IEOR E4602: Quantitative Risk Management

Risk Measures

Martin Haugh

Department of Industrial Engineering and Operations Research
Columbia University
Email: martin.b.haugh@gmail.com

Reference: Chapter 8 of 2^{nd} ed. of MFE's *Quantitative Risk Management*.

Risk Measures

- Let $\mathcal{M}$ denote the space of random variables representing portfolio losses over some fixed time interval, Δ.
- Assume that $\mathcal{M}$ is a convex cone so that
  - If $L_1, L_2 \in \mathcal{M}$ then $L_1 + L_2 \in \mathcal{M}$
  - And $\lambda L_1 \in \mathcal{M}$ for every $\lambda > 0$.
- A risk measure is a real-valued function, $\varrho : \mathcal{M} \to \mathbb{R}$, that satisfies certain desirable properties.
- $\varrho(L)$ may be interpreted as the riskiness of a portfolio or ...
- ... the amount of capital that should be added to the portfolio so that it can be deemed **acceptable**
  - Under this interpretation, portfolios with $\varrho(L) < 0$ are already acceptable
  - In fact, if $\varrho(L) < 0$ then capital could even be *withdrawn*.

Axioms of Coherent Risk Measures

Translation Invariance For all $L \in \mathcal{M}$ and every constant $a \in \mathbb{R}$, we have

$$\varrho(L + a) \;=\; \varrho(L) + a.$$

- necessary if earlier risk-capital interpretation is to make sense.

Subadditivity: For all $L_1, L_2 \in \mathcal{M}$ we have

$$\varrho(L_1 + L_2) \;\leq\; \varrho(L_1) + \varrho(L_2)$$

- reflects the idea that pooling risks helps to diversify a portfolio
- the most debated of the risk axioms
- allows for the decentralization of risk management.

Axioms of Coherent Risk Measures

Positive Homogeneity For all $L \in \mathcal{M}$ and every $\lambda > 0$ we have

$$\varrho(\lambda L) \;=\; \lambda \varrho(L).$$

- also controversial: has been criticized for not penalizing concentration of risk
- e.g. if $\lambda > 0$ very large, then perhaps we should require $\varrho(\lambda L) > \lambda \varrho(L)$
- but this would be inconsistent with subadditivity:

$$\varrho(nL) \;=\; \varrho(L + \cdots + L) \;\leq n\varrho(L) \tag{1}$$

- positive homogeneity implies we must have equality in (1).

Monotonicity For $L_1, L_2 \in \mathcal{M}$ such that $L_1 \leq L_2$ almost surely, we have

$$\varrho(L_1) \;\leq\; \varrho(L_2)$$

- clear that any risk measure should satisfy this axiom.

Coherent Risk Measures

Definition: A risk measure, ϱ, acting on the convex cone $\mathcal{M}$ is called coherent if it satisfies the translation invariance, subadditivity, positive homogeneity and monotonicity axioms. Otherwise it is incoherent.

Coherent risk measures were introduced in 1998

- and a large literature has developed since then.

Convex Risk Measures

- Criticisms of subadditivity and positive homogeneity axioms led to the study of convex risk measures.
- A convex risk measure satisfies the same axioms as a coherent risk measure except that subadditivity and positive homogeneity axioms are replaced by the convexity axiom:

Convexity Axiom For $L_1, L_2 \in \mathcal{M}$ and $\lambda \in [0,1]$

$$\varrho(\lambda L_1 \;+\; (1-\lambda)L_2) \;\leq\; \lambda\varrho(L_1) \;+\; (1-\lambda)\varrho(L_2)$$

It is possible to find risk measures within the convex class that satisfy $\varrho(\lambda L) > \lambda\varrho(L)$ for $\lambda > 1$.

Value-at-Risk

Recall ...

Definition: Let $\alpha \in (0, 1)$ be some fixed confidence level. Then the VaR of the portfolio loss, L, at the confidence level, α, is given by

$$\mathsf{VaR}_\alpha := q_\alpha(L) \; = \; \inf\{x \in \mathbb{R} \; : \; F_L(x) \geq \alpha\}$$

where $F_L(\cdot)$ is the CDF of the random variable, L.

Value-at-Risk is **not** a coherent risk measure since it **fails** to be subadditive!

