

**Universidade Estadual de Campinas**

INSTITUTO DE MATEMÁTICA ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA

Departamento de Matemática Aplicada

---

Dissertação de Mestrado

**Medidas de risco em otimização de portfólios**

por

**Luís Felipe C. da R. Bueno<sup>†</sup>**

Mestrado em Matemática Aplicada - Campinas - SP

**Orientador: Prof. Dr. José Mario Martínez**

<sup>†</sup>Este trabalho contou com apoio financeiro da Capes.

# Medidas de risco em otimização de portfólios

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por **Luís Felipe C. da R. Bueno** e aprovada pela comissão julgadora.

Campinas, 25 de Fevereiro de 2008.



---

**Prof. Dr. José Mario Martínez**

Orientador

Banca examinadora:

Prof. Dr. José Mario Martínez.

Prof. Dr. Paulo José da Silva e Silva.

Profa. Dr. Roberto Andreani .

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de **MESTRE em MATEMÁTICA APLICADA**.

**FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DO IMECC DA UNICAMP**  
Bibliotecária: Maria Júlia Milani Rodrigues – CRB8a / 2116

Bueno, Luís Felipe Cesar da Rocha  
B862m      Medidas de risco em otimização de portfólios / Luís Felipe Cesar da Rocha Bueno -- Campinas, [S.P. :s.n.], 2008.

Orientador : José Mario Martínez  
Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Modelamento matemático. 2. Otimização matemática. 3. Order-value optimization (OVO). 4. Value at risk (VaR). 5. Conditional value at risk (CVaR). I. Martínez, José Mario. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título. (mjmr/imecc)

Título em inglês: Risk measures in portfolio optimization.

Palavras-chave em inglês (Keywords): 1. Mathematical modelling. 2. Mathematical optimization. 3. Order-value optimization (OVO). 4. Value at risk (VaR). 5. Conditional value at risk (CVaR)

Área de concentração: Otimização

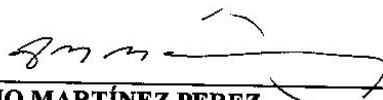
Titulação: Mestre em Matemática Aplicada

Banca examinadora: Prof. Dr. José Mario Martínez (IMECC-UNICAMP)  
Prof. Dr. Roberto Andreani (IMECC-UNICAMP)  
Prof. Dr. Paulo José da Silva e Silva (IME-USP)

Data da defesa: 25/02/2008

Programa de Pós-Graduação: Mestrado em Matemática Aplicada

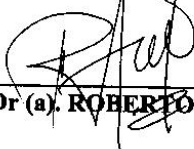
**Dissertação de Mestrado defendida em 25 de fevereiro de 2008 e aprovada  
Pela Banca Examinadora composta pelos Profs. Drs.**



\_\_\_\_\_  
**Prof. (a). Dr (a). JOSÉ MARIO MARTÍNEZ PEREZ**



\_\_\_\_\_  
**Prof. (a). Dr (a). PAULO JOSÉ DA SILVA E SILVA**



\_\_\_\_\_  
**Prof. (a). Dr (a). ROBERTO ANDREANI**

---

# AGRADECIMENTOS

Gostaria de poder agradecer a cada pequeno desvio no meu caminho que, de um modo ou de outro, me conduziu até aqui. Infelizmente isso não é possível, pois a quantidade de desvios pelos quais passamos é imensa, e as vezes esses desvios são tão pequenos que os tornam imperceptíveis. De todo modo, aceitarei mais um desafio ao escrever esse trabalho e tentarei agradecer a todos os principais grupos que contribuíram nessa caminhada. Dessa forma os que tiverem a paciência de ler até o final encontrarão-se com um texto diferente, e sobretudo maior, do que o habitual.

Primeiramente agradeço aos meus pais, Newton e Maria de Lourdes, por terem me proporcionado a oportunidade de estudar em boas escolas e pelo incentivo e amor que sempre me deram. O apoio que recebi durante toda a minha vida, e principalmente quando mudei para Campinas, foram fundamentais para o desenvolvimento deste trabalho.

Agradeço ao governo brasileiro, especialmente à CAPES, pelo auxílio financeiro fundamental para que pudesse me dedicar à pesquisa durante esse período.

Agradeço também aos meus irmãos, Martha e Paulo, pela confiança que sempre depositaram em mim, fazendo com que eu nunca me desanimasse procurando não desapontá-los.

À minha família por parte de pai por ser um porto seguro nos meus retornos a Minas

Gerais. Em especial aos meus primos Léo com quem, além de ter uma boa amizade, convivi por alguns meses; Dudu por não me deixar esquecer da alegria e amizade verdadeira de uma criança e Cacá por ser uma amiga pra todas as horas e com quem sempre matava as saudades, mesmo pelo telefone.

Aos meus amigos de república que me aguentaram por todo esse tempo, em especial Tiago e Brenner.

Aos meus amigos do CSA, com quem nunca perdi o contato e que me fazem lembrar não só das bagunças do colégio mas também dos ideais que nos foram passados. Não poderia deixar de destacar nesse grupo os amigos Rodrigo e Pedrinho que sempre foram confidentes e confessores e que, apesar da distância, estiveram presentes nas mais diversas situações nesses seis anos em Campinas.

Aos amigos e professores nos cursos no IMPA que muito contribuíram para minha formação. E aos companheiros de Florianópolis no CNMAC, em especial à dona Margarida, mãe do meu amigo César, por ter me acolhido tão bem em sua casa.

Aos meus amigos da graduação com os quais aprendi a me virar sozinho superando as dificuldades existenciais durante o curso e a distância da família. Gostaria de citar alguns nomes fundamentais na minha formação Ângela, Mariana, Márcia, Gláucia, Ana Cláudia, Ana Paula, Priscila, Letícia, Ricardo, Poli, Lipinho, Batata, Lonardo, Daniel Cariello, Maurício ressaltando Paula, Marília, Ana Paula, Léo, Squizo, Élton, André, Régis, Anne, e principalmente Ulisses.

Aos meus amigos da pós graduação, dentre eles os companheiros de sala Jair, Ederson e Flávia; à equipe de F. de tênis Cristiano, João e à torcedora Mael (ela torcia pra que chovesse) e aos colegas nas disciplinas Túlio, Liliana, Rodrigo, Larissa, Momoe, Giovane, Feodor, Leandro, Maik, Sílvia, Diego, João, Rafa, Alê, Thiago, entre outros.

Aos professores do IMECC, em especial à Valéria e Cheti por terem sido verdadeiras mães, à Sandra pela paciência de ter lido toda a dissertação, Lúcio, Aurélio, Moretti e

Clóvis pela ajuda com PL, Laura e Nancy pela ajuda com probabilidade. Ao grupo de otimização (professores e alunos) por quem fui muito bem recebido e com quem muito aprendi. Ao professor Paulo da USP por ter aceitado participar da banca e cujos comentários enriquecem este trabalho.

Ao Nino por ter depositado sua confiança em mim, convidando-me, exatamente no dia do meu aniversário, a fazer parte do melhor grupo de otimização do país.

Ao Martínez por ter aceitado me orientar, e tê-lo feito da melhor forma possível. Além disso foi um grande amigo, proporcionando vários momentos divertidos durante todo esse período.

Aos alunos dos quais fui monitor e às turmas de MS 416, em especial à aluna Andrea Mello, que me incentivaram a escrever este texto da forma mais simples possível.

Aos funcionários do IMECC, em especial Júlio, Alice, Ester, Tânia e Dona Fátima pelo apoio que me deram.

Às ATPs por terem me proporcionado momentos inesquecíveis. Na profissional destaco Gustavo Kuerten, exemplo de luta, dedicação, sucesso e humildade e a dupla André Sa e Marcelo Melo pelas alegrias nesses últimos anos. A do parque considero como uma família, onde aprendi muito sobre valores da vida. Por lá fiz grandes amigos e destaco os meus afilhados de casamento Cícero e Ana pra quem sempre ligo pra matar as saudades.

À Joana, Cristina, Nely, Val e Isabel por toda sua simplicidade e amizade cativante.

Às minhas cachorras, especialmente Mika e Preta neste período, pelo exemplo de amor e lealdade incondicional.

Por fim agradeço à minha família por parte de mãe. À Almerinda, por ter me recebido tão bem em São Paulo por diversas vezes e ao seu marido, afinal meu sonho quando crescer é ser Tio Galdino.

Agradeço também a todos que lêem esse trabalho e a todos que esqueci neste momento mas que por várias vezes foram lembrados durante essa trajetória.



# Resumo

Nesta dissertação fazemos uma exposição sobre alguns modelos matemáticos com aplicações em economia. Dentre os modelos estudados destacamos a versão discreta das populares medidas de risco VaR (*Value at Risk*) e C-VaR (*Conditional Value at Risk*). Discutimos algumas propriedades de tais medidas, e, principalmente, expomos sobre algumas idéias para otimizá-las sob uma formulação do tipo OVO (*Order Value Optimization*) e propomos uma nova formulação para o problema de minimizar a VaR.

**Palavras-chave:** Modelagem Matemática, Otimização, OVO (*Order Value Optimization*), VaR (*Value at Risk*), C-VaR (*Conditional Value at Risk*).

# Abstract

In this dissertation we make a presentation on some mathematical models with applications in economics. Among the studied models we highlight a discrete version of the popular risk measures VaR (Value at Risk) and C-VaR (Conditional Value at Risk). We discuss about some properties of such measures, and, above all, expose on some ideas for optimizing the VaR and CVaR under a OVO (Order Value Optimization) formulation and propose a new formulation to the problem of minimizing the VaR.

**Keywords:** Mathematical Modeling, Optimization, OVO (Order Value Optimization), VaR (Value at Risk), C-VaR (Conditional Value at Risk).

---

# ÍNDICE

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Perda média <math>\times</math> risco</b>	<b>3</b>
2.1	Introdução . . . . .	3
2.2	Minimizar a perda média . . . . .	4
2.3	Risco como distância até a média . . . . .	8
2.3.1	Modelo de Markowitz . . . . .	9
2.3.2	Modelo de Konno . . . . .	20
<b>3</b>	<b>VaR e C-VaR como medidas de risco</b>	<b>23</b>
3.1	A VaR ( <i>Value at Risk</i> ) . . . . .	23
3.2	A C-VaR ( <i>Conditional Value at Risk</i> ) . . . . .	26
3.3	Coerência . . . . .	28
3.4	Outras propriedades da VaR e da C-VaR . . . . .	34
3.5	Otimização da VaR e da C-VaR . . . . .	37
<b>4</b>	<b>Noções de probabilidade e simulação de cenários futuros</b>	<b>52</b>
4.1	Introdução . . . . .	52
4.2	Simulações probalísticas . . . . .	53
<b>5</b>	<b>Introdução à fórmula Black Scholes</b>	<b>63</b>

<b>6 Testes computacionais e conclusões</b>	<b>70</b>
6.1 Conclusões finais e perspectivas futuras . . . . .	80
<b>Bibliografia</b>	<b>81</b>

---

# CAPÍTULO 1

---

## Introdução

Nesta dissertação fazemos uma exposição sobre alguns modelos matemáticos com aplicações em economia. Nosso objetivo no texto é apresentar algumas maneiras de como tomar uma decisão financeira de modo a obtermos um lucro razoável sem nos expor a um risco maior do que o aceitado. As ferramentas que usaremos para desenvolver nossos raciocínios serão os conceitos básicos de probabilidade, programação linear e não linear, e uma idéia intuitiva de economia.

No capítulo 2 apresentamos os conceitos básicos a serem usados por todo o texto tais como o vetor de decisão  $x$ , indicando a quantidade de ativos que devemos adquirir, vetores de preços futuros  $y^j$  e as funções de perda a eles associados  $f(x, y^j)$ . Feitas essas considerações, introduziremos o conceito de risco e trataremos dos modelos de Markowitz e Konno.

O capítulo 3 é a parte fundamental deste trabalho. Nele apresentamos a versão discreta das populares medidas de risco VaR (*Value at Risk*) e C-VaR (*Conditional Value at Risk*). Discutimos a idéia de coerência exposta em [10] e propomos um novo conceito de coerência. Ainda nesse capítulo destacamos a seção 3.5. Nela discutimos algumas idéias para otimizar a VaR e a CVaR sob uma formulação do tipo OVO (*Order Value Optimization*), apresentadas em [2, 6, 14], e propomos uma nova formulação para o problema de

minimizar a VaR.

No capítulo 4 relembramos os principais resultados da probabilidade elementar, comentamos sobre o *movimento browniano geométrico* e suas aplicações em modelos probabilísticos ligados a economia. Por fim, apresentamos a maneira que usamos para simular nossos cenários futuros.

No capítulo 5 introduzimos os conceitos básicos do mercado de opções e daremos uma idéia sobre a precificação de uma opção do tipo Call Européia através da fórmula original desenvolvida por Black e Scholes [12].

Por fim, no capítulo 6 apresentamos os resultados computacionais referentes ao capítulo 3 e expomos nossas conclusões.

---

# CAPÍTULO 2

---

## Perda média $\times$ risco

---

### 2.1 Introdução

---

Suponhamos que existissem  $n$  ativos e que devêssemos tomar a decisão de comprar a quantidade  $x_i$  do ativo  $i$ . O sucesso de nossa decisão depende do futuro da economia, porém o futuro nunca pode ser previsto com exatidão. Nos nossos modelos existirão  $m$  vetores  $y \in \mathbb{R}^n$  que representarão  $m$  cenários possíveis para a situação da economia. No capítulo 5 discutiremos algumas idéias de como obter esses vetores  $y$ . Nesse capítulo assumiremos que eles já são dados de alguma forma, por exemplo por um economista ou uma simulação probabilística. Desse modo, para cada decisão  $x$  tomada teremos uma função de lucro  $l(x, y)$ , a qual é uma variável aleatória. Ou seja,  $l(x, y^j)$  fornece o lucro que teremos se tomarmos a decisão  $x$  e acontecer o cenário  $j$ . Por motivo de tradição da literatura de otimização, a qual trata do problema de minimização, definiremos a função  $f(x, y) = -l(x, y)$  chamada função de perda da decisão  $x$  sob o cenário  $y$ . Por simplicidade escreveremos algumas vezes  $f_j(x) = f(x, y^j)$ . Definamos também  $\Omega$  como sendo o conjunto das decisões possíveis de serem tomadas. Por exemplo, aquelas cujos gastos não ultrapassam uma quantidade de dinheiro disponível. Visto isso, como tomar uma decisão?

---

## 2.2 Minimizar a perda média

---

Nossa primeira idéia seria tomar a decisão que dará a menor perda possível na média dos cenários.

$$\begin{aligned} & \text{minimizar} \quad E[f(x, y)] \\ & \text{sujeito a} \quad x \in \Omega \end{aligned}$$

Supondo os  $m$  cenários equiprováveis (portanto cada um com probabilidade  $\frac{1}{m}$ ),  $y_i^0$  representando o preço inicial do ativo  $i$  e  $y_i^j$  o preço do ativo  $i$  sob o cenário  $j$ , temos que:

$$f(x, y^j) = \text{gastos} - \text{retorno} = \sum_{i=1}^n y_i^0 x_i - \sum_{i=1}^n y_i^j x_i = \sum_{i=1}^n (y_i^0 - y_i^j) x_i$$

Logo, tomando a decisão  $x$ , a perda média seria:

$$\begin{aligned} E[f(x, y)] &= \sum_{j=1}^m \frac{1}{m} f(x, y^j) \\ &= \sum_{j=1}^m \frac{1}{m} \sum_{i=1}^n (y_i^0 x_i - y_i^j x_i) \\ &= \sum_{i=1}^n \sum_{j=1}^m \frac{1}{m} (y_i^0 x_i - y_i^j x_i) \\ &= \sum_{i=1}^n \left( y_i^0 x_i - x_i \frac{1}{m} \sum_{j=1}^m y_i^j \right) \end{aligned}$$

Chamando  $\bar{y}_i = \frac{1}{m} \sum_{j=1}^m y_i^j$  o preço futuro médio do ativo  $i$  temos que:

$$E(f(x, y)) = \sum_{i=1}^n (y_i^0 - \bar{y}_i) x_i$$

Podemos ver que  $(y_i^0 - \bar{y}_i)$  é a perda média por unidade do ativo  $i$ .

No caso em que  $\Omega$  é o conjunto das decisões cujos gastos são menores ou iguais a quantidade de dinheiro disponível ( $M$ ), temos que:

$$\Omega = \left\{ x \in \mathbb{R}^n; \sum_{i=1}^n y_i^0 x_i \leq M \text{ e } x_i \geq 0 \ \forall i \right\}$$

Portanto nossa primeira idéia consiste em tomar a decisão  $x$  que é a solução do PL:

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^n (y_i^0 - \bar{y}_i) x_i \\ \text{Suj.a} \quad & \sum_{i=1}^n y_i^0 x_i \leq M; \\ & x_i \geq 0 \ \forall i \end{aligned}$$

A solução do problema acima, supondo  $y_i^0$  e  $\bar{y}_i$  maiores do que zero, é:

- $x = 0$ ; se  $y_i^0 - \bar{y}_i \geq 0 \ \forall i$

Ou seja, não devo gastar nada se espero que o preço futuro de todos os ativos será menor do que o preço atual. Vou ter prejuízo onde quer que eu invista!!!

