are going beyond parameter sensitivities to understand how the worst-case error changes the structure of the model. The sources of model uncertainty directly lies in X and indirectly originates from $Z, \epsilon_1, ..., \epsilon_n$.
The worst case change of measure- $m^*(\theta) \propto exp(\theta I_{L>x})$ This change of measure lifts the probabilities of losses greater than x and lowers the probability of all other scenarios

Experiment Setup

$n = 100$

$c_i \equiv 1$, i.e, We are taking the equal LGD amount, It can be scaled to a real world value.

$x = 5$

Hence,$P(L > x) = 3.8\%$Calculated using monte–carlo Estimate

Remember we can't use a binomial distribution for L, as the Y_i's are mutually dependent. Hence, we can use the equations given in section 3.1: paragraph 1 and (13), with N = 10^6 Samples. The dotted red line shows results under parameter changes only; these are determined as follows.

- At each relative entropy level, we estimate all model parameters (the means, standard deviations and correlations for the normal random variables $Z, \epsilon_1, ..., \epsilon_n$).
- $\mu_i = E[m^*(\theta)\epsilon_i]$ and $\sigma_{ij} = E[m^*(\theta)(\epsilon_i - \mu_i)(\epsilon_j - \mu_j)]$
- we then simulate the Gaussian copula model with these modified parameters in the model.

Conclusion

We see that the worst case change under the new model parameters curve lies below the worst case scenario below robust measurement. Thus, focusing on parameter changes only does not fully utilize the relative entropy budget. At this point, we can get slight hints at the worst case densities does not follow normal distribution strictly.
Its also been found that with 95% confidence, Jarque-Bera and Anderson-Darling test reject normality of Z at $\theta \geq 1$ but fail to reject normality of the ϵ_i even at $\theta = 2$.

Hence the idea of changing the model parameters while retaining the same copula structure of independent idiosyncratic factors does not count as the worst case model. In fact, the greatest vulnerability to model error takes us outside the Gaussian copula model, creating greater dependence between obligors in the direction of more likely defaults, rather than just through a change of parameters within the Gaussian copula framework.
To be continued