Example 1

Consider two IID assets, X and Y where

$$X = \epsilon + \eta \quad \text{where} \quad \epsilon \sim \mathsf{N}(0,1)$$

$$\text{and} \quad \eta = \begin{cases} 0, & \text{with prob } .991 \\ -10, & \text{with prob } .009. \end{cases}$$

Consider a portfolio consisting of X and Y. Then

$$\begin{aligned} \mathsf{VaR}_{.99}(X+Y) &= 9.8 \\ &> \mathsf{VaR}_{.99}(X) + \mathsf{VaR}_{.99}(Y) \\ &= 3.1 + 3.1 \\ &= 6.2 \end{aligned}$$

- thereby demonstrating the non-subadditivity of VaR.

Example 2: Defaultable Bonds

Consider a portfolio of $n = 100$ defaultable corporate bonds

- Probability of default over next year identical for all bonds and equal to 2%.
- Default events of different bonds are **independent**.
- Current price of each bond is 100.
- If bond does not default then will pay 105 one year from now
  - otherwise there is no repayment.

Therefore can define the loss on the i^{th} bond, L_i, as

$$L_i \ := \ 105\,Y_i \ - \ 5$$

where $Y_i = 1$ if the bond defaults over the next year and $Y_i = 0$ otherwise.

By assumption also see that $P(L_i = -5) = .98$ and $P(L_i = 100) = .02$.

Example 2: Defaultable Bonds

Consider now the following two portfolios:

- A: A fully concentrated portfolio consisting of 100 units of bond 1.
- B: A completely diversified portfolio consisting of 1 unit of each of the 100 bonds.

We want to compute the 95% VaR for each portfolio.

Obtain $\text{VaR}_{.95}(L_A) = -500$, representing a gain(!) and $\text{VaR}_{.95}(L_B) = 25$.

So according to $\text{VaR}_{.95}$, portfolio B is riskier than portfolio A

- absolute nonsense!

Have shown that

$$\text{VaR}_{.95}\left(\sum_{i=1}^{100} L_i\right) \;\geq\; 100\,\text{VaR}_{.95}(L_1) \;=\; \sum_{i=1}^{100} \text{VaR}_{.95}(L_i)$$

demonstrating again that VaR is not sub-additive.

Example 2: Defaultable Bonds

Now let ϱ be any coherent risk measure depending only on the distribution of L.

Then obtain (why?)

$$\varrho\left(\sum_{i=1}^{100} L_i\right) \;\leq\; \sum_{i=1}^{100} \varrho(L_i) \;=\; 100\varrho(L_1)$$

- so ϱ would correctly classify portfolio A as being riskier than portfolio B.

We now describe a situation where VaR is always sub-additive ...

Subadditivity of VaR for Elliptical Risk Factors

Theorem

Suppose that $\mathbf{X} \sim \mathsf{E}_n(\mu, \boldsymbol{\Sigma}, \psi)$ and let $\mathcal{M}$ be the set of linearized portfolio losses of the form

$$\mathcal{M} \; := \; \{L : L \; = \; \lambda_0 \; + \; \sum_{i=1}^{n} \lambda_i X_i, \; \lambda_i \in \mathbb{R}\}.$$

Then for any two losses $L_1, L_2 \in \mathcal{M}$, and $0.5 \leq \alpha < 1$,

$$\mathsf{VaR}_\alpha(L_1 + L_2) \; \leq \; \mathsf{VaR}_\alpha(L_1) \; + \; \mathsf{VaR}_\alpha(L_2).$$

Proof of Subadditivity of VaR for Elliptical Risk Factors

Without (why?) loss of generality assume that $\lambda_0 = 0$.

Recall if $\mathbf{X} \sim \mathsf{E}_n(\mu, \mathbf{\Sigma}, \psi)$ then $\mathbf{X} = \mathbf{AY} + \mu$ where $\mathbf{A} \in \mathbb{R}^{n \times k}$, $\mu \in \mathbb{R}^n$ and $\mathbf{Y} \sim S_k(\psi)$ is a spherical random vector.

Any element $L \in \mathcal{M}$ can therefore be represented as

$$\begin{aligned} L \;=\; \boldsymbol{\lambda}^T \mathbf{X} \;&=\; \boldsymbol{\lambda}^T \mathbf{AY} \;+\; \boldsymbol{\lambda}^T \mu \\ &\sim\; ||\boldsymbol{\lambda}^T \mathbf{A}||\; Y_1 \;+\; \boldsymbol{\lambda}^T \mu \end{aligned} \qquad (2)$$

- (2) follows from part 3 of Theorem 2 in *Multivariate Distributions* notes.