- $x_i = \begin{cases} \frac{M}{y_p^0}, & \text{se } i=p \text{ onde } p \text{ é tal que } y_p^0 < \bar{y}_p \text{ e } \frac{\bar{y}_p - y_p^0}{y_p^0} > \frac{\bar{y}_i - y_i^0}{y_i^0} \ \forall i \\ 0, & \text{se } i \neq p \end{cases}$

Ou seja, se para algum ativo eu espero ter lucro ( $\bar{y}_p > y_p^0$ ) então invisto todo o meu dinheiro no ativo que dá o maior retorno percentual ( $\frac{\bar{y}_p - y_p^0}{y_p^0} > \frac{\bar{y}_i - y_i^0}{y_i^0}$ ).

### Exemplo 1:

Suponhamos que um investidor tivesse 100 mil reais para investir em ações da Cemig, cujo preço inicial é de 100 reais, e da Vale do Rio Doce cujo preço é 50 reais. Suponhamos que existam 4 cenários futuros possíveis. São eles:



- O governo vai dar um subsídio às empresas de energia elétrica e a as empresas de mineração crescerão seguindo o fluxo geral do aquecimento da economia. Nesse caso os preços, daqui a um ano, das ações da Cemig e da Vale estarão em 130 e 58 reais respectivamente.
- O governo vai dar o subsídio para as empresas de energia porém a entrada no Brasil de uma multinacional na produção de aço faz com que os lucros da Vale não sejam tão expressivos. Como consequência as ações da Cemig estarão negociadas a 140 reais e da Vale a 54 reais.
- O governo não dá o subsídio às empresas de energia e os lucros da Cemig continuam no ritmo dos últimos anos. Já a Vale consolida sua liderança no mercado interno e consegue bons acordos internacionais. As cotações de Cemig e Vale passarão a ser 115 e 62 reais.
- A economia brasileira como um todo não vai bem e as ações da Cemig caem para 75 reais, enquanto as da Vale chegam a 46 reais.

Nossas variáveis de decisão seriam:

$x_1$   $\mapsto$  quantidade de ações da Cemig que compraríamos.

$x_2$   $\mapsto$  quantidade de ações da Vale que compraríamos.

Os preços iniciais seriam:

$$y_1^0 = 100 \text{ e } y_2^0 = 50$$

Os possíveis preços futuros seriam:

$$y^1 = \begin{pmatrix} 130 \\ 58 \end{pmatrix}; y^2 = \begin{pmatrix} 140 \\ 54 \end{pmatrix}; y^3 = \begin{pmatrix} 115 \\ 62 \end{pmatrix}; y^4 = \begin{pmatrix} 75 \\ 46 \end{pmatrix}$$

O preço médio de cada ação seria:

$$\bar{y}_1 = \frac{1}{4}(130 + 140 + 115 + 75) = 115;$$

$$\bar{y}_2 = \frac{1}{4}(58 + 54 + 62 + 46) = 55;$$

Restrições:

$$\text{gasto máximo: } 100 \cdot x_1 + 50 \cdot x_2 \leq 100.000$$

$$\text{positividade: } x_1 \geq 0; x_2 \geq 0$$

Problema a ser resolvido:

$$\begin{aligned} \min \quad & (100 - 115)x_1 + (50 - 55)x_2 \\ S.a \quad & 100 \cdot x_1 + 50 \cdot x_2 \leq 100.000 \\ & x_1 \geq 0; x_2 \geq 0 \end{aligned}$$

Solução:

$$x_1 = 1.000 \text{ e } x_2 = 0$$

Compraremos então 1.000 ações da Cemig e nenhuma da Vale.

É importante olhar para o lucro percentual e não para o absoluto. Isso fica claro no seguinte exemplo: Suponhamos que eu compre uma mansão hoje por 10 milhões de reais e venda daqui a um ano por 11 milhões de reais. Meu lucro absoluto seria de 1 milhão de reais, ou seja, um retorno percentual de 10%. Vejamos agora o caso em que eu pudesse comprar casas de 100 mil reais e vendê-las por 120 mil reais após um ano. Nesse caso meu lucro absoluto seria de 20 mil reais por casa (muito menor do que 1 milhão) e um retorno percentual de 20% que é o dobro da situação anterior. Porém com os mesmos 10 milhões que eu comprava uma mansão eu posso comprar 100 casas de 100 mil e meu lucro com o investimento seria de  $100 \times 20$  mil reais = 2 milhões de reais. Teria o dobro do lucro!

Um defeito grave dessa primeira idéia é que não fazemos a pergunta:

“Vale a pena tomar uma decisão que em média seja muito boa mas que tenha grandes perdas em alguns cenários?”

Essa situação ocorre no exemplo da Cemig e da Vale. Embora a Cemig desse um lucro médio de 15% e a Vale de 10% as ações da Cemig poderiam se desvalorizar até 25% enquanto com as ações da Vale, em qualquer que seja o cenário futuro, não perderíamos mais de 8%.

Isso nos remete ao conceito de risco. Vale a pena correr um risco grande em troca de uma boa média? Como medir esse risco?

---

## 2.3 Risco como distância até a média

---

Uma das alternativas de medir o risco de um investimento é olhar quanto a perda relacionada a uma decisão  $x$  pode se ‘distanciar’ da média. Quanto menor for essa distância mais segurança temos sobre o investimento.

Por exemplo, se tomarmos uma decisão  $x$  tal que  $f_j(x) = f$  para todo  $j$  então nossa perda não se distancia da média. Nesse caso teremos a certeza de que aconteça o que acontecer com o futuro da economia teremos um lucro de  $-f$ . O risco seria, portanto, nulo.

Tomando a decisão  $x$  indicamos por  $f_j$  a perda no cenário  $j$  e  $\bar{f}$  a perda média para essa decisão ( $\bar{f} = E(f) = \sum_{j=1}^m \frac{1}{m} f_j$ ).

Pensamos, então, na distância entre  $\begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$  e  $\begin{pmatrix} \bar{f} \\ \vdots \\ \bar{f} \end{pmatrix}$ .

Se optarmos por medir essa distância pela norma euclidiana teremos:

$$d = \left\| \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} - \begin{pmatrix} \bar{f} \\ \vdots \\ \bar{f} \end{pmatrix} \right\|_2 = \left\| \begin{pmatrix} f_1 - \bar{f} \\ \vdots \\ f_m - \bar{f} \end{pmatrix} \right\|_2 = \left( \sum_{j=1}^m (f_j - \bar{f})^2 \right)^{\frac{1}{2}} = \sqrt{m} \sigma(f).$$

Como  $\sqrt{m}$  é constante trabalharemos com o desvio padrão.

Se pensarmos no quadrado da norma euclidiana trabalharíamos com a variância que, por razões de diferenciabilidade, ajudaria na otimização das funções envolvidas.

Outra alternativa seria usar a norma da soma ( $d = \sum_{j=1}^m |f_j - \bar{f}|$ ) o que vai ser útil no modelo de Konno, o qual veremos mais adiante.

Em todos os casos caímos em um problema multiobjetivo pois queremos minimizar a perda média e o risco ao mesmo tempo. No caso de problemas desse tipo, geralmente, não

temos um ponto ótimo no sentido habitual e sim uma curva eficiente onde os pontos desta não podem ser melhorados nos dois objetivos ao mesmo tempo. Ou seja uma decisão  $x^*$  representa um ponto da curva eficiente se, e somente se, não existem decisões viáveis que apresentem simultaneamente perda média e risco menores do que as de  $x^*$ . Isso nos remete a noção de ótimo a Pareto.

Podemos notar que fixado  $p$  o par  $(p, r) \in C$  somente se

$$r = \inf \{ \text{risco}(x) \text{ tais que } x \in \Omega \text{ e } \bar{f}(x) \leq p \}$$

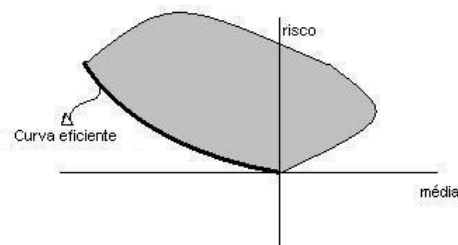


Figura 2.1: Cada ponto  $(p, r)$  na região em cinza representa uma decisão factível com perda média  $p$  e risco  $r$ .

Pensando em contornar o inconveniente do problema multiobjetivo chegamos aos modelos de Markowitz e Konno.

### 2.3.1 Modelo de Markowitz

Nessa seção vamos evitar o problema multiobjetivo procurando diminuir ao máximo um dos objetivos de tal modo que possamos garantir que o outro esteja em um nível aceitável. Faremos isso minimizando um dos objetivos e colocando como restrição que o outro esteja abaixo de nossa tolerância.

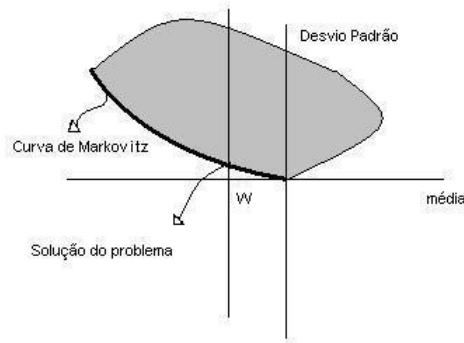
$$\begin{aligned} \text{Min} \quad & \text{Objetivo 1} \\ \text{S.a} \quad & \text{Objetivo 2} \leq \text{tolerância}; \\ & x \in \Omega \end{aligned}$$

No modelo de Markowitz tomamos como objetivo a ser minimizado o risco, e colocamos como restrição que a perda média não supere uma constante  $W$ . Por razões de

simplicidade do problema de otimização a ser resolvido, as quais veremos mais adiante, vamos usar a variância como medida de risco.

Nosso problema passa a ser:

$$\begin{aligned} \text{Min} \quad & \text{Variância } [f(x, y)] \\ \text{Suj.a} \quad & E[f(x, y)] \leq W; \\ & x \in \Omega \end{aligned}$$



Sob as hipóteses das funções de perda lineares podemos calcular a variância da seguinte maneira:

$$\begin{aligned} f(x, y) - E[f(x, y)] &= \sum_{i=1}^n (y_i^0 - y_i^j) x_i - \sum_{i=1}^n (y_i^0 - \bar{y}_i) x_i \\ &= \sum_{i=1}^n (\bar{y}_i - y_i^j) x_i \end{aligned}$$

Daí:

$$\begin{aligned}
\text{Variância } [f(x, y)] &= \sigma^2[f(x, y)] = E [(f(x, y) - E[f(x, y)])^2] \\
&= \sum_{j=1}^m \frac{1}{m} \left[ \sum_{i=1}^n (\bar{y}_i - y_i^j) x_i \right]^2 = \sum_{j=1}^m \frac{1}{m} \left( \bar{Y}^{jT} x \right)^2 \\
&= \sum_{j=1}^m \frac{1}{m} x^T \left( \bar{Y}^j \bar{Y}^{jT} \right) x = x^T \left( \frac{1}{m} \sum_{j=1}^m \bar{Y}^j \bar{Y}^{jT} \right) x \\
&= x^T \left( \frac{1}{m} \sum_{j=1}^m M^j \right) x = x^T A x,
\end{aligned}$$

$$\text{onde } \bar{Y}^j = \begin{pmatrix} y_1^j - \bar{y}_1 \\ \vdots \\ y_n^j - \bar{y}_n \end{pmatrix}; M^j = \bar{Y}^j \bar{Y}^{jT} \text{ e } A = \frac{1}{m} \sum_{j=1}^m M^j.$$

Como cada matriz  $M^j$  é semi definida positiva temos que  $A$  também é, portanto nossa função objetivo é convexa.

O modelo clássico de Markowitz é:

$$\begin{aligned}
\text{Min } & x^T A x \\
\text{Suj.a } & \sum_{i=1}^n y_i^0 x_i \leq M; \\
& \sum_{i=1}^n (y_i^0 - \bar{y}_i) x_i \leq W; \\
& x_i \geq 0 \quad \forall i
\end{aligned} \tag{2.1}$$

O problema acima é quadrático e convexo. A literatura garante que problemas desse tipo são relativamente fáceis de serem resolvidos. Isso justifica as escolhas dos objetivos e de como medir o risco.

No modelo de Markowitz conseguimos de certa forma atender nossa expectativa de encontrar um ponto ótimo do problema multiobjetivo uma vez que se (2.1) admite solução  $x$  então provavelmente essa solução correspondente a um ponto da curva eficiente. De fato,

para  $W \geq 0$  a solução ótima seria  $x = 0$  pois claramente é factível e  $x^T Ax = 0$ , o que pela definição de semi-positividade garante a otimalidade do problema. Intuitivamente se não estamos exigindo lucro, a decisão de menor risco é não investir o dinheiro em lugar algum. No caso em que  $W < 0$  se  $x$  não pertencesse à curva eficiente existiria  $x_1 \in \Omega$  tal que  $E[f(x_1, y)] \leq E[f(x, y)] \leq W$  e portanto  $x_1$  seria factível, e  $x_1^T Ax_1 \leq x^T Ax$  com alguma das desigualdades estritas. Supondo  $x_1^T Ax_1 < x^T Ax$  então  $x$  claramente não é solução de (2.1), pois estamos exibindo um ponto factível com menor valor de função objetivo, que é um absurdo uma pela hipótese.

Podemos obter outra característica de  $A$  que nos fornece uma idéia do comportamento de um ativo em relação a outro. Com isso podemos conseguir informações sobre a sensibilidade e o comportamento geral do problema. Para isso definimos  $Y_i$  como a variável aleatória que a cada cenário  $j$  associa o preço do ativo  $i$  na situação  $j$ , ou seja  $Y_i(j) = y_i^j$  e conseqüentemente  $\bar{y}_i = E(Y_i)$ . Voltando na definição de  $M^j$  temos que:

$M_{ik}^j = \bar{Y}_i^j \cdot \bar{Y}_k^j = (y_i^j - \bar{y}_i) \cdot (y_k^j - \bar{y}_k)$ . Daí:

$$A_{ik} = \frac{1}{m} \sum_{j=1}^m (y_i^j - \bar{y}_i) \cdot (y_k^j - \bar{y}_k) = \frac{1}{m} \sum_{j=1}^m (Y_i(j) - E(Y_i)) \cdot (Y_k(j) - E(Y_k))$$

Conseguimos assim apresentar duas relações para  $A_{ik}$

$$A_{ik} = \begin{cases} \frac{1}{m} \langle Z_i, Z_k \rangle; \text{ onde } Z_i = \begin{pmatrix} (y_i^1 - \bar{y}_i) \\ \vdots \\ (y_i^m - \bar{y}_i) \end{pmatrix} \\ E[(Y_i - E(Y_i)) \cdot (Y_k - E(Y_k))] = Cov(Y_i, Y_k) \end{cases}$$

Em conseqüência da segunda relação  $A$  é denominada matriz de covariância do problema.

A covariância entre duas variáveis aleatórias nos dá a idéia de como elas estão ligadas [13]. Porém a covariância não é adimensional o que, várias vezes, a torna difícil de ser interpretada. Pensando nisso definimos o coeficiente de correlação

$$\begin{aligned}
\rho(Y_i, Y_k) &= \frac{\text{cov}(Y_i, Y_k)}{\sigma(Y_i)\sigma(Y_k)} = \frac{A_{ik}}{[E[(Y_i - E(Y_i))^2]^{1/2} \times [E[(Y_k - E(Y_k))^2]^{1/2}]} \\
&= \frac{\frac{1}{m} \langle Z^i, Z^k \rangle}{[\sum_{j=1}^m \frac{1}{m} (y_i^j - \bar{y}_i)^2]^{1/2} \times [\sum_{j=1}^m \frac{1}{m} (y_k^j - \bar{y}_k)^2]^{1/2}} \\
&= \frac{\frac{1}{m} \langle Z^i, Z^k \rangle}{(\frac{1}{m})^{1/2} \|Z^i\| \times (\frac{1}{m})^{1/2} \|Z^k\|} \\
&= \frac{\langle Z^i, Z^k \rangle}{\|Z^i\| \|Z^k\|} \\
&= \text{cosseno do ângulo entre } Z^i \text{ e } Z^k
\end{aligned}$$

Portanto  $\rho(Y_i, Y_k)$  é um parâmetro entre  $-1$  e  $1$  e, então, podemos definir a matriz de correlação  $A^*$  onde  $A_{ik}^* = \rho(Y_i, Y_k)$ . Se  $\rho(Y_i, Y_k) = 0$  ou, equivalentemente, se  $Z_i$  é perpendicular a  $Z_k$  dizemos que  $Y_i$  e  $Y_k$  não estão correlacionadas. Isso acontece por exemplo quando são independentes, ou seja, a informação sobre o preço do ativo  $i$  não nos diz nada sobre o preço do ativo  $k$ . Se quando o preço do ativo  $i$  aumentar tivermos que o preço do ativo  $k$  também aumenta (na mesma proporção) teremos que  $\rho(Y_i, Y_k) = 1$  e se diminui (também na mesma proporção) teremos  $\rho(Y_i, Y_k) = -1$ . No primeiro caso  $Z_i$  e  $Z_k$  estão na mesma direção e sentido e no segundo, um tem sentido oposto do outro.

## Exemplo 2

Um grande comerciante agropecuário dispõe de um milhão de reais para comprar boi, frango, leite e queijo a serem entregues daqui a 1 mês. O preço inicial dos produtos é respectivamente 50 reais a arroba, 2 reais o quilo, 1 real o litro e 5 reais o quilo. Ao receber os produtos ele revende-os pelo preço de mercado.

Querendo arriscar-se no mercado de ações ele inclui a possibilidade de investir parte do dinheiro em ações da IBM cujo preço inicial é de 200 reais. Os cenários previstos para daqui a um mês são:

- **Cenário 1**

Há um surto de febre aftosa



As condições climáticas são boas  
A IBM consegue bons acordos internacionais

- **Cenário 2**

Há um surto de febre aftosa  
As condições climáticas são boas  
Os acordos internacionais da IBM não se concretizam

- **Cenário 3**

Há um surto de febre aftosa  
As condições climáticas não são boas  
A IBM consegue bons acordos internacionais

- **Cenário 4**

Há um surto de febre aftosa  
As condições climáticas não são boas  
Os acordos internacionais da IBM não se concretizam

- **Cenário 5**

A suspeita de febre aftosa é afastada  
As condições climáticas são boas  
A IBM consegue bons acordos internacionais

- **Cenário 6**

A suspeita de febre aftosa é afastada  
As condições climáticas são boas  
Os acordos internacionais da IBM não se concretizam

- **Cenário 7**

A suspeita de febre aftosa é afastada  
As condições climáticas não são boas  
A IBM consegue bons acordos internacionais

- **Cenário 8**

A suspeita de febre aftosa é afastada

As condições climáticas não são boas

Os acordos internacionais da IBM não se concretizam

A tabela abaixo mostra os preços futuros dos ativos em cada um dos cenários:

	ativo 1	ativo 2	ativo 3	ativo 4	ativo 5
	arroba boi	quilo frango	litro leite	quilo queijo	ação IBM
Cenário 1	35	2.50	0.85	4.20	220
Cenário 2	36	2.45	0.87	4.25	202
Cenário 3	40	2.35	1.10	5.50	218
Cenário 4	41	2.33	1.09	5.45	203
Cenário 5	55	1.85	0.90	4.50	219
Cenário 6	55	1.87	0.92	4.55	201
Cenário 7	60	1.80	1.15	5.70	219
Cenário 8	59	1.82	1.15	5.75	202

A matriz de covariância do problema é:

$$\begin{pmatrix} 98.4844 & -2.8733 & 0.5352 & 2.7766 & -3.4375 \\ -2.8733 & 0.0848 & -0.0130 & -0.0679 & 0.1081 \\ 0.5352 & -0.0130 & 0.0148 & 0.0765 & -0.0419 \\ 2.7766 & -0.0679 & 0.0765 & 0.3948 & -0.1562 \\ -3.4375 & 0.1081 & -0.0419 & -0.1562 & 72.7500 \end{pmatrix}$$

A matriz de correlação é:

$$\begin{pmatrix} 1.0000 & -0.9945 & 0.4425 & 0.4453 & -0.0406 \\ -0.9945 & 1.0000 & -0.3666 & -0.3713 & 0.0435 \\ 0.4425 & -0.3666 & 1.0000 & 0.9989 & -0.0403 \\ 0.4453 & -0.3713 & 0.9989 & 1.0000 & -0.0292 \\ -0.0406 & 0.0435 & -0.0403 & -0.0292 & 1.0000 \end{pmatrix}$$

Podemos então tirar algumas conclusões sobre o comportamento dos ativos.

1. O coeficiente de correlação entra as ações da IBM e qualquer outro ativo é, em módulo, muito pequeno. Isso indica não há relação entre o preço das ações da IBM e dos produtos agropecuários. Esse fato já era esperado uma vez que os produtos não são concorrentes, não utilizam a mesma matéria prima e atingem a sociedade com intuits diferentes.
2. O coeficiente próximo a -1 entre a arroba bovina e o quilo do frango mostra que esses produtos seguem tendências opostas, quando um fica mais caro o preço do outro diminui. Isso deve-se ao fato de serem produtos de substituição. Com um problema no rebanho bovino as exportações são suspensas e substituídas pelo consumo de frango. Esses produtos em conjunto podem diminuir o risco em nosso investimento pois quando perdemos em um ganhamos com outro.
3. O coeficiente próximo de 1 entre queijo e leite reflete a forte ligação do preço entre esses produtos. Isso já era esperado uma vez que eles usam a mesma matéria prima e estão submetidos a fiscalizações sanitárias muito semelhantes.
4. Os coeficientes intermediários de aproximadamente 0.44 entre boi-leite e boi-queijo mostram que há uma ligação positiva entre esses produtos embora não seja muito forte. Já os coeficientes por volta de -0.37 entre frango-queijo e frango-leite mostram que os preços destes produtos seguem caminhos relativamente contrários. Isso já era esperado uma vez que leite e queijo acompanham moderadamente o boi que por sua vez rivaliza com o frango.

A seguir apresentamos as soluções do problema para três valores da constante  $W$  do modelo de Markowitz. Podemos ver que aumentando  $|W|$  estamos exigindo um lucro maior, conseqüentemente estamos sujeitos a um risco maior.

Para  $W = -10$  mil reais a solução ótima é comprar 3.635 arrobas de boi, 126.132 quilos de frango, não comprar nem leite nem queijo e comprar 318 ações da IBM. Para essa decisão nossa perda média é de exatamente -10 mil reais, o dinheiro investido é 502.350,15 reais, ou seja, para esse pequeno retorno médio só é preciso investir a metade do dinheiro.

Com isso, como a parte não aplicada não tem risco, conseguimos diminuir a variância do investimento. Nesse caso nosso risco foi de  $2,3 \times 10^7$ .

Para  $W = -25$  mil já somos obrigados a investir todo nosso dinheiro. A solução do problema seria comprar 6.222 arrobas de boi, 222.187 quilos de frango, não comprar nem leite nem queijo e comprar 1.222 ações da IBM. Nesse caso, mesmo investindo todo o dinheiro, tomamos uma decisão que dividia boa parte do dinheiro entre frango e boi (44% e 31% respectivamente). O risco para essa decisão foi de  $1,7 \times 10^8$ .

Para  $W = -50$  mil reais a decisão ótima é comprar 530 arrobas de boi, 59.304 quilos de frango, 16.739 litros de leite, não comprar queijo e comprar 4.191 ações da IBM. Nesse caso, além de investir todo o dinheiro, não temos a opção de dividir uma alta porcentagem da aplicação entre boi e frango. Para obtermos um lucro médio tão grande devemos por 84% de nosso dinheiro em ações da IBM. Essa alternativa, embora dê uma perda média muito baixa, é bastante arriscada uma vez que se a ação da IBM cair teremos um grande prejuízo. Com isso nosso risco sobe para  $1,4 \times 10^9$ .

Visto a importância da correlação entre os ativos voltamos ao exemplo 1, cuja decisão pertence ao  $\mathbb{R}^2$  (quantidade de ações da Vale e da Cemig), para compararmos os métodos estudados de maneira que possamos fazer uma análise gráfica do modelo de Markowitz. A matriz de covariância do problema seria:

$$A = \begin{pmatrix} 612,5 & 95,0 \\ 95,0 & 35,0 \end{pmatrix}$$

Logo nosso problema seria:

$$\min (x_1, x_2) \begin{pmatrix} 612,5 & 95,0 \\ 95,0 & 35,0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$S.a \quad -15x_1 - 5x_2 \leq W$$

$$100x_1 + 50x_2 \leq 100.000$$

$$x_1 \geq 0; x_2 \geq 0$$

Se exigirmos um lucro médio mínimo de 11.500,00 reais a solução seria comprar 300

ações da Cemig e 1.400 da Vale. Percebemos que essa é uma decisão muito mais equilibrada do que a obtida anteriormente, pois investimos nos dois ativos e, além disso, colocamos a maioria de nosso dinheiro no ativo mais estável, o qual sabemos que não sofrerá grandes perdas.

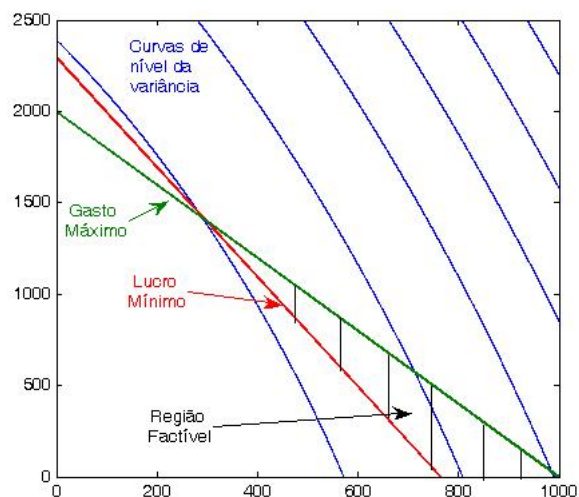


Figura 2.2: A figura mostra as curvas de nível da função objetivo e as restrições do problema

Da figura 2.2 fica clara a relação entre a solução do problema e a posição relativa da restrição da perda média mínima e da quantidade de dinheiro disponível, incluindo a situação de infactibilidade, que não apresentamos em nossos exemplos. Um comentário importante a ser feito aqui é que essa distância não foi tomada percentualmente o que a princípio é um problema pois é diferente se um investimento pode gerar uma perda que se distancie 10 reais de sua perda média se o lucro médio para essa decisão for de 100 reais ou de 1 milhão de reais.

Felizmente não precisamos nos preocupar com essa sutileza uma vez que o risco é uma função decrescente da perda média na fronteira eficiente. De fato, caso contrário existiriam dois pontos na fronteira eficiente sendo um deles com maior perda média e maior risco do que o outro, o que contradiz o conceito de eficiência. Logo, para  $W \leq 0$ , o mínimo

do problema 2.1 sempre acontece quando  $\bar{f}(x) = W$ . Outra maneira de ver isso é que se  $x$  é factível com  $\bar{f}(x) = K < W < 0$  então  $r = W/K \in (0, 1)$  e conseqüentemente  $y = rx$  é viável e possui valor de função objetivo  $\sigma^2(y) = r^2\sigma^2(x) \leq \sigma^2(x)$ . Portanto podemos pensar que em (2.1) estamos minimizando o risco sob a restrição de que a perda média seja igual a  $W$ . Dessa forma, como o referencial é constante, podemos tratar a distância relativa como absoluta.

Uma outra opção seria substituir  $A$  por  $A^*$  em (2.1), obtendo assim:

$$\begin{aligned} \text{Min} \quad & x^T A^* x \\ \text{Suj.a} \quad & \sum_{i=1}^n y_i^0 x_i \leq M; \\ & \sum_{i=1}^n (y_i^0 - \bar{y}_i) x_i \leq W; \\ & x_i \geq 0 \quad \forall i \end{aligned}$$

O problema acima difere do anterior apenas por uma mudança de escala na função objetivo e ainda é convexo. A razão disso vem do fato de  $A_{ik}^* = m \frac{A_{ik}}{\|Z^i\| \|Z^k\|}$  implicar que  $A^* = mD^T A D$ , onde  $D$  é a matriz diagonal tal que  $D_i = 1/\|Z^i\|$ .

A idéia dual ao modelo de Markowitz seria:

$$\begin{aligned} \text{Min} \quad & E[f(x, y)] \\ \text{Suj.a} \quad & \text{Variância } [f(x, y)] \leq W; \\ & x \in \Omega \end{aligned}$$

Ou seja:

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^n (y_i^0 - \bar{y}_i) x_i \\ \text{Suj.a} \quad & x^T A x \leq W; \\ & x \in \Omega \end{aligned}$$

Nesse caso não teríamos a conveniência do problema quadrático.

### 2.3.2 Modelo de Konno

Outra alternativa para contornar o problema multi-objetivo é definir uma função de mérito, onde damos peso ( $p > 0$ ) a um dos objetivos a serem minimizados. Poderíamos tomar a decisão que fosse solução do problema:

$$\begin{aligned} \text{Min} \quad & \text{Objetivo 1 } (x) + p \cdot \text{Objetivo 2 } (x) \\ \text{Suj.a} \quad & x \in \Omega \end{aligned}$$

Podemos demonstrar que a solução ( $x$ ) do problema formulado dessa forma corresponde a um ponto da curva eficiente.

#### Demonstração:

*De fato, caso contrário existiria  $x^* \in \Omega$  tal que*

$$\text{Objetivo 1 } (x^*) \leq \text{Objetivo 1 } (x)$$

*e*

$$\text{Objetivo 2 } (x^*) \leq \text{Objetivo 2 } (x),$$

*com uma das desigualdades estritas.*

*Daí,  $\text{Objetivo 1 } (x^*) \leq \text{Objetivo 1 } (x)$  e  $p \cdot \text{Objetivo 2 } (x^*) \leq p \cdot \text{Objetivo 2 } (x)$  com uma das desigualdades estritas.*

*Logo  $\text{Objetivo 1 } (x^*) + p \cdot \text{Objetivo 2 } (x^*) < \text{Objetivo 1 } (x) + p \cdot \text{Objetivo 2 } (x)$ .*

*Essa última afirmação é um absurdo pois é equivalente a falar que a função objetivo em  $x^*$  é menor do que em  $x$  o que contraria o fato de  $x$  ser minimizador do problema acima. Sendo assim  $x$  corresponde a um ponto da fronteira eficiente.*

Uma escolha possível para os objetivos 1 e 2 seria, respectivamente,  $E[f(x, y)]$  e  $\sigma[f(x, y)]$ .

$$\begin{aligned} \text{Min} \quad & E[f(x, y)] + p \cdot \sigma[f(x, y)] \\ \text{Suj.a} \quad & x \in \Omega, \end{aligned}$$

onde  $p$  é um parâmetro que reflete nossa aversão ao risco, Quanto maior  $p$  mais importância estamos dando ao risco. Optamos aqui por trabalhar com  $\sigma[f(x, y)]$  e não com  $\sigma^2$  pois  $\sigma$  e  $E[f(x, y)]$  estão com a mesma unidade, logo é mais plausível olhar o problema dessa forma.

Uma reformulação, devido a Konno, do problema anterior seria tomar a norma da soma no lugar da norma euclidiana. Nesse caso, nosso problema se tornaria:

$$\begin{aligned} \text{Min} \quad & E[f(x, y)] + p \cdot \sum_{j=1}^m |f(x, y_j) - E[f(x, y)]| \\ \text{Suj.a} \quad & x \in \Omega \end{aligned}$$

Sob as hipóteses de linearidade podemos escrever nosso problema como:

$$\begin{aligned} \text{Min} \quad & E[f(x, y)] + p \cdot \sum_{j=1}^m \left| \sum_{i=1}^n (\bar{y}_i - y_i^j) x_i \right| \\ \text{Suj.a} \quad & \sum_{i=1}^n y_i^0 x_i \leq M \\ & x_i \geq 0 \quad \forall i \end{aligned}$$

Podemos adaptar o problema anterior, acrescentando  $m$  variáveis  $w_j$ , para contornar a não diferenciabilidade:



$$\begin{aligned}
\text{Min} \quad & \sum_{i=1}^n (y_i^0 - \bar{y}_i)x_i + p \cdot \sum_{j=1}^m w_j \\
\text{Suj.a} \quad & w_j \geq \sum_{i=1}^n (\bar{y}_i - y_i^j)x_i; j=1\dots m \\
& w_j \geq -\sum_{i=1}^n (\bar{y}_i - y_i^j)x_i; j=1\dots m \\
& \sum_{i=1}^n y_i^0 x_i \leq M \\
& x_i \geq 0 \quad \forall i
\end{aligned}$$

Visto dessa forma a decisão a ser tomada volta a ser a solução de PL, assim como quando minimizávamos apenas a perda média. Uma desvantagem dessa abordagem é o inconveniente de que aumentamos uma variável para cada cenário, o que geralmente aumenta demasiadamente a dimensão do problema, uma vez que se desejarmos ter confiabilidade no modelo, quanto maior o número de cenários, melhor.

Outras medidas de risco importantes, não baseadas em distância até a média, são a VaR e a C-VaR. Pelo fato delas representarem a parte fundamental desse trabalho vamos expô-las em um capítulo a parte.

---

---

# CAPÍTULO 3

---

## VaR e C-VaR como medidas de risco

---

### 3.1 A VaR (*Value at Risk*)

---

Uma das medidas de risco mais consagradas no mercado financeiro é a *VaR* (*Value at Risk*). A *VaR* com um nível de confiança  $\beta$  é uma função que a cada decisão  $x$  associa a resposta da seguinte pergunta:

“Com um nível de confiança  $\beta$  qual a maior perda que podemos ter se tomarmos a decisão  $x$ ?”

Formalmente para cada  $x \in \Omega$  ordenamos  $f_{i_1(x)}(x) \leq f_{i_2(x)}(x) \leq \dots \leq f_{i_m(x)}(x)$ , tomamos  $p \approx \beta m$  de forma que  $\frac{p}{m} \leq \beta \leq \frac{p+1}{m}$  e definimos  $VaR_\beta(x) = f_{i_p(x)}(x)$ .

Por simplicidade denotaremos algumas vezes  $VaR_\beta(x) = f(x)$ .