Translation invariance and positive homogeneity of VaR imply

$$\mathsf{VaR}_\alpha(L) \;=\; ||\boldsymbol{\lambda}^T \mathbf{A}||\, \mathsf{VaR}_\alpha(Y_1) \;+\; \boldsymbol{\lambda}^T \mu.$$

Suppose now that $L_1 := \boldsymbol{\lambda}_1^T \mathbf{X}$ and $L_2 := \boldsymbol{\lambda}_2^T \mathbf{X}$. Triangle inequality implies

$$||(\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2)^T \mathbf{A}|| \;\leq\; ||\boldsymbol{\lambda}_1^T \mathbf{A}|| \;+\; ||\boldsymbol{\lambda}_2^T \mathbf{A}||$$

Since $\mathsf{VaR}_\alpha(Y_1) \geq 0$ for $\alpha \geq .5$ (why?), result follows from (2). $\square$

Subadditivity of VaR

Widely believed that if individual loss distributions under consideration are continuous and symmetric then VaR is sub-additive.

This is not true(!)

- Counterexample may be found in Chapter 8 of *MFE*
- The loss distributions in the counterexample are smooth and symmetric but the copula is highly asymmetric.

VaR can also fail to be sub-additive when the individual loss distributions have heavy tails.

Expected Shortfall

Recall ...

Definition: For a portfolio loss, L, satisfying $\mathsf{E}[|L|] < \infty$ the expected shortfall (ES) at confidence level $\alpha \in (0,1)$ is given by

$$\mathsf{ES}_\alpha \;:=\; \frac{1}{1-\alpha} \int_\alpha^1 q_u(F_L)\; du.$$

Relationship between ES_α and VaR_α therefore given by

$$\mathsf{ES}_\alpha \;:=\; \frac{1}{1-\alpha} \int_\alpha^1 \mathsf{VaR}_u(L)\; du \tag{3}$$

- clear that $\mathsf{ES}_\alpha(L) \geq \mathsf{VaR}_\alpha(L)$.

When the CDF, F_L, is continuous then a more well known representation given by

$$\mathsf{ES}_\alpha \;=\; \mathsf{E}\left[L \mid L \geq \mathsf{VaR}_\alpha\right].$$

Expected Shortfall

Theorem: Expected shortfall is a coherent risk measure.

Proof: Translation invariance, positive homogeneity and monotonicity properties all follow from the representation of ES in (3) and the same properties for quantiles.

Therefore only need to demonstrate subadditivity

- this is proven in lecture notes. □

There are many other examples of risk measures that are coherent

- e.g. risk measures based on generalized scenarios
- e.g. spectral risk measures
  - of which expected shortfall is an example.

Risk Aggregation

Let $\mathbf{L} = (L_1, \ldots, L_n)$ denote a vector of random variables

- perhaps representing losses on different trading desks, portfolios or operating units within a firm.

Sometimes need to aggregate these losses into a random variable, $\psi(\mathbf{L})$, say.

Common examples include:

1. The total loss so that $\psi(\mathbf{L}) = \sum_{i=1}^{n} L_i$.
2. The maximum loss where $\psi(\mathbf{L}) = \max\{L_1, \ldots, L_n\}$.
3. The excess-of-loss treaty so that $\psi(\mathbf{L}) = \sum_{i=1}^{n} (L_i - k_i)^+$.
4. The stop-loss treaty in which case $\psi(\mathbf{L}) = \left(\sum_{i=1}^{n} L_i \ - \ k\right)^+$.

Risk Aggregation

Want to understand the risk of the aggregate loss function, $\varrho(\psi(\mathbf{L}))$

- but first need the distribution of $\psi(\mathbf{L})$.

Often know only the distributions of the L_i's

- so have little or no information about the dependency or copula of the L_i's.

In this case can try to compute lower and upper bounds on $\varrho(\psi(\mathbf{L}))$:

$$\begin{aligned}\varrho_{min} &:= \inf\{\varrho(\psi(\mathbf{L})) \ : \ L_i \sim F_i, \ i = 1, \ldots, n\} \\ \varrho_{max} &:= \sup\{\varrho(\psi(\mathbf{L})) \ : \ L_i \sim F_i, \ i = 1, \ldots, n\}\end{aligned}$$

where F_i is the CDF of the loss, L_i.