O problema de minimizar a *VaR* pode ser visto como uma evolução do problema *minimax*, o qual é um caso particular correspondente a *VaR* com nível de confiança de 100%. Ou seja a  $VaR_{100\%}(x) = \max\{f_i(x)\}$ . Como exemplo ilustrativo consideremos o caso em que nossa única variável de decisão seja a quantidade de dinheiro que investiremos em

um fundo de ações do Banco do Brasil. Suponhamos que nosso capital se limite a 10 mil reais. Suponhamos também que existam 5 cenários possíveis para o futuro econômico. As funções de perda em cada um deles são:

$$f_1(x) = x^2$$

$$f_2(x) = 3.5x$$

$$f_3(x) = 40.000 \operatorname{sen} \left( \frac{\pi x}{5.000} \right)$$

$$f_4(x) = (x - 12.000)^2$$

$$f_5(x) = (x - 12.000)^2 - 60.000$$

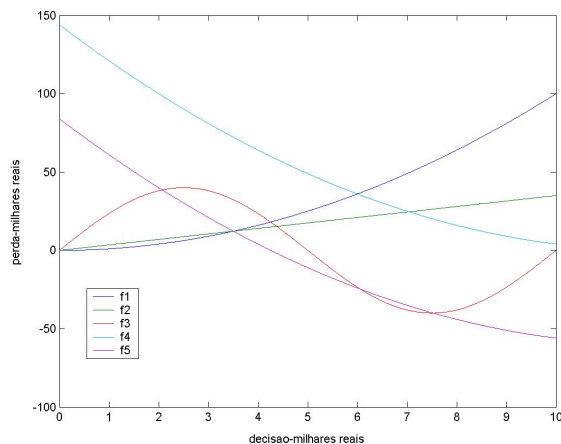


Figura 3.1: Gráfico das funções de perda correspondentes aos cinco cenários.

As figuras 3.2 e 3.3 a seguir mostram a VaR para  $\beta = 100\%$ , correspondente ao máximo das  $f_i(x)$ , e para  $\beta = 80\%$ . Nesse último caso  $80\%$  de  $5 = 4$ , logo  $VaR_{80\%}$  corresponde à segunda maior perda que podemos ter em cada uma das decisões.

A  $VaR$  não leva em consideração que queremos também, ao mesmo tempo que diminuamos o risco, obter uma média de lucro razoável. Isso poderia ser feito como nas formulações anteriores usando a  $VaR$  como medida de risco. Um exemplo disso seria minimizar  $\bar{f} + p \cdot VaR$ . Consideraremos aqui o caso de minimizar apenas a  $VaR$  (e pos-

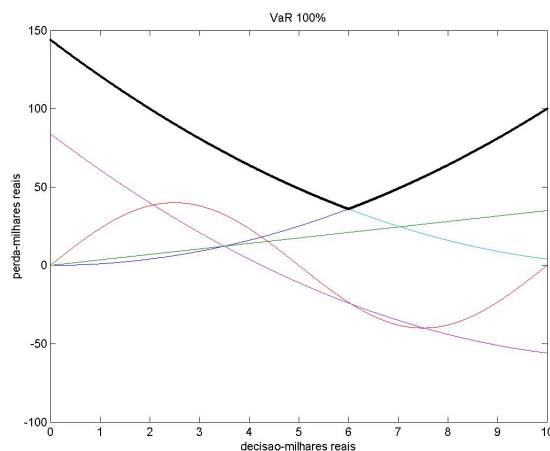


Figura 3.2: A curva destacada corresponde a  $VaR_{100\%}$ .

teriormente a  $C-VaR$ ) pois esse problema por si só é de grande interesse em otimização, sendo parte de uma ampla classe de problemas chamada OVO (*Order-Value-Optimization*) cujas aplicações se estendem muito além de problemas ligados a economia.

A  $VaR$  é uma função não diferenciável em alguns pontos e com vários mínimos locais não globais, essas características tornam sua minimização uma tarefa difícil. Na seção 3.5 apresentaremos algoritmos duais suaves que tratam esses problemas. Algoritmos primais e teoria para minimizar a  $VaR$ , inclusive com hipóteses mais fracas do que as apresentadas aqui (exige-se apenas que cada  $f_i$  seja continuamente diferenciável e que  $\Omega$  seja fechado e convexo), são encontrados em [7] e [9].

Os resultados encontrados nos artigos mencionados acima baseiam-se nas definições e no teorema abaixo:

**Definição:**  $I_\epsilon(x) = \{j \in \{1 \cdots m\} \mid f(x) - \epsilon \leq f_j(x) \leq f(x) + \epsilon\}$

Intuitivamente estamos escolhendo as funções são candidatas a se tornar  $VaR_\beta(x')$  para  $x'$  perto de  $x$ .

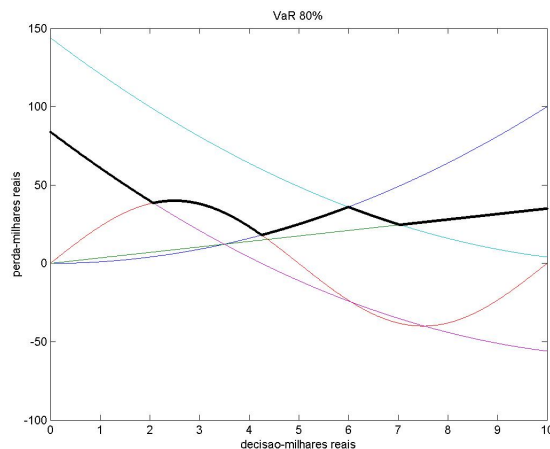


Figura 3.3: A curva destacada corresponde a  $VaR_{80\%}$ .

**Definição:** Dizemos que  $x^*$  é  $\epsilon$ -estacionário se:

$$D \equiv \{d \in \mathbb{R}^n \mid x + d \in \Omega \text{ e } \nabla f_j(x)^T d < 0 \ \forall j \in I_\epsilon(x^*)\} = \emptyset.$$

$D$  é o conjunto das direções factíveis ( $x^* + d \in \Omega$ ) e de descida para as funções  $f_j$  que, numa vizinhança de  $x^*$ , podem assumir o valor da  $VaR$ . Logo é natural definir que  $x^*$  será  $\epsilon$ -estacionário se não houver direção de descida para todas  $f_j(x)$  candidatas a se tornarem  $f(x)$ .

**Teorema 3.1.1.** *Se  $x^*$  é um minimizador local de  $f(x)$  sujeita a  $x \in \Omega$  e  $\epsilon \geq 0$  então,  $x^*$  é  $\epsilon$ -estacionário.*

A idéia do teorema é que se  $x^*$  é um minimizador local então não existe direção factível de descida para  $f$  em  $x^*$ . Com mais forte razão não pode haver direção factível de descida para toda  $f_j$  com  $j \in I_\epsilon(x^*)$  uma vez que para todo  $x$  numa vizinhança de  $x^*$  temos  $f(x) = f_{j(x)}(x)$  para algum  $j(x) \in I_\epsilon(x^*)$ .

---

## 3.2 A C-VaR (*Conditional Value at Risk*)

---

Uma medida de risco com propriedades de otimização muito mais atraentes é a *C-VaR* (*Conditional Value at Risk*). A *C-VaR* com nível de confiança  $\beta$  é uma função que associa

a cada  $x$  a média dos valores  $f_i(x)$  maiores do que a  $VaR_\beta(x)$ . Ou seja, a  $C-VaR$  mede qual é a perda média que teremos nos cenários descartados pela VaR.

$$\begin{aligned} C-VaR_\beta(x) &= \frac{1}{m-p} \sum_{f_i(x) > VaR_\beta(x)} f_i(x) \\ &= \frac{1}{m-p} \sum_{j=p+1}^m f_{i_{j(x)}}(x) \end{aligned}$$

No caso ilustrativo da seção anterior em que temos cinco cenários sabemos que para  $\beta = 80\%$  a  $VaR_\beta(x)$  corresponde à segunda maior perda que podemos ter, desse modo só existe uma perda acima da  $VaR_\beta(x)$  logo a média desta é igual a ela própria.

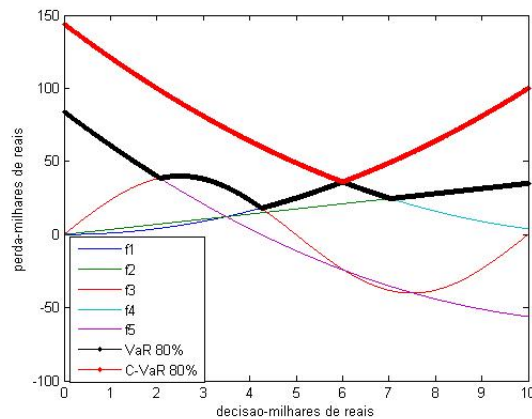


Figura 3.4: A  $C-VaR_{80\%}$  corresponde a  $VaR_{100\%}$ .

Já a  $VaR_{60\%}$  corresponde à terceira maior perda que podemos ter, nesse caso existem duas perdas acima. Sendo assim, a  $C-VaR_{60\%}$  é a média entre a segunda maior e a maior perda para cada decisão.

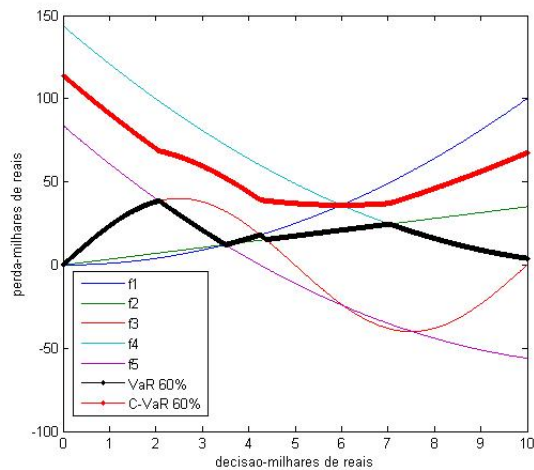


Figura 3.5: A  $C\text{-VaR}_{60\%}$  corresponde à média das 2 maiores perdas que podemos ter.

---

### 3.3 Coerência

---

Nesta seção apresentaremos e discutiremos sobre o conceito de coerência segundo Artzner, Delbaen, Eber e Heath [10]. A  $C\text{-VaR}$  é uma medida de risco coerente nesse sentido, o que para vários autores é uma redundância uma vez que, para eles, a definição de medida de risco é exatamente cumprir os axiomas de coerência.

Dentre os axiomas de coerência destaca-se a sub-aditividade. Porém acreditamos que a sub-aditividade na verdade não reflete a característica que Artzner, Delbaen, Eber e Heath queriam impor a uma medida de risco. Isso deve-se ao fato de que uma medida ( $\rho$ ) não está definida no conjunto de decisões factíveis ( $\Omega$ ) e sim num conjunto abstrato de variáveis aleatórias ( $V$ ).

Relembramos aqui que a uma decisão  $x \in \Omega$  podemos associar a variável aleatória  $X$  tal que  $X(j) = f_j(x)$ . Como  $X$  está definida em um domínio finito podemos vê-la como um vetor.

$$x \rightarrow \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

Podemos ver que  $V$  define classes de equivalência em  $\Omega$ . Diremos que duas decisões são equivalentes se estão associadas a uma mesma variável aleatória. Essa é uma definição intuitiva pois é natural ser indiferente a duas decisões cujas perdas são as mesmas em qualquer que seja o cenário futuro. Esse fato fez com que Artzner, Delbaen, Eber e Heath definissem uma medida de risco em  $V$  e não em  $\Omega$ . Claramente existe uma relação intrínseca entre  $\Omega$  e  $V$ . Portanto poderíamos redefinir naturalmente  $\rho : \Omega \rightarrow \mathbb{R}$ . Porém, nesse caso, teríamos que a **C-VaR NÃO é sub-aditiva**, a não ser sob certas hipóteses. Acreditamos que esses casos sim trazem a essência da sub-aditividade e refletem que o problema esteja bem modelado.

Vamos definir aqui o que é uma medida de risco coerente segundo Artzner, Delbaen, Eber e Heath. Faremos algumas adaptações para tratarmos de funções de perda e não de lucro.

**Definição 3.3.1.** *Seja  $V$  um conjunto de variáveis aleatórias a valores reais.*

*Uma função  $\rho : V \rightarrow \mathbb{R}$  é dita uma medida de risco coerente se satisfaz os axiomas:*

- i) monotonicidade:  $X, Y \in V, X \geq Y \Rightarrow \rho(X) \geq \rho(Y)$*
- ii) invariância por translação:  $a \in \mathbb{R}, X, X + a \in V \Rightarrow \rho(X + a) = \rho(X) + a$*
- iii) positividade homogênea:  $h > 0, X, hX \in V \Rightarrow \rho(hX) = h\rho(X)$*
- iv) sub-aditividade:  $X, Y, X + Y \in V \Rightarrow \rho(X + Y) \leq \rho(X) + \rho(Y)$*

O axioma da sub-aditividade procura modelar o fato de que a diversificação de um portfólio ajuda a diminuir o risco do investimento. Outra interpretação seria que caso *iv)* não valesse seria possível uma empresa conseguir um menor risco, com um mesmo investimento, se o quebrasse em dois e o fizesse de forma independente através de departamentos diferentes.



A demonstração de que tanto a VaR quanto a C-VaR satisfazem os três primeiros axiomas é imediata. Já a sub-aditividade só vale para a C-VaR. A seguir daremos um exemplo onde tal axioma falha para a VaR mesmo quando as funções de perda são lineares e em seguida demonstraremos que realmente a C-VaR é sub-aditiva.

**Exemplo 3.3.2.** *Suponhamos que as funções de perda nos cinco cenários possíveis sejam as funções lineares abaixo:*

$$f_1(x) = 0x_1 + 3x_2$$

$$f_2(x) = 1x_1 + 0x_2$$

$$f_3(x) = 2x_1 + 0x_2$$

$$f_4(x) = 0x_1 + 0x_2$$

$$f_5(x) = 0x_1 + 0x_2$$

*Podemos ver que a VaR não é sub-aditiva pois:*

$$\begin{aligned} x &= (1, 0)^T \Rightarrow X = (0, 1, 2, 0, 0)^T \\ y &= (0, 1)^T \Rightarrow Y = (3, 0, 0, 0, 0)^T \\ x + y &= (1, 1)^T \Rightarrow X + Y = (3, 1, 2, 0, 0)^T \end{aligned}$$

*Logo:*

$$VaR_{80\%}(X) = 1, VaR_{80\%}(Y) = 0 \text{ e } VaR_{80\%}(X + Y) = 2$$

*conseqüentemente,*

$$VaR_{80\%}(X + Y) > VaR_{80\%}(X) + VaR_{80\%}(Y).$$

**Teorema 3.3.3.** *A C-VaR é sub-aditiva.*

*Demonstração:*

$$\begin{aligned}
C\text{-VaR}_\beta(X + Y) &= \frac{1}{m - p} \sum_{j=p+1}^m [f_i(x) + f_i(y)]_{j(X+Y)} \\
&= \frac{1}{m - p} \left( \sum_{j=p+1}^m f_{i_{j(X+Y)}}(x) + \sum_{j=p+1}^m f_{i_{j(X+Y)}}(y) \right) \\
&\leq \frac{1}{m - p} \sum_{j=p+1}^m f_{i_{j(X)}}(x) + \frac{1}{m - p} \sum_{j=p+1}^m f_{i_{j(Y)}}(y) \\
&= C\text{-VaR}_\beta(X) + C\text{-VaR}_\beta(Y)
\end{aligned}$$

Uma análise mais crítica nos leva a pensar “O que significa  $X, Y, X + Y \in V \Rightarrow \rho(X + Y) \leq \rho(X) + \rho(Y)$ ”. “Não deveria ser  $x, y, x + y \in \Omega \Rightarrow \rho(x + y) \leq \rho(x) + \rho(y)$ ?”. Quando estamos pensando em quebrar uma decisão em duas e efetuá-la separadamente estamos trabalhando em  $\Omega$  não em  $V$ . O exemplo a seguir mostra que, no caso geral, não é equivalente trabalhar com  $\Omega$  ou  $V$ .

Suponhamos que nosso conjunto de decisões possíveis  $\Omega = \mathbb{R}^2$ , e que existam 5 cenários possíveis para o futuro econômico. A função de perda para cada um deles é:

$$\begin{aligned}
f_1(x) = f_2(x) = f_3(x) = f_4(x) &= 0, \\
f_5(x) &= 4x_1^2 + 4x_2^2 - 3
\end{aligned}$$

Consideremos agora as seguintes decisões:

$$x = (1, 0) \text{ associada a } X = (0, 0, 0, 0, 1)$$

$$y = (0, 1) \text{ associada a } Y = (0, 0, 0, 0, 1)$$

A variável  $Z = X + Y = (0, 0, 0, 0, 2) \in V$  pois  $Z$  está associada à decisão  $z = (\sqrt{5/8}, \sqrt{5/8})$ .

Podemos ver que:

$$C\text{-VaR}_{80\%}(X + Y) = 2 \leq 1 + 1 = C\text{-VaR}_{80\%}(X) + C\text{-VaR}_{80\%}(Y)$$

Este resultado era esperado, uma vez já foi provado (Teorema 3.3.3) que a  $C\text{-VaR}$  é sub-aditiva.

Porém, o que acontece com a decisão  $w = x + y = (1, 1)$ ?

A variável aleatória  $W$  associada a  $w$  é  $W = (0, 0, 0, 0, 5)$ .

$$C\text{-VaR}_{80\%}(W) = 5 > 1 + 1 = C\text{-VaR}_{80\%}(X) + C\text{-VaR}_{80\%}(Y)!!!$$

Nesse exemplo vemos que podemos quebrar uma decisão em duas e assim obter um risco menor, portanto chegamos a conclusão que o intuito do axioma da sub-aditividade não foi atingido!

Por que isso aconteceu?

A culpa de tal paradoxo foi de  $f_5(x)$  uma vez que em todas decisões consideradas no exemplo tínhamos que  $C\text{-VaR}_{80\%}(x) = f_5(x)$ . A característica fundamental de  $f_5(x)$  é o fato não ser sub-aditiva, ou seja existem  $x$  e  $y \in \Omega$  tais que  $f_5(x + y) > f_5(x) + f_5(y)$ . Acreditamos que um problema desse tipo esteja mal modelado uma vez que, para qualquer que seja o cenário futuro, é difícil de aceitar que a perda em relação a um investimento pode diminuir se o dividirmos e realizarmos separadamente. Acreditamos também que essa característica, assim como várias outras, das funções de perda sejam conhecidas dos economistas.