Problems of this type are referred to as Frechet problems

- solutions are available in some circumstances, **e.g.** attainable correlations.

Have been studied in some detail when $\psi(\mathbf{L}) = \sum_{i=1}^{n} L_i$ and $\varrho(\cdot)$ is the VaR function.

Capital Allocation

Total loss given by $L = \sum_{i=1}^{n} L_i$.

Suppose we have determined the risk, $\varrho(L)$, of this loss.

The capital allocation problem seeks a decomposition, $AC_1, \ldots, AC_n$, such that

$$\varrho(L) \;=\; \sum_{i=1}^{n} AC_i \tag{4}$$

- AC_i is interpreted as the risk capital allocated to the i^{th} loss, L_i.

This problem is important in the setting of performance evaluation where we want to compute a risk-adjusted return on capital (RAROC).

e.g. We might set $\text{RAROC}_i \;=\; \text{Expected Profit}_i \;/\; \text{Risk Capital}_i$

- must determine risk capital of each L_i in order to compute RAROC_i.

Capital Allocation

More formally, let $L(\boldsymbol{\lambda}) := \sum_{i=1}^{n} \lambda_i L_i$ be the loss associated with the portfolio consisting of λ_i units of the loss, L_i, for $i = 1, \ldots, n$.

Loss on actual portfolio under consideration then given by $L(\mathbf{1})$.

Let $\varrho(\cdot)$ be a risk measure on a space $\mathcal{M}$ that contains $L(\boldsymbol{\lambda})$ for all $\boldsymbol{\lambda} \in \Lambda$, an open set containing $\mathbf{1}$.

Then the associated risk measure function, $r_\varrho : \Lambda \to \mathbb{R}$, is defined by

$$r_\varrho(\boldsymbol{\lambda}) = \varrho(L(\boldsymbol{\lambda})).$$

We have the following definition ...

Capital Allocation Principles

Definition: Let r_ϱ be a risk measure function on some set $\Lambda \subset \mathbb{R}^n \setminus \mathbf{0}$ such that $\mathbf{1} \in \Lambda$.

Then a mapping, $f^{r_\varrho} : \Lambda \to \mathbb{R}^n$, is called a per-unit capital allocation principle associated with r_ϱ if, for all $\boldsymbol{\lambda} \in \Lambda$, we have

$$\sum_{i=1}^{n} \lambda_i f_i^{r_\varrho}(\boldsymbol{\lambda}) \; = \; r_\varrho(\boldsymbol{\lambda}). \tag{5}$$

- We then interpret $f_i^{r_\varrho}$ as the amount of capital allocated to one unit of L_i when the overall portfolio loss is $L(\boldsymbol{\lambda})$.
- The amount of capital allocated to a position of $\lambda_i L_i$ is therefore $\lambda_i f_i^{r_\varrho}$ and so by (5), the total risk capital is fully allocated.

The Euler Allocation Principle

Definition: If r_ϱ is a positive-homogeneous risk-measure function which is differentiable on the set Λ, then the per-unit Euler capital allocation principle associated with r_ϱ is the mapping

$$f^{r_\varrho} : \Lambda \to \mathbb{R}^n \ : \ f_i^{r_\varrho}(\boldsymbol{\lambda}) \ = \ \frac{\partial r_\varrho}{\partial \lambda_i}(\boldsymbol{\lambda}).$$

- The Euler allocation principle is a full allocation principle since a well-known property of any positive homogeneous and differentiable function, $r(\cdot)$ is that it satisfies

$$r(\boldsymbol{\lambda}) = \sum_{i=1}^{n} \lambda_i \frac{\partial r}{\partial \lambda_i}(\boldsymbol{\lambda}).$$

- The Euler allocation principle therefore gives us different risk allocations for different positive homogeneous risk measures.
- There are good economic reasons for employing the Euler principle when computing capital allocations.

Value-at-Risk and Value-at-Risk Contributions

Let $r^{\alpha}_{VaR}(\boldsymbol{\lambda}) = \mathsf{VaR}_{\alpha}(L(\boldsymbol{\lambda}))$ be our risk measure function.