Desse modo diremos que um problema está bem modelado somente se  $f_i(x)$  for sub-aditiva para toda função de perda  $f_i$  envolvida na modelagem. Afirmamos que só tem sentido falar em medida de risco coerente para problemas bem modelados. Para essa classe de problemas vamos redefinir que  $\rho : \Omega \rightarrow \mathbb{R}$  é uma medida de risco coerente se

cumpra os axiomas abaixo:

1. monotonicidade:  $x, y \in \Omega, X \geq Y \Rightarrow \rho(x) \geq \rho(y)$
2. invariância por translação:  $a \in \mathbb{R}, x, y \in \Omega, Y = X + a \Rightarrow \rho(y) = \rho(x) + a$
3. positividade homogênea:  $h > 0, x, y \in \Omega, Y = hX \Rightarrow \rho(y) = h\rho(x)$
4. sub-aditividade:  $x, y, x + y \in \Omega \Rightarrow \rho(x + y) \leq \rho(x) + \rho(y)$

É importante notar que quando cada  $f_i$  é linear, e se  $X + Y \in V$  implicar  $x + y \in \Omega$  a definição acima coincide com a anterior (3.3.1). Portanto, pelo exemplo 3.3.2, a VaR ainda é não coerente nesse novo contexto.

No caso geral temos que 1, 2, 3 são equivalentes a *i)* *ii)* e *iii)* uma vez que  $X \in V \Leftrightarrow \exists x \in \Omega$  tal que a variável aleatória associada a  $x$  é  $X$ . Portanto a C-VaR cumpre tais axiomas.

Provemos que com a hipótese do problema estar bem definido a C-VaR também cumpre 4 e conseqüentemente é uma medida de risco coerente no novo sentido apresentado aqui.

### **Demonstração:**

Sejam  $x, y \in \Omega$ , então pelo fato do problema estar bem definido temos que  $f_i(x + y) \leq$

$f_i(x) + f_i(y) \forall i$ . Sendo assim:

$$\begin{aligned}
 C\text{-VaR}_\beta(x+y) &= \frac{1}{m-p} \sum_{j=p+1}^m f_{i_j(x+y)}(x+y) \\
 &\leq \frac{1}{m-p} \sum_{j=p+1}^m [f_{i_j(x+y)}(x) + f_{i_j(x+y)}(y)] \\
 &= \frac{1}{m-p} \sum_{j=p+1}^m f_{i_j(x+y)}(x) + \frac{1}{m-p} \sum_{j=p+1}^m f_{i_j(x+y)}(y) \\
 &\leq \frac{1}{m-p} \sum_{j=p+1}^m f_{i_j(x)}(x) + \frac{1}{m-p} \sum_{j=p+1}^m f_{i_j(y)}(y) \\
 &= C\text{-VaR}_\beta(x) + C\text{-VaR}_\beta(y)
 \end{aligned}$$

Como comentário final gostaríamos de ressaltar que a sub-aditividade junto com a positividade homogênea implicam uma propriedade mais forte que a convexidade, a convexidade de epígrafe cônica. Entretanto a sub-aditividade por si só não tem relação com convexidade. A função  $f = x - 10$  mostra que convexidade não implica sub-aditividade. Já a função  $g(x) = -\exp(x)$  definida em  $[\ln(2), +\infty)$  é sub-aditiva. De fato,  $\exp(x) \geq 2 \Rightarrow \exp(x+y) = \exp(x) \cdot \exp(y) \geq 2 \cdot \max\{\exp(x), \exp(y)\} \geq \exp(x) + \exp(y) \Rightarrow g(x+y) \leq g(x) + g(y)$ . Porém  $g$  não é convexa pois sua segunda derivada é menor do que zero. Um exemplo que mostra a dificuldade de se fazer grandes interpretações geométricas da sub-aditividade é o fato de que uma translação de uma função sub-aditiva não é necessariamente sub-aditiva. Podemos ver isso tomando a função  $f$  e a função  $h(x) = x$ , claramente  $f$  não é sub-aditiva mas  $h$  é.

---

### 3.4 Outras propriedades da VaR e da C-VaR

---

Nessa seção vamos expor mais algumas propriedades da VaR e da C-VaR. Um conjunto dessas características são puramente matemáticas, outro tem um apelo econômico, entretanto acreditamos que algumas destas idéias podem ser questionáveis do ponto de vista prático.

A primeira propriedade que apresentaremos aqui é que se as funções de perda são convexas então a C-VaR também é.

**Demonstração:**

$$\begin{aligned}
C\text{-VaR}_\beta(\alpha x + (1 - \alpha)y) &= \frac{1}{m - p} \sum_{j=p+1}^m f_{i_j(\alpha x + (1 - \alpha)y)}(\alpha x + (1 - \alpha)y) \\
&\leq \alpha \frac{1}{m - p} \sum_{j=p+1}^m f_{i_j(\alpha x + (1 - \alpha)y)}(x) + \\
&\quad (1 - \alpha) \frac{1}{m - p} \sum_{j=p+1}^m f_{i_j(\alpha x + (1 - \alpha)y)}(y) \\
&\leq \alpha \frac{1}{m - p} \sum_{j=p+1}^m f_{i_j(x)}(x) + \\
&\quad (1 - \alpha) \frac{1}{m - p} \sum_{j=p+1}^m f_{i_j(y)}(y) \\
&= \alpha C\text{-VaR}_\beta(x) + (1 - \alpha) C\text{-VaR}_\beta(y)
\end{aligned}$$

Já a VaR pode não ser convexa nem no caso clássico em que as funções de perda são lineares.

**Exemplo:**

Suponhamos que as funções de perda nos três cenários possíveis sejam as funções lineares abaixo:

$$f_1(x) = 0x_1 + 6x_2$$

$$f_2(x) = 1x_1 + 1x_2$$

$$f_3(x) = 4x_1 + 0x_2$$

Logo:

$$VaR_{66\%}(1, 0) = 1, VaR_{66\%}(0, 1) = 1 \text{ e } VaR_{66\%}(0.5, 0.5) = 2$$

conseqüentemente,

$$VaR_{66\%}(0.5, 0.5) > 0.5 \cdot VaR_{66\%}(1, 0) + 0.5 \cdot VaR_{66\%}(0, 1)$$

Confirmamos então que a VaR pode não ser convexa nem no caso linear.

A convexidade é de suma importância pois garante que para a C-VaR todo minimizador local também é minimizador global, quando  $\Omega$  é convexo.

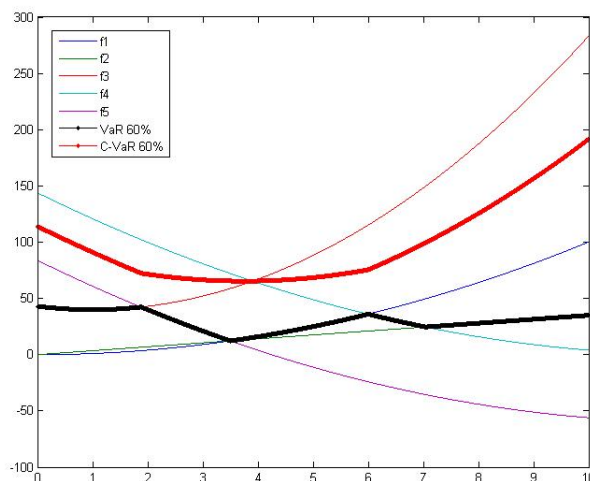


Figura 3.6: Nessa figura podemos ver que a  $VaR_{60\%}$  possui dois minimizadores locais não globais (1 e 7.0375), já a  $C-VaR_{60\%}$  é convexa e seu minimizador global é 3.75.

Essas características, junto com o fato de que o problema de minimizar a C-VaR pode ser transformado num PL (veremos como mais adiante), tornam a C-VaR muito mais viável de ser otimizada do que a VaR. Como a  $C-VaR_{\beta}(x) \geq VaR_{\beta}(x)$ , uma vez que é a média das funções acima da VaR, temos que diminuindo a C-VaR diminuimos um limitante para o valor mínimo de  $VaR(x)$ . Além disso, vários autores (por exemplo [2]) afirmam que experimentos numéricos mostram que em situações reais o portfólio ótimo em relação a minimizar a VaR não está longe do relativo à minimização da C-VaR.

Outra propriedade da C-VaR é não menosprezar cenários catastróficos. Por exemplo, suponhamos que em um problema com 10 cenários futuros existam duas decisões  $x$  e  $y$  tais que  $f_i(x) = f_i(y) = -1$  para  $i = 1, \dots, 8$ ;  $f_9(x) = f_{10}(x) = 0$  e  $f_9(y) = f_{10}(y) = 10^6$ . Nesse caso temos que  $VaR_{80\%}(x) = VaR_{80\%}(y) = -1$ . Porém, é justo que essas decisões tenham o mesmo risco? A C-VaR mediria essa diferença. Essa característica, embora alarmante, não nos preocupa muito uma vez que não acreditamos que em situações práticas existam decisões com tal comportamento. Além disso o que aconteceria se avaliarmos a  $VaR_{90\%}$ ? Para esse percentil teríamos uma idéia clara da diferença entre as decisões. De fato é comum em situações concretas avaliar a VaR para vários valores de  $\beta$  ou colocar como restrição do problema que nossas decisões não apresentem  $VaR_{\beta_i}$  que ultrapassem certas constantes  $M_i$  para alguns valores de  $\beta_i$  e  $M_i$  diferentes. Para alguns investidores essa é inclusive uma desvantagem da C-VaR. Eles acreditam que a C-VaR é muito pessimista e decisões tomadas sob sua influência são demasiadamente conservadoras, diminuindo assim o risco mas limitando muito a chance de altos rendimentos.

---

### 3.5 Otimização da VaR e da C-VaR

---

Nessa seção vamos expor sobre a reformulação clássica, devida a Rockafeller e Uryasev [14], do problema de minimizar a C-VaR. Em seguida vamos apresentar a idéia de Andreani, Dunder e Martínez [6] para minimizar a VaR, posteriormente vamos falar da iniciativa de Alexander, Coleman e Li [2] de aproximar o problema de minimizar a C-VaR por outro de menor porte e por fim juntaremos essas duas últimas idéias propondo uma nova maneira de trabalhar com o problema de minimizar a VaR.

Faremos agora um apanhado das idéias expostas em [14], adaptando-as ao nosso caso particular de cenários discretos. Seguindo o espírito das seções anteriores temos que o problema de minimizar a C-VaR é muito mais fácil do que o problema de minimizar a VaR. Porém um detalhe muito importante não foi dito até aqui. Na definição da C-VaR está implícita a expressão da VaR, logo as dificuldades de manipulação da VaR são automaticamente passadas a C-VaR. Felizmente conseguimos contornar esse problema com as observações abaixo e com a ajuda do teorema chave 3.5.1.



Podemos ver que:

$$\begin{aligned} C\text{-VaR}_\beta(x) &= \frac{1}{m-p} \sum_{j=p+1}^m f_{i_{j(x)}}(x) \\ &= \text{VaR}_\beta(x) + \frac{1}{m-p} \sum_{i=1}^m (f_i(x) - \text{VaR}_\beta(x))^+, \end{aligned}$$

onde  $(x)^+ = \max\{0, x\}$ .

Pensando nisso podemos definir  $F(x, t)$  da seguinte maneira:

$$F(x, t) \equiv t + \frac{1}{m-p} \sum_{i=1}^m (f_i(x) - t)^+.$$

Temos então que a  $C\text{-VaR}_\beta(x) = F(x, \text{VaR}_\beta(x))$ . Muitas vezes vamos preferir trabalhar com  $F$  em vez da C-VaR. A razão disso vem do teorema a seguir:

**Teorema 3.5.1. Teorema chave**

Sejam  $f_1 \cdots f_m$  números reais tais que  $f_1 \leq \cdots \leq f_m$  e  $p \in \{1 \cdots m\}$ . Definamos  $\bar{F}(t) = (m-p)t + \sum_{i=1}^m (f_i - t)^+$ .

Dessa forma temos que, se  $p < m$ , o conjunto dos minimizadores de  $\bar{F}(t)$  é o intervalo  $[f_p, f_{p+1}]$  com valor mínimo de  $\bar{F}$  igual a  $\sum_{i=p+1}^m f_i$ . Se  $p = m$  o conjunto de minimizadores é  $[f_m, \infty)$  com valor mínimo igual a zero.

**Demonstração:**

Faremos apenas o caso  $p < m$ . O caso  $p = m$  é análogo e trivial. Primeiramente vamos achar a expressão para  $\bar{F}(t)$

- $t < f_1 \Rightarrow (f_i - t)^+ = f_i - t \quad \forall i$

$$\bar{F}(t) = (m-p)t + \left( \sum_{i=1}^m f_i \right) - mt = -pt + \sum_{i=1}^m f_i$$

$$\bullet f_k \leq t \leq f_{k+1} \Rightarrow (f_i - t)^+ = \begin{cases} 0, & \text{se } i < k + 1 \\ f_i - t & \text{se } i \geq k + 1 \end{cases}$$

$$\bar{F}(t) = (m - p)t + \left( \sum_{i=k+1}^m f_i \right) - (m - k)t = (k - p)t + \sum_{i=k+1}^m f_i$$

$$\bullet t > f_m \Rightarrow (f_i - t)^+ = 0 \quad \forall i$$

$$\bar{F}(t) = (m - p)t$$

Substituindo  $k$  por  $p$  vemos que se  $t \in [f_p, f_{p+1}]$  então  $\bar{F}(t) = \sum_{i=p+1}^m f_i$ . Provemos agora

que se  $t \notin [f_p, f_{p+1}]$  então  $\bar{F}(t) > \sum_{i=p+1}^m f_i$

$$\bullet t < f_p$$

$$\text{a) } f_k \leq t < f_{k+1} \Rightarrow (f_j - t) > 0 \quad \forall j = (k + 1) \cdots p$$

$$\begin{aligned} \bar{F}(t) &= (k - p)t + \sum_{i=k+1}^m f_i \\ &= (f_{k+1} - t) + \cdots + (f_p - t) + \sum_{i=p+1}^m f_i \\ &> \sum_{i=p+1}^m f_i \end{aligned}$$

$$\text{b) } t < f_1 \Rightarrow (f_j - t) > 0 \quad \forall j = 1 \cdots p$$

$$\begin{aligned} \bar{F}(t) &= -pt + \sum_{i=1}^m f_i \\ &= (f_1 - t) + \cdots + (f_p - t) + \sum_{i=p+1}^m f_i \\ &> \sum_{i=p+1}^m f_i \end{aligned}$$

$$\bullet t > f_{p+1}$$

$$\text{a) } f_k < t \leq f_{k+1} \Rightarrow (f_j - t) < 0 \quad \forall j = (p+1) \cdots k$$

$$\begin{aligned} \bar{F}(t) &= (k-p)t + \sum_{i=k+1}^m f_i \\ &= -(f_{p+1} - t) - \cdots - (f_k - t) + \sum_{i=p+1}^m f_i \\ &> \sum_{i=p+1}^m f_i \end{aligned}$$

$$\text{b) } t > f_m \Rightarrow (f_j - t) < 0 \quad \forall j = (p+1) \cdots m$$

$$\begin{aligned} \bar{F}(t) &= (m-p)t \\ &= -(f_{p+1} - t) - \cdots - (f_m - t) + \sum_{i=p+1}^m f_i \\ &> \sum_{i=p+1}^m f_i \end{aligned}$$

Podemos notar também que  $F(x, t)$ , como função de  $t$ , é convexa, uma vez que a soma de e o máximo entre funções convexas é uma função convexa, e é linear por partes. O coeficiente linear de cada parte aumenta, em módulo, em uma unidade ao distanciarmos do intervalo  $[f_p, f_{p+1}]$ .

A figura abaixo ilustra esses fatos e o teorema anterior.

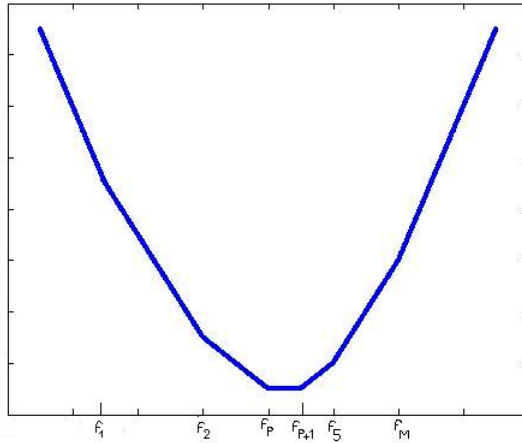


Figura 3.7: Gráfico de  $F(x, t)$  em função de  $t$ .

Claramente se cada  $f_i$  é convexa então  $F(x, t)$  também o é com respeito a  $(x, t)$  pelo mesmo argumento anterior.