Then subject to technical conditions can be shown that

$$\begin{aligned} f_i^{r^{\alpha}_{VaR}}(\boldsymbol{\lambda}) &= \frac{\partial r^{\alpha}_{VaR}}{\partial \lambda_i}(\boldsymbol{\lambda}) \\ &= \mathsf{E}\left[L_i \mid L(\boldsymbol{\lambda}) = \mathsf{VaR}_{\alpha}(L(\boldsymbol{\lambda}))\right], \quad \text{for} \ \ i = 1, \ldots, n. \end{aligned} \tag{6}$$

Capital allocation, AC_i, for L_i is then obtained by setting $\boldsymbol{\lambda} = \mathbf{1}$ in (6).

Will now use (6) and Monte-Carlo to estimate the VaR contributions from each security in a portfolio.

- Monte-Carlo is a general approach that can be used for complex portfolios where (6) cannot be calculated analytically.

An Application: Estimating Value-at-Risk Contributions

Recall total portfolio loss is $L = \sum_{i=1}^{n} L_i$.

According to (6) with $\lambda = 1$ we know that

$$
\begin{aligned}
AC_i &= \mathsf{E}\left[L_i \mid L = \mathsf{VaR}_\alpha(L)\right] && (7) \\
&= \left.\frac{\partial\, \mathsf{VaR}_\alpha(\boldsymbol{\lambda})}{\partial \lambda_i}\right|_{\lambda=1} \\
&= w_i \frac{\partial\, \mathsf{VaR}_\alpha}{\partial w_i} && (8)
\end{aligned}
$$

for $i = 1, \ldots, n$ and where w_i is the number of units of the i^{th} security held in the portfolio.

Question: How might we use Monte-Carlo to estimate the VaR contribution, AC_i, of the i^{th} asset?

Solution: There are three approaches we might take:

First Approach: Monte-Carlo and Finite Differences

As AC_i is a (mathematical) derivative we could estimate it numerically using a finite-difference estimator.

Such an estimator based on (8) would take the form

$$\widehat{AC}_i := \frac{\mathrm{VaR}_\alpha^{i,+} - \mathrm{VaR}_\alpha^{i,-}}{2\delta_i} \tag{9}$$

where $\mathrm{VaR}_\alpha^{i,+}$ ($\mathrm{VaR}_\alpha^{i,-}$) is the portfolio VaR when number of units of the i^{th} security is increased (decreased) by $\delta_i w_i$ units.

Each term in numerator of (9) can be estimated via Monte-Carlo

- same set of random returns should be used to estimate each term.

What value of δ_i should we use? There is a bias-variance tradeoff but a value of $\delta_i = .1$ seems to work well.

This estimator will not satisfy the additivity property so that $\sum_i^n \widehat{AC}_i \neq \mathrm{VaR}_\alpha$

- but easy to re-scale estimated $\widehat{AC}_i$'s so that the property will be satisfied.

Second Approach: Naive Monte-Carlo

Another approach is to estimate (7) directly. Could do this by simulating N portfolio losses $L^{(1)}, \ldots, L^{(N)}$ with $L^{(j)} = \sum_{i=1}^{n} L_i^{(j)}$

- $L_i^{(j)}$ is the loss on the i^{th} security in the j^{th} simulation trial.

Could then set (why?) $AC_i = L_i^{(m)}$ where m denotes the VaR_α scenario, i.e. $L^{(m)}$ is the $\lceil N(1-\alpha) \rceil^{th}$ largest of the N simulated portfolio losses.

Question: Will this estimator satisfy the additivity property, i.e. will $\sum_i^n AC_i = \text{VaR}_\alpha$?

Question: What is the problem with this approach? Will this problem disappear if we let $N \to \infty$?

A Third Approach: Kernel Smoothing Monte-Carlo

An alternative approach that resolves the problem with the second approach is to take a weighted average of the losses in the i^{th} security around the VaR_α scenario.

A convenient way to do this is via a kernel function.