Com a ajuda desse teorema temos que  $C\text{-VaR}(x) = \text{mínimo } F(x, t)$  em relação a  $t$ , logo o problema de minimizar a  $C\text{-VaR}$  passa a ser:

$$\underset{x \in \Omega}{\text{minimizar}} \quad \underset{t \in \mathbb{R}}{\text{mínimo}} \quad t + \frac{1}{m-p} \sum_{i=1}^m (f_i(x) - t)^+$$

Ou, equivalentemente:

$$\underset{x \in \Omega; t \in \mathbb{R}}{\text{minimizar}} \quad t + \frac{1}{m-p} \sum_{i=1}^m (f_i(x) - t)^+ \quad (3.1)$$

Essa maneira de minimizar a  $C\text{-VaR}$  é um problema convexo no caso em que as funções de perda são convexas, em particular quando são dadas da maneira clássica (linearmente), porém não diferenciável já que envolve o máximo entre duas funções.

Apresentaremos aqui duas idéias para contornar esse inconveniente. A primeira é uma reformulação, também proposta em [14], que consiste em acrescentar variáveis  $z$  restringindo-as de modo a satisfazer  $z_i = (f_i(x) - t)^+$  na solução e nos leva a:

$$\begin{aligned}
& \underset{x, t, z}{\text{minimizar}} && t + \frac{1}{m-p} \sum_{i=1}^m z_i \\
& \text{S. a} && x \in \Omega \\
& && z_i \geq 0 \quad \forall i = 1 \dots m \\
& && z_i \geq f_i(x) - t \quad \forall i = 1 \dots m
\end{aligned} \tag{3.2}$$

Se as funções  $f_i$  forem lineares e  $\Omega$  for um politopo, o que é o caso nos problemas anteriores, temos que o problema acima é um *PL*.

Pelo teorema anterior o trio  $(x^*, t^*, z^*)$  solução de (3.2) é tal que  $x^*$  é o minimizador da  $C\text{-VaR}_\beta(x)$ ,  $t^* + \frac{1}{m-p} \sum_{i=1}^m z_i^*$  é o valor mínimo da  $C\text{-VaR}_\beta(x)$  e  $t^*$  é uma aproximação para a  $VaR_\beta$  associada à decisão  $x^*$ , uma vez que  $t^* \in [f_{i_p}(x^*), f_{i_{p+1}}(x^*)]$  e para um grande número de cenários é natural que  $f_{i_p}(x^*)$  esteja próximo de  $f_{i_{p+1}}(x^*)$ , ocorrendo muitas vezes até a igualdade. Portanto, sem perda de generalidade, vamos considerar em muitos casos  $t^* = VaR_\beta(x^*)$ . É importante notar que  $x^*$  não é necessariamente solução do problema de minimizar  $VaR_\beta(x)$ .

A modelagem acima apresenta o mesmo problema do modelo de Konno. Novamente acrescentamos uma variável para cada cenário previsto. Para evitar isso, e ainda assim deixar o problema diferenciável temos a segunda idéia. Em [2] foi proposta por Alexander, Coleman e Li a definição de:

$$\phi_\epsilon(x) = \begin{cases} x; & \text{se } x \geq \epsilon \\ \frac{x^2}{4\epsilon} + \frac{x}{2} + \frac{\epsilon}{4}; & \text{se } -\epsilon < x < \epsilon \\ 0; & \text{se } x \leq -\epsilon \end{cases}$$

e a seguinte reformulação de 3.1:

$$\begin{aligned}
& \underset{x, t}{\text{Min}} && t + \frac{1}{m-p} \sum_{i=1}^m \phi_\epsilon(f_j(x) - t) \\
& \text{S. a} && x \in \Omega
\end{aligned} \tag{3.3}$$

A intenção da introdução da função  $\phi_\epsilon$  é suavizar o “bico” da função máximo mantendo a maior semelhança possível com a função original.

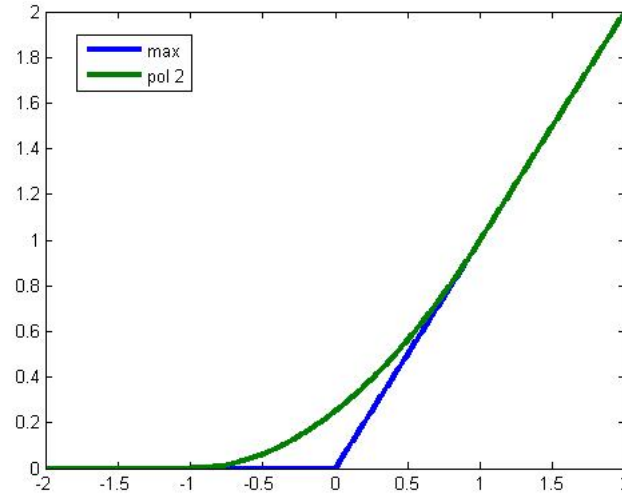


Figura 3.8: Gráfico da suavização para  $\epsilon = 1$ .

Podemos perceber que  $\phi_\epsilon(x)$  difere do  $\max\{0, x\}$  somente no intervalo  $(-\epsilon, \epsilon)$  e o módulo dessa diferença é limitada por  $\frac{\epsilon}{4}$ . Isso nos permite concluir que  $\phi_\epsilon(x)$  converge uniformemente para  $\max\{0, x\}$  quando  $\epsilon$  tende a zero. O teorema a seguir garante algumas propriedades teóricas para o problema aproximado.

**Teorema 3.5.2.** *Seja  $\{f_1(x), f_2(x), \dots\}$  uma seqüência de funções de  $\Omega$  em  $\mathbb{R}$  uniformemente convergente para uma função contínua  $f : \Omega \rightarrow \mathbb{R}$ . Suponhamos que cada  $f_i$  tenha um minimizador global  $m_i$ . Logo todo ponto de aderência de  $\{m_1, m_2, \dots\}$  é um minimizador global de  $f$ .*

**Demonstração:**

Seja  $m$  um ponto de aderência de  $\{m_1, m_2, \dots\}$ . Suponhamos por absurdo que exista  $z \in \Omega$  tal que  $f(z) < f(m)$ .

Tomemos  $\epsilon = \frac{f(m) - f(z)}{3}$ .

Como  $f$  é contínua existe  $\delta > 0$  tal que  $f(x) \in (f(m) - \epsilon, f(m) + \epsilon)$ ,  $\forall x \in B(m, \delta) \cap \Omega$ .

Pela convergência uniforme existe  $N \in \mathbb{N}$  tal que  $\forall n > N$  e  $\forall x \in \Omega$  temos  $|f_n(x) -$

$f(x)| < \epsilon$ . Em particular  $f_n(z) \leq f(z) + \epsilon$  e  $f(m_n) < f_n(m_n) + \epsilon$ .

Conseqüentemente  $f(m_n) \leq f_n(m_n) + \epsilon \leq f_n(z) + \epsilon < f(z) + 2\epsilon \leq f(m) - \epsilon$ .

Portanto  $m_n \notin B(m, \delta) \forall n > N$ . Sendo assim  $m$  não é ponto de aderência de  $\{m_1, m_2, \dots\}$  o que contraria a hipótese. Concluimos assim que  $f(m) \leq f(x) \forall x \in \Omega$  e dessa forma temos que  $m$  é um minimizador global de  $f$  em  $\Omega$ .

Outra propriedade importante de  $\phi_\epsilon$  é que ela preserva a convexidade do problema quando as funções de perda são convexas. Isso vem diretamente do fato de que se  $f$  é crescente e convexa e  $g$  é convexa então  $f(g(x))$  é convexa. É natural supor  $\Omega$  compacto em situações práticas, nesse caso o problema (3.3) tem um minimizador global. Para funções de perda convexas espera-se que encontremos esses minimizadores pois nossos algoritmos são eficazes na busca por minimizadores locais e sob essas hipóteses esses serão globais também. Dessa forma se tivermos uma seqüência  $\epsilon_n$  tendendo a zero então o teorema 3.5.2 acima garante que os minimizadores dos problema 3.3 convergirão para a solução de minimizar a C-VaR. Embora esse seja um resultado forte, em situações práticas escolhemos  $\epsilon$  pequeno e nos contentamos com essa aproximação. Além disso, valores de  $\epsilon$  abaixo de certo patamar deixa o problema numericamente muito próximo de um problema não diferenciável, o que, como veremos nos testes computacionais, dificulta sua resolução.

Acreditamos que a utilidade da suavização vai muito além de competir com um PL, permitindo resolver problemas cujas funções de perda são não lineares, oriundas de uma modelagem qualquer. Validaremos essa idéia usando, em alguns de nossos testes funções de perda não lineares motivadas pela fórmula Black Scholes que será introduzida mais adiante.

Nos testes computacionais usaremos também outras suavizações possíveis para  $(x)^+$  pois acreditamos que a exposta em [2] não seja tão confiável uma vez que  $\phi$  não é duas vezes diferenciável, o que pode ser um empecilho para métodos do tipo Newton ou Quasi-Newton.

Uma alternativa para isso seria introduzir um polinômio no intervalo  $[-\epsilon, \epsilon]$  com grau

suficientemente grande de modo que ao menos valha:

- $\phi(-\epsilon) = (-\epsilon)^+$
- $\phi(\epsilon) = (\epsilon)^+$
- $\phi'(-\epsilon) = (-\epsilon)^{+'}$
- $\phi'(\epsilon) = (\epsilon)^{+'}$
- $\phi''(-\epsilon) = (-\epsilon)^{+''}$  e
- $\phi''(\epsilon) = (\epsilon)^{+''}$ .

Um exemplo seria tomar

$$\phi_\epsilon(x) = \begin{cases} x; & \text{se } x \geq \epsilon \\ -\frac{x^4}{16\epsilon^3} + \frac{3x^2}{8\epsilon} + \frac{x}{2} + \frac{3\epsilon}{16}; & \text{se } -\epsilon < x < \epsilon \\ 0; & \text{se } x \leq -\epsilon \end{cases} \quad (3.4)$$

Para esse tipo de suavização temos que  $\phi(x)$  difere de  $(x)^+$  apenas no intervalo  $[-\epsilon, \epsilon]$ , em particular, no caso da função acima, o máximo dessa diferença vale  $3\epsilon/16$  e é atingido em  $x = 0$ . Porém, essa classe de adaptações tem um número finito de derivadas.

Pensando nisso procuramos definir  $\phi$  de modo a obtermos uma função  $C^\infty$ . Sabemos que

$$(x)^+ = \frac{x + |x|}{2} = \frac{x + \sqrt{x^2}}{2},$$

como  $\sqrt{x}$  só não é diferenciável em  $x = 0$  podemos obter a aproximação

$$(x)^+ \approx \phi_\epsilon(x) = \frac{x + \sqrt{x^2 + \epsilon}}{2}.$$

Conseguimos assim uma função  $C^\infty$  mas que difere da função original em toda a reta. É fácil ver que nesse caso  $\phi_\epsilon(x) \geq (x)^+$ , logo o erro da aproximação seria  $E(x) = \phi_\epsilon(x) - (x)^+$ . Mais ainda,



$$E'(x) = \begin{cases} \frac{1}{2} \left( 1 + \frac{x}{\sqrt{x^2+\epsilon}} \right) > 0; & \text{se } x < 0 \\ \frac{1}{2} \left( \frac{x}{\sqrt{x^2+\epsilon}} - 1 \right) < 0; & \text{se } x > 0 \end{cases},$$

logo  $E(x)$  é crescente para  $x < 0$  e decrescente para  $x > 0$ , portanto, seu máximo é atingido em  $x = 0$  e vale  $\sqrt{\epsilon}/2$ . Sendo assim, todas as suavizações apresentadas convergem uniformemente para  $(x)^+$ .

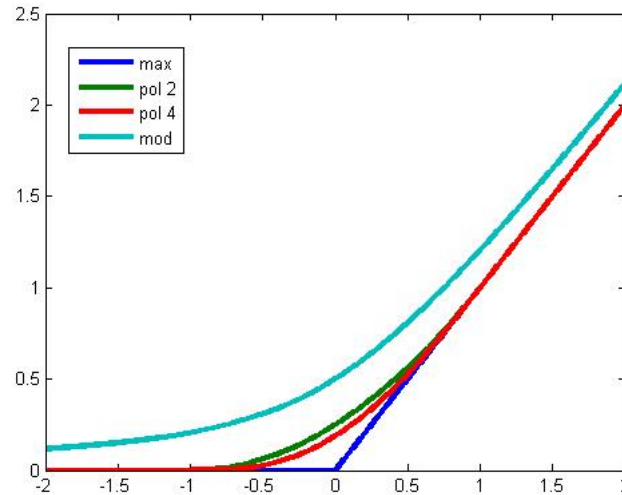


Figura 3.9: Gráfico das suavizações para  $\epsilon = 1$ .

Seguindo ainda espírito do Teorema 3.5.1 lembramos que:

$$\underset{t}{\operatorname{argmin}} \quad F(x, t) = [f_{i_p(x)}, f_{i_{p+1}(x)}],$$

logo podemos modelar o problema de minimizar  $VaR_\beta(x)$  da seguinte forma:

Minimizar  $t$   
 $x, t$

Sujeito a  $x \in \Omega$

$t$  resolve o Problema:

$$\text{Minimizar}_t \quad t + \frac{1}{m-p} \sum_{i=1}^m (f_i - t)^+$$

No caso em que acrescentamos a variável  $z$  para evitarmos a não diferenciabilidade temos:

Min  $t$

S. a  $x \in \Omega$

$t$  resolve o Problema:

$$\text{Min}_{t,z} \quad t + \frac{1}{m-p} \sum_{i=1}^m z_i \tag{3.5}$$

$$\text{S. a} \quad z_i \geq 0 \quad \forall i = 1 \cdots m$$

$$z_i \geq f_i - t \quad \forall i = 1 \cdots m$$

A reformulação acima trata-se de um problema de dois níveis. Podemos substituir o problema de baixo por suas condições de otimalidade. No caso em que funções de perda e  $\Omega$  são convexas essas condições são necessárias e suficientes. Porém, quando fazemos isso, caímos num problema de programação não linear (provavelmente de grande porte pois estamos acrescentando uma variável a cada cenário) mesmo quando tratamos do problema clássico com restrições lineares.

Colocando as condições KKT no problema do segundo nível temos:

$$\begin{array}{ll} \min_{x,z,t,\lambda,\alpha} & t \\ S.a & x \in \Omega \\ & (z_i + t - f_i(x))\alpha_i = 0 \\ & z_i\lambda_i = 0 \\ & 1 = \lambda_i + \alpha_i \\ & m - p = \sum_{i=1}^m \alpha_i \\ & \lambda_i \geq 0 \\ & \alpha_i \geq 0 \\ & z_i \geq 0 \\ & z_i + t - f_i(x) \geq 0 \end{array}$$

Usando as relações  $\alpha_i = 1 - \lambda_i$ , denotando  $u_i = z_i - t - f_i(x)$  e com algumas manipulações algébricas chegamos exatamente na formulação proposta por Andreani, Dunder e Martínez em [6].

$$\begin{aligned}
& \min_{x,z,t,u,\alpha} && t \\
& S.a && x \in \Omega \\
& && u_i = z_i - t - f_i(x) \\
& && \sum_{i=1}^m u_i(1 - \lambda_i) = 0 \\
& && \sum_{i=1}^m z_i \lambda_i = 0 \\
& && \sum_{i=1}^m \lambda_i = p \\
& && 0 \leq \lambda_i \leq 1 \\
& && \alpha_i \geq 0 \\
& && z_i \geq 0 \\
& && u_i \geq 0
\end{aligned}$$

Nesse artigo os autores conseguem resultados ainda mais fortes do que os encontrados em [7] com respeito à qualidade da solução do problema.

Uma alternativa natural seria colocar uma suavização na resolução do problema do segundo nível.

Min  $t$

S. a  $x \in \Omega$

$t$  resolve o Problema:

$$\underset{t}{Min} \quad t + \frac{1}{m-p} \sum_{i=1}^m \phi_{\epsilon}(f_i - t)$$

Substituindo as condições de otimalidade:

$$\begin{aligned}
\min_{x,t} \quad & t \\
S.a \quad & x \in \Omega \\
& (m-p) - \sum_{j=1}^m \phi'_\epsilon(f_j(x) - t) = 0
\end{aligned}$$

Essa aproximação nos parece ainda mais vantajosa aqui do que no caso da minimização da C-VaR. Isso porque no problema da C-VaR trocamos um PL de grande porte por um PNL menor. Já aqui de toda forma seríamos obrigados a trabalhar com um problema de programação não linear, pois as condições KKT de um PL assim são. Portanto, é de se esperar que o problema de menor porte seja mais eficiente.

Entretanto, a questão das suavizações aqui é extremamente delicada. A razão da preocupação vem do fato da condição de otimalidade do problema 3.1 ser  $t \in [f_{i_p(x)}, f_{i_{p+1}(x)}]$ . Como  $(x)^+$  só não é diferenciável em  $x = 0$ , um abuso de notação nos levaria a restrição

$$\sum_{j=1}^m (f_j(x) - t)^{+'} = m - p,$$

onde  $(x)^{+'} = 0$  se  $x < 0$  e  $(x)^{+'} = 1$  se  $x > 0$ . Ou seja, a restrição conta quantas funções de perda estão acima de  $t$  e exige que esse valor seja exatamente  $(m - p)$ , conseqüente exige-se que  $t \in [f_{i_p(x)}, f_{i_{p+1}(x)}]$ . Dessa forma estamos trabalhando com uma restrição descontínua, e mais ainda, binária. Vejamos como as suavizações propostas aproximam tal restrição.