In particular, say $K(x;h) := K\left(\frac{x}{h}\right)$ is a kernel function if it is:

1. Symmetric about zero
2. Takes a maximum at $x = 0$
3. And is non-negative for all x.

A simple choice is to take the triangle kernel so that

$$K(x;h) := \max\left(1 - \left|\frac{x}{h}\right|,\, 0\right).$$

A Third Approach: Kernel Smoothing Monte-Carlo

The kernel estimate of AC_i is then given by

$$\widehat{AC}_i^{ker} := \frac{\sum_{j=1}^{N} K\left(L^{(j)} - \hat{\mathsf{VaR}}_\alpha; h\right) L_i^{(j)}}{\sum_{j=1}^{N} K\left(L^{(j)} - \hat{\mathsf{VaR}}_\alpha; h\right)} \tag{10}$$

where $\widehat{\mathsf{VaR}}_\alpha := L^{(m)}$ with m as defined above.

One minor problem with (10) is that the additivity property doesn't hold. Can easily correct this by instead setting

$$\widehat{AC}_i^{ker} := \widehat{\mathsf{VaR}}_\alpha \frac{\sum_{j=1}^{N} K\left(L^{(j)} - \hat{\mathsf{VaR}}_\alpha; h\right) L_i^{(j)}}{\sum_{j=1}^{N} K\left(L^{(j)} - \hat{\mathsf{VaR}}_\alpha; h\right) L^{(j)}}. \tag{11}$$

Must choose an appropriate value of smoothing parameter, h.

Can be shown that an optimal choice is to set

$$h = 2.575\,\sigma\,N^{-1/5}$$

where $\sigma = \mathsf{std}(L)$, a quantity that we can easily estimate.

When Losses Are Elliptically Distributed

If $L_1, \ldots, L_N$ have an elliptical distribution then it may be shown that

$$AC_i \;=\; \mathsf{E}\left[L_i\right] + \frac{\mathsf{Cov}\left(L, L_i\right)}{\mathsf{Var}\left(L\right)}\left(\mathsf{VaR}_\alpha(L) - \mathsf{E}\left[L\right]\right). \tag{12}$$

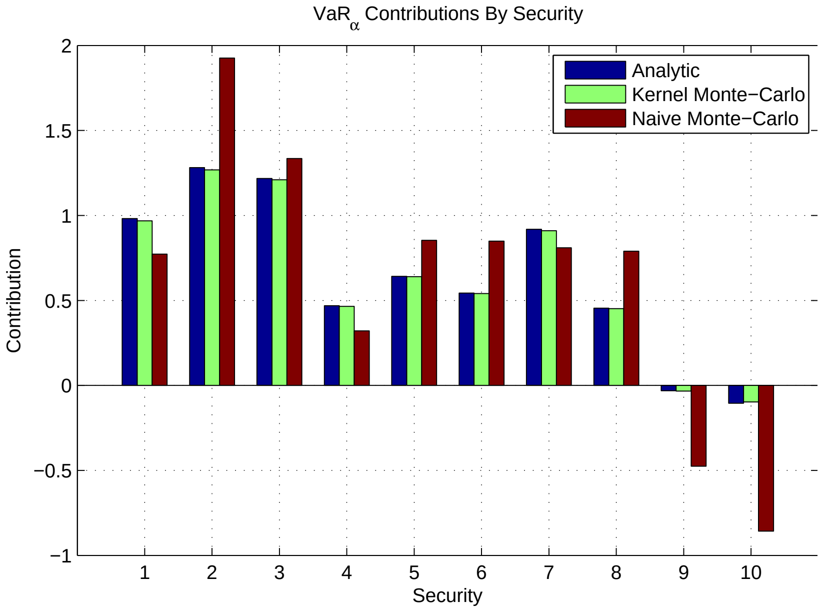
In numerical example below, we assume 10 security returns are elliptically distributed. In particular, losses satisfy $(L_1, \ldots, L_n) \sim \mathsf{MN}_n(\mathbf{0}, \mathbf{\Sigma})$.

Other details include:

1. First eight securities were all positively correlated with one another.
2. Second-to-last security uncorrelated with all other securities.
3. Last security had a correlation of -0.2 with the remaining securities.
4. Long position held on each security.

Estimated $VaR_{\alpha=.99}$ contributions of the securities displayed in figure below

- last two securities have a negative contribution to total portfolio VaR
- also note how inaccurate the "naive" Monte-Carlo estimator is
- but kernel Monte-Carlo is very accurate!


VaR$_\alpha$ Contributions By Security
Analytic
Kernel Monte−Carlo
Naive Monte−Carlo
Contribution
Security
2
1.5
1
0.5
0
−0.5
−1
1
2
3
4
5
6
7
8
9
10