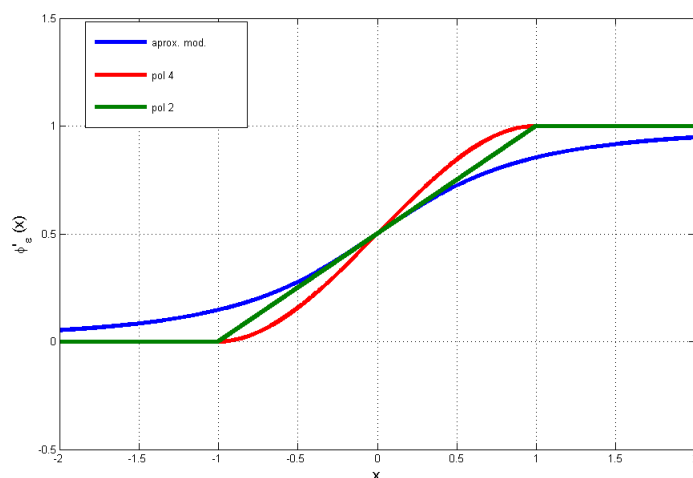


Figura 3.10: Gráfico das derivadas das suavizações para  $\epsilon = 1$ .

Mais uma vez a suavização proposta em [2] não parece muito adequada pois teríamos uma restrição não diferenciável, o que deixaria o problema ainda não suave. Já a aproximação  $\phi \approx (x + \sqrt{x^2 + \epsilon})/2$ , mostra-se muito bem comportada, provavelmente por se tratar de uma função  $C^\infty$ . Isso deve ser refletido nos resultados computacionais com algoritmos mais robustos.

---

---

# CAPÍTULO 4

---

## Noções de probabilidade e simulação de cenários futuros

---

### 4.1 Introdução

---

Nesse capítulo vamos expor algumas maneiras de prever possíveis cenários para o futuro da economia. Basicamente as alternativas para construirmos esses cenários podem ser divididas em três grupos.

O primeiro é totalmente subjetivo onde o responsável pela previsão dos cenários utiliza sua própria experiência e intuição, além de informações dadas por outras pessoas nas quais se tem confiança, para construir um panorama sobre o futuro econômico.

Esse tipo de abordagem é muito difícil de ser tratado matematicamente além de exigir um amplo conhecimento específico de cada ativo envolvido. Essas características fazem com que esse pensamento seja deixado de lado em nossa exposição.

Um exemplo dessa filosofia é apresentado a seguir:  
Suponhamos que um grande comerciante de madeira do norte do Paraná acredite, baseado nos seus trinta anos de envolvimento no ramo, que o único fator que influenciará

o preço do seu produto nesse semestre será o crescimento da produção de pinheiros na Finlândia. Um amigo escandinavo lhe garante que esse crescimento será de 5%. Sendo assim ele prevê que o único preço possível para a madeira daqui a seis meses é 30 reais por metro cúbico.

O segundo grupo consiste em fazer uma formulação estritamente matemática do assunto. Essas formulações são feitas com os dados históricos dos preços dos ativos e geralmente usam um modelo probabilístico para prever o futuro. Voltaremos nossas atenções para esse grupo. Mais adiante apresentaremos algumas distribuições de probabilidade que nos serão úteis e as características principais de cada uma delas para que possamos usá-las em nossas previsões.

O terceiro grupo é na verdade um misto dos dois anteriores onde um modelo matemático é criado de acordo com os dados do passado porém seus parâmetros são escolhidos levando em consideração a expectativa do responsável pela previsão. Esse tipo de modelo é o mais usado nas instituições financeiras, entretanto a parte subjetiva foge do escopo deste trabalho.

---

## 4.2 Simulações probalísticas

---

Nessa seção vamos fazer algumas definições básicas para usá-las ao longo deste capítulo. Inicialmente definimos que  $X$  é uma variável aleatória se é uma função real definida no espaço amostral e está associada a uma probabilidade. Uma variável aleatória é dita discreta quando assume um número finito de valores  $\{x_1 \cdots x_n\}$ , a ela está relacionado um conjunto

$$\{P[X = x_1], \cdots, P[X = x_n]\},$$

chamado conjunto de distribuição de probabilidade de  $X$ . Aqui  $P[X = x_i]$  denota a probabilidade de  $X$  ser igual a  $x_i$ . Já a uma variável aleatória contínua costuma-se associar uma função de densidade de probabilidade ( $f$ ). Essa função determina que a área abaixo do gráfico de  $f$  entre  $a$  e  $b$  é igual à probabilidade de  $X$  assumir valores nesse intervalo. Portanto  $f$  deve ser não negativa e  $\int_{-\infty}^{\infty} f(t)dt = 1$ .



A primeira função de densidade que apresentaremos é a resultante de uma variável aleatória com distribuição uniforme.

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{se } x \in [a, b] \\ 0, & \text{caso contrário.} \end{cases}$$

Esse tipo de distribuição pode ser apropriada quando se tem estimativa dos limitantes ( $a$  e  $b$ ) que  $X$  pode atingir mas não do comportamento de  $X$  entre  $a$  e  $b$ . Assume-se, portanto, que a probabilidade é igual para cada valor intermediário.

A esperança de uma variável aleatória com essa distribuição é  $\frac{a+b}{2}$  e o desvio padrão é  $\frac{b-a}{\sqrt{12}}$ .

Vamos usar a distribuição uniforme para simular a presença de resíduos em nossos testes numéricos.

A segunda distribuição que consideraremos será a distribuição normal com parâmetros  $\mu$  e  $\sigma$  cuja notação será dada por  $N(\mu, \sigma^2)$  e função de densidade por:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

A esperança de uma variável aleatória normal vale  $\mu$  e a variância  $\sigma^2$ .

No caso especial em que  $\mu = 0$  e  $\sigma = 1$ , ou seja  $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ , a distribuição chama-se normal padrão.

Como uma função de densidade normal não tem primitiva em termos de funções elementares são de suma importância as tabelas que fornecem a probabilidade de que uma variável aleatória com distribuição normal padrão  $X$  seja menor do que  $x$  para diversos valores de  $x$ . Tais tabelas estão presentes em praticamente qualquer livro introdutório de probabilidade e aproximações numéricas para sua integral estão programadas em diversos softwares computacionais.

É possível mostrar que se  $Z$  tem distribuição normal  $N(\mu, \sigma^2)$  então  $X = \frac{Z-\mu}{\sigma}$  tem distribuição normal padrão.

A grande importância da distribuição normal é dada pelo teorema central do limite e em muitos casos por um de seus corolários imediatos. Enunciaremos tais resultados a seguir:

**Teorema 4.2.1. Teorema Central do Limite de Lindeberg**

Sejam  $X_1, X_2 \dots$  variáveis aleatórias independentes, isto é:

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i),$$

tais que  $E[X_i] = \mu_i$ , variância  $[X_i] = \sigma_i^2$  onde  $0 < \sigma_i^2 < \infty$  e  $f_i(x)$  é função de densidade de  $X_i$ .

Sejam também  $S_n = X_1 + \dots + X_n$  e  $s_n = \sqrt{\sigma^2[S_n]} = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$ .

Então, para que  $\frac{S_n - E[S_n]}{s_n}$  convirja a uma normal padrão quando  $n$  tende a infinito é suficiente que a seguinte condição, chamada condição de Lindeberg, seja satisfeita:

$$\forall \epsilon, \lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \int_{|x - \mu_i| > \epsilon s_n} (x - \mu_i)^2 f_i(x) dx = 0$$

A condição de Lindeberg significa, basicamente, que as parcelas  $\frac{X_k - \mu_k}{s_n}$  da soma  $\frac{S_n - E[S_n]}{s_n}$  são uniformemente pequenas para  $n$  grande. Portanto o teorema central do limite garante que se uma variável aleatória pode ser vista como a soma de várias outras onde o peso de cada parcela é pequeno então essa variável aleatória terá uma distribuição aproximadamente normal. Ou seja, quando um fator depende de vários outros ele deve ter distribuição semelhante a uma normal. Além desse suporte teórico muitas observações práticas confirmam que a distribuição normal de fato está ligada a vários fenômenos de diversas áreas, incluindo a economia.

Enunciaremos agora um de seus colorários imediatos.

**Corolário 4.2.2.** Sejam  $X_1, X_2 \dots X_n$  variáveis aleatórias independentes e identicamente distribuídas (ie. todas  $X_i$  têm a mesma distribuição) com esperança  $\mu$  e desvio padrão  $\sigma$ . Então,  $\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$  converge a uma normal padrão quando  $n$  tende a infinito.

As variáveis  $X_1, X_2 \dots X_n$  podem ser vistas como observações independentes de uma mesma variável aleatória. Nesse caso, o corolário nos diz que, qualquer que seja a variável aleatória, a média de suas observações segue uma distribuição normal.

A terceira distribuição que apresentaremos será a distribuição lognormal a qual se baseia na distribuição normal porém tem algumas características que nos chamam mais atenção. A distribuição lognormal está associada ao *movimento browniano geométrico* que é uma das hipóteses de Black e Scholes em sua famosa fórmula sobre a qual faremos uma pequena exposição no próximo capítulo.

Dizemos que uma variável aleatória positiva  $V$  tem distribuição lognormal com parâmetros  $\mu$  e  $\sigma$  se  $\log(V)$  tem distribuição normal  $N(\mu, \sigma^2)$ .

Suponhamos que  $y(t)$  denote o preço de um ativo no instante de tempo  $t \in [0, \infty)$ , onde  $y(0)$  é o preço do ativo hoje. Por definição, temos que  $y$  segue um *movimento browniano geométrico* com parâmetros de tendência  $\mu$  e de volatilidade  $\sigma$  se para cada  $z$  e  $t$  a variável aleatória

$$\frac{y(t+z)}{y(z)}$$

independe de  $z$  e tem distribuição lognormal com parâmetros  $\mu t$  e  $t\sigma^2$ . É importante notar que  $y$  não é uma função dos reais na reta. A cada  $t$  real  $y(t)$  é uma variável aleatória pois se trata do preço do ativo no instante de tempo futuro  $t$  podendo assumir diversos valores, dependendo do cenário geral da economia nesse instante.

Estamos fazendo hipóteses sobre  $y(t+z)/y(z)$ , ou equivalentemente sobre  $y(t+z)/y(z) - 1$ , e não sobre  $y(t+z) - y(z)$ , para tratarmos com o lucro relativo e não do absoluto. A razão disso é que falar de um aumento de 10% é uma coisa universal, adimensional. Porém, se falarmos que um preço de um ativo aumentou 10 reais com certeza seria propícia a pergunta “e quanto ele custava?”. Tal pergunta é pertinente uma vez que é muito diferente se o preço do ativo era de um real ou de mil reais.

O fato de  $y(t+z)/y(z)$  independe de  $z$  significa que o aumento percentual do preço do ativo não depende da data que estamos falando e sim do tempo que falta para o instante da previsão. Geralmente vamos considerar  $z = 0$  uma vez que estaremos fazendo previsões hoje para o preço do ativo no instante  $t$ .

É de se esperar que a variação do preço de um ativo dependa de vários fatores, como

por exemplo uma nova legislação federal, o desempenho de empresas concorrentes, a valorização do dólar etc, logo, pelo que apresentamos anteriormente, seria natural modelar

$$\frac{y(t+z) - y(z)}{y(z)} = \frac{y(t+z)}{y(z)} - 1$$

com distribuição normal. Como  $\log(1+x) \approx x$  para  $x$  pequeno temos que é razoável que

$$\log\left(1 + \frac{y(t+z) - y(z)}{y(z)}\right) = \log\left(1 + \frac{y(t+z)}{y(z)} - 1\right) = \log\left(\frac{y(t+z)}{y(z)}\right)$$

siga uma distribuição normal. Dessa forma temos que a distribuição lognormal é adequada para modelos econômicos. Ou seja é razoável adotar para previsões financeiras que  $y(t+z) = y(z)e^N$ , em particular  $y(t) = y(0)e^N$ , onde  $N$  é uma variável aleatória com distribuição normal. Conseguimos assim a vantagem de assegurar a positividade de  $y(t)$ , não obtida se supusermos apenas que  $\frac{y(t+z)-y(z)}{y(z)}$  tem distribuição normal.

Os parâmetros da distribuição normal associada ao *movimento browniano geométrico* segue do fato de que se  $y(z+1) = y(z)e^{W_z}$ , onde  $W_z \sim N(\mu_z, \sigma_z)$  então:

$$\begin{aligned} y(1) &= y(0)e^{W_0} \\ y(2) &= y(1)e^{W_1} = y(0)e^{W_0}e^{W_1} = y(0)e^{W_0+W_1} \\ &\vdots \\ y(t) &= y(0)e^W, \text{ onde } W = \sum_{j=0}^{t-1} W_j. \end{aligned}$$

Pela independência em  $z$  de  $y(t+z)/y(z)$  temos que as  $W_z$  são independentes e  $\mu_z = \mu$  além de  $\sigma_z = \sigma$  para todo  $z$ . Portanto  $W \sim N(\mu t, \sigma\sqrt{t})$ , como havíamos afirmado.

Outro resultado teórico de grande importância é a lei dos grandes números apresentada a seguir.

**Teorema 4.2.3. (Lei Forte de Kolmogorov)**

Sejam  $X_1, X_2, \dots$  variáveis aleatórias independentes, identicamente distribuídas (ie. todas  $X_i$  têm a mesma distribuição) e integráveis, com  $E[X_i] = \mu$ . Então

$$\frac{X_1 + \dots + X_n}{n} \rightarrow \mu, \text{ quando } n \rightarrow \infty$$

As variáveis aleatórias  $X_1, X_2 \dots$  podem ser vistas como observações independentes de uma única variável aleatória.

Esses fatos justificam que um histograma normalizado das observações de uma variável aleatória  $X$  aproxima sua função de densidade.

Para ver isso definamos para cada  $a \in \mathbb{R}$  e  $\epsilon > 0$

$$Y_{ei} = \begin{cases} \frac{1}{2\epsilon}, & \text{se } X_i \in [a - \epsilon, a + \epsilon] \\ 0, & \text{caso contrário} \end{cases}$$

Logo

$$E[Y_{ei}] = \frac{1}{2\epsilon} P(X_i \in [a - \epsilon, a + \epsilon]) = \frac{1}{2\epsilon} P(X \in [a - \epsilon, a + \epsilon])$$

A altura do histograma normalizado em  $[a - \epsilon, a + \epsilon]$  com  $n$  observações vale  $\frac{Y_{e1} + \dots + Y_{en}}{n}$ .

Passando o limite quando  $n \rightarrow \infty$  temos pela lei dos grandes números que

$$\frac{Y_{e1} + \dots + Y_{en}}{n} \longrightarrow \frac{1}{2\epsilon} P(X \in [a - \epsilon, a + \epsilon]) = \frac{1}{2\epsilon} \int_{a-\epsilon}^{a+\epsilon} f(x) dx.$$

Fazendo agora o limite quando  $\epsilon \rightarrow 0^+$  e supondo que exista  $F$  primitiva de  $f$ , o que é o caso quando  $f$  é contínua, obtemos que

$$\begin{aligned} \lim_{\epsilon \rightarrow 0^+} \lim_{n \rightarrow \infty} \frac{Y_{e1} + \dots + Y_{en}}{n} &= \lim_{\epsilon \rightarrow 0^+} \frac{1}{2\epsilon} \int_{a-\epsilon}^{a+\epsilon} f(x) dx \\ &= \lim_{\epsilon \rightarrow 0^+} \frac{F(a + \epsilon) - F(a - \epsilon)}{2\epsilon} \\ &= F'(a) = f(a). \end{aligned}$$

O resultado garante que se quisermos ter uma estimativa sobre a função de densidade do comportamento da variação do preço de um ativo podemos fazer um histograma de suas variações para alguns dados históricos.

A seguir vamos analisar se para alguns dados reais o modelo lognormal de fato é consistente. Para isso usamos o preço passado das ações de 5 empresas durante 1001 dias, obtendo assim 1000 variações de preço.

Antes de averiguar a aplicabilidade da distribuição log-normal vamos verificar que a aproximação  $x \approx \log(1 + x)$  é válida para  $x = \frac{y(t+1) - y(t)}{y(t)}$  da ordem da variação do preço

dos ativos em dias consecutivos. Para isso calculamos o erro relativo que cometemos na aproximação em cada observação  $\left(\left|\frac{x - \log(1+x)}{x}\right|\right)$ . A tabela abaixo mostra que realmente a aproximação é razoável para todas as quatro empresas analisadas.

Empresa	média	percentil 90%	máximo
Ambev (PN)	0.0062	0.0129	0.0832
Telemar (PN)	0.0072	0.0160	0.0515
Banco do Brasil (ON)	0.0091	0.0194	0.0698
C. S. N (ON).	0.0098	0.0205	0.0968

Tabela 4.1: A tabela mostra o erro médio, o limitante para 90% dos erros e o erro máximo para cada uma das empresas

A comprovação desse fato não era de fato necessária uma vez que não é uma hipótese direta do nosso modelo. Porém ela é filosoficamente importante pois foi o que motivou tal modelagem. Mais adiante usaremos uma estratégia, baseada em  $\frac{y(t+1)-y(t)}{y(t)}$  para gerar os cenários de nossos testes numéricos. Teremos, então, a garantia de que de certa maneira nosso modelo se assemelha a um *movimento browniano geométrico*.

Para verificar que  $\log\left(\frac{y(t+1)}{y(t)}\right)$  realmente pode ter distribuição normal vamos assumir que vale a independência em  $z$  para  $\frac{y(t+z)}{y(z)}$  suposta no *movimento browniano geométrico*. Sendo assim a variação do preço em cada dia pode ser considerada como uma observação de uma mesma variável aleatória. Na figura 4.1 apresentamos os histogramas e um ajuste normal para quatro grandes empresas dos mais diferentes ramos da economia brasileira. O ajuste é feito tomando os estimadores

$$\bar{\mu} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{e} \quad \bar{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{\mu})^2}{n-1}$$

como os parâmetros da distribuição normal.

A grande semelhança entre os histogramas e as funções de densidade normal parecem atestar pela validade da hipótese lognormal. Por esses motivos o modelo clássico para o comportamento do preço de um ativo é o *movimento browniano geométrico*. Entretanto, para a voraz competição do mercado financeiro atual, tais aproximações muitas vezes já

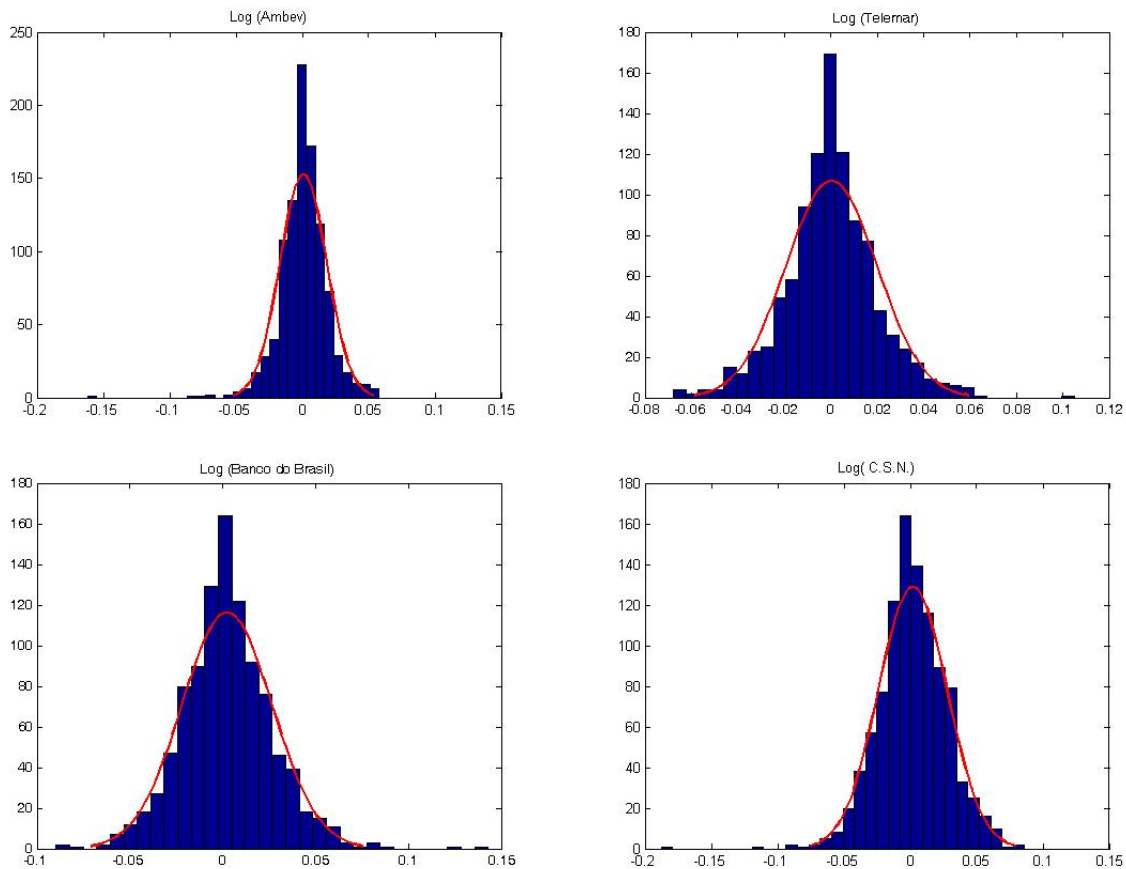


Figura 4.1: Histogramas de  $\log\left(\frac{y(t+1)}{y(t)}\right)$  para Ambev, Telemar, Banco do Brasil e C.S.N.

são consideradas grosseiras e modelos ainda mais realistas já estão sendo usados hoje em dia.

Apesar disso a previsão de cenários dessa forma deixa de lado uma das principais características de um problema de análise de risco, a correlação entre os ativos. Pensando nisso chegamos a nossa alternativa de previsão dos cenários futuros:

Sejam  $H(0), \dots, H(p)$  os dados históricos dos preços de um ativo durante  $p$  dias. Para simular um cenário futuro para daqui a  $D$  dias sortearmos  $\{j_0, j_1, \dots, j_{D-1}\}$ , onde  $j_i \in \{0, \dots, p-1\}$  e faremos por recorrência, partindo de  $y(0) = H(p)$ ,

$$y(t+1) = y(t) \frac{H(j_t+1)}{H(j_t)}$$

até obtermos  $y(D)$ .

A correlação entre os ativos é mantida se usarmos o mesmo conjunto sorteado para simular o cenário futuro para cada um dos ativos. Para obtermos  $m$  cenários basta sortear  $m$  conjuntos  $\{j_0, \dots, j_{D-1}\}$ .

### Exemplo

A tabela 4.2 contém os dados históricos dos preços de 3 ativos durante 11 dias.

dia	Petrobras (PN)	Bradesco (PN)	Vale (ON)
1	45,44	48,72	43,51
2	46,59	50,13	44,83
3	47,17	49,66	45,15
4	47,47	49,33	45,18
5	46,40	47,87	44,05
6	46,89	46,83	43,47
7	47,48	47,48	43,79
8	47,60	46,62	42,43
9	48,65	46,91	43,58
10	49,46	47,59	44,08
11	50,20	48,92	44,76

Tabela 4.2: Preço histórico dos ativos. Dia 11 é o último preço histórico, portanto refere-se ao dia de hoje.

Vamos fazer duas previsões para o preço dos ativos para daqui a dois dias.

Cenário 1 - dias sorteados  $\{2,8\}$

Os preços previstos nesse cenário são, respectivamente, 51,95; 48,76 e 46,30 para Petrobras, Bradesco e Vale.

Cenário 2 - dias sorteados  $\{6,3\}$



ativo	dia 1	dia 2
Petrobras (PN)	$50,20 \times \frac{46,59}{45,44} = 50,82$	$50,82 \times \frac{48,65}{47,60} = 51,95$
Bradesco (PN)	$48,92 \times \frac{49,66}{50,13} = 48,46$	$48,46 \times \frac{46,91}{46,62} = 48,76$
Vale (ON)	$44,76 \times \frac{45,15}{44,83} = 45,08$	$45,08 \times \frac{43,58}{42,43} = 46,30$

ativo	dia 1	dia 2
Petrobras (PN)	$50,20 \times \frac{47,48}{46,89} = 50,83$	$50,83 \times \frac{47,47}{47,17} = 51,15$
Bradesco (PN)	$48,92 \times \frac{47,48}{46,83} = 49,60$	$49,60 \times \frac{49,33}{49,66} = 49,27$
Vale (ON)	$44,76 \times \frac{43,79}{43,47} = 45,12$	$45,12 \times \frac{45,18}{45,15} = 45,12$

Para o segundo cenário os preços previstos são, respectivamente, 51,15; 49,27 e 45,12 para Petrobras, Bradesco e Vale.

Esse tipo de simulação, embora não tenha um forte embasamento teórico, assemelha-se ao *movimento browniano geométrico* por estar estreitamente ligado a suas hipóteses, além de ser de uma simplicidade muito grande.

---

---

# CAPÍTULO 5

---

## Introdução à fórmula Black Scholes

Nesse capítulo vamos fazer uma breve exposição sobre o mercado de opções e sobre a fórmula Black Scholes para que possamos usá-la em nossos testes computacionais como função de perda não linear.

Uma definição importante nesse contexto é a de derivativo o qual nada mais é que um tipo especial de ativo que depende de um outro ativo base. Uma classe especial de derivativos é chamada de opções. Uma opção está relacionada com o direito de comprar ou vender um ativo base por um preço pré-fixado.

Opções se dividem basicamente em 4 categorias:

- Call Européia
- Put Européia
- Call Americana
- Put Americana

Um investidor que compre uma opção do tipo Call  $O_c(A, K, T)$  adquire o direito de comprar um ativo  $A$  por preço  $K$  (preço de exercício). Para uma opção do tipo Put  $O_p(A, K, T)$  o investidor adquire o direito de vender um ativo  $A$  por um preço  $K$ . No

caso de uma opção européia o investidor pode exercer seu direito apenas no instante de tempo  $T$ . Para uma opção americana ele pode exercer o direito a qualquer momento até o instante de expiração  $T$ . As expressões européia e americana não significam que essas opções sejam negociadas apenas na Europa ou nos EUA, apenas denotam a especificação da opção.

### **Exemplo:**

Suponhamos que investimos nosso dinheiro em uma opção do tipo Call Européia cujo ativo base é uma saca de café, o preço de exercício é 130,00 reais e o prazo de expiração de um ano. Então daqui a um ano teremos o direito de comprar uma saca de café por 130,00 reais. Um dos motivos por termos adquirido essa opção pode ter sido fazer uma espécie de seguro contra um investimento duvidoso que tenhamos feito. Por exemplo, podemos vender hoje uma saca de café, a ser entregue daqui a um ano, por 130,00 reais. Fazemos isso pois acreditamos que após esse período será possível comprar a saca no mercado por 100,00 reais. Se nossa previsão estiver certa compraremos a saca no mercado por 100,00 reais e teremos um lucro. Porém, como não somos especialistas no assunto, existe a dúvida “E se a saca passar a valer 180,00 reais?”. Nesse caso exercemos a opção e a compramos pelos 130,00 reais estipulados como valor de exercício e podemos honrar nossos compromissos sem um grande prejuízo.

Porém, quanto vale direito? Essa precificação é o que a fórmula Black Scholes tenta fornecer. Ou seja, no exemplo anterior teríamos de pagar uma certa quantia a um investidor para que esse nos desse o direito de comprar uma saca de café daqui a um ano por 130,00 reais qualquer que seja o preço de mercado do produto na época.

Se hoje fosse a data de exercício não há dúvida que esse direito valeria  $\max\{0, y(T) - K\} = (y(T) - K)^+$  onde  $y(t)$  denota o preço do ativo no instante  $T$ . De fato, ninguém pagaria nada pelo direito de comprar um ativo por  $k$  reais se ele estivesse disponível no mercado por menos de  $k$ . Claramente se o preço do ativo tivesse maior do que o preço de exercício uma pessoa só se sujeitaria a dar o direito a outra de lhe comprar o ativo por  $k$  se este indivíduo lhe pagar a diferença até o valor de mercado. Por outro lado, não há dúvida de que esse seria o maior valor que uma pessoa estaria disposta a pagar por esse

direito, pois caso contrário seria mais vantajoso adquirir o ativo diretamente do mercado.

A situação não é tão simples assim quando não estamos na data de exercício pois há uma incerteza de quanto custará o ativo no futuro. Black e Scholes em [12] fizeram algumas hipóteses para tratar do assunto.

A primeira delas é que não há oportunidade de arbitragem, ou seja não há chance de um lucro certo. Um importante teorema nessa direção é o teorema da arbitragem. O teorema trata da seguinte situação: consideremos um experimento cujo conjunto de resultados possíveis seja  $\{1, 2, \dots, m\}$  e existam  $n$  itens nos quais podemos apostar. O retorno para cada real apostado no item  $i$ , caso aconteça o resultado  $j$  no experimento, é dado por  $r_i(j)$ . Permitimos que as apostas possam ser positivas, negativas ou nulas. Nesse caso, se acontecer o resultado  $j$  no experimento, o retorno associado a uma estratégia  $x \in \mathbb{R}^n$ , onde  $x_i$  é a quantidade apostada no item  $i$ , é:

$$\text{retorno de } x = \sum_{i=1}^n x_i r_i(j)$$

Podemos então enunciar o resultado.

**Teorema 5.0.4.** *Teorema da Arbitragem*

*Exatamente uma das afirmações é verdadeira:*

a) *Existe um vetor de probabilidades  $p = (p_1, \dots, p_m)$  tal que:*

$$\sum_{j=1}^m p_j r_i(j) = 0 \quad \forall i = 1, \dots, n$$

b) *Existe uma estratégia  $x = (x_1, \dots, x_n)$  para a qual:*

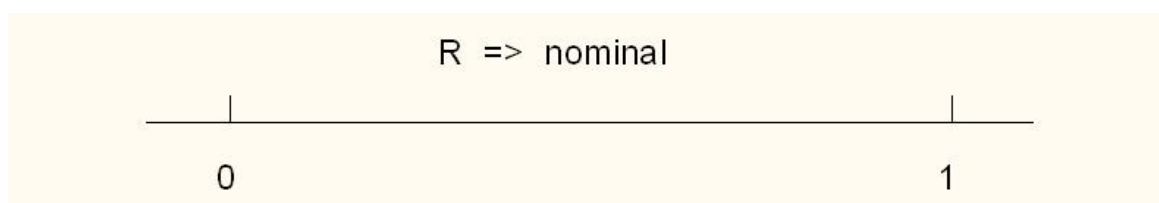
$$\sum_{j=1}^m x_i r_i(j) > 0 \quad \forall j = 1, \dots, m$$

Uma demonstração deste teorema pode ser encontrada em [15] e baseia-se na dualidade de um problema de programação linear associado.

É fácil relacionar o teorema com o fato de termos  $m$  cenários futuros e podermos investir em  $n$  ativos uma quantidade  $x_i$  que aqui pode ser negativa, indicando a venda do ativo, o que é bastante natural, principalmente quando se trata de opções. Nesse contexto o teorema está dizendo que ou existe um vetor de probabilidade para que os cenários ocorram de tal forma que o lucro esperado será zero para todos os ativos, ou existiria uma oportunidade de arbitragem, ou seja existiria um investimento onde ganharíamos em qualquer que fosse o cenário futuro.

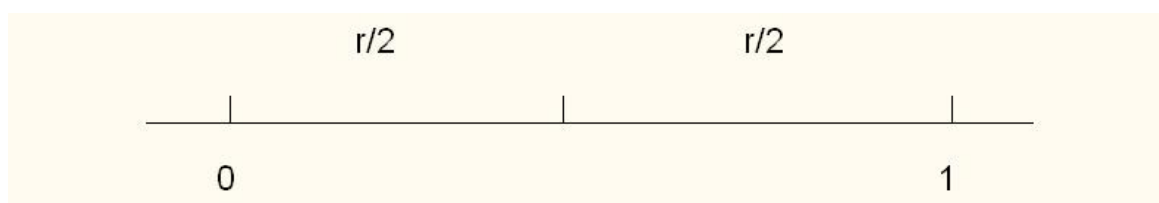
A segunda hipótese feita por Black e Scholes é que os incrementos no preço de um determinado ativo são calculados continuamente. Expliquemos o que isso quer dizer.

Suponhamos que um investidor possua um capital  $c(t)$  aplicado em um fundo que lhe paga uma taxa de juros nominal de  $100r\%$  por período de tempo  $T$ . Logo, decorrido o período  $P$ , o capital do investidor será:



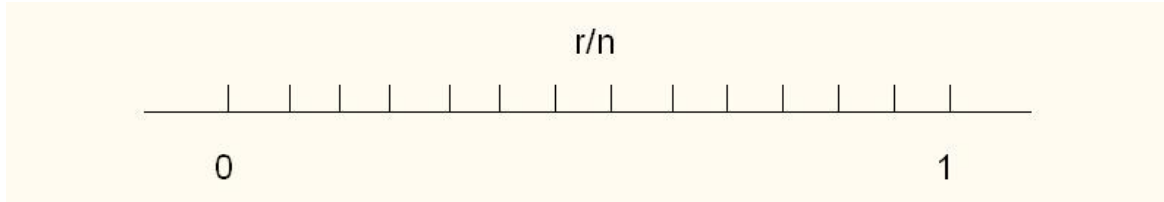
$$c(T) = c(0) (1 + r)$$

Se o juros fosse computado duas vezes durante o período (uma vez na metade do período e outra no final) o capital do investidor seria:



$$c(T) = c(0) \left(1 + \frac{r}{2}\right)^2$$

Se o juro fosse computado  $n$  vezes durante o período:



$$c(T) = c(0) \left(1 + \frac{r}{n}\right)^n$$

Quando  $n$  tende a infinito, ou seja estamos computando os incrementos a todo instante, temos que:

$$c(T) = c(0) e^r$$

Sem perda de generalidade podemos supor  $T$  como a unidade de tempo e teríamos:

$$c(1) = c(0) e^r$$

Se tivéssemos  $t$  períodos de tempo:

$$c(t) = c(0) \left[\left(1 + \frac{r}{n}\right)^n\right]^t \Rightarrow c(t) = c(0) e^{rt}$$

Quando os incrementos são feitos dessa forma dizemos que são calculados continuamente.

A terceira hipótese diz que o preço do ativo base segue um *movimento browniano geométrico* com parâmetros  $\mu$  e  $\sigma$  (explicado na seção anterior). Um comentário pertinente aqui é observar que a exponencial envolvida na distribuição log-normal concorda com o fato dos incrementos serem calculados continuamente.

Sob as hipóteses anteriores se  $r$  for a taxa de juros referente a uma aplicação sem risco, por exemplo poupança, é possível mostrar [15], através do teorema da arbitragem, que  $\mu = r - \frac{\sigma^2}{2}$ . Dessa forma o preço do ativo base será  $y(t) = y(0)e^W$  onde  $W \sim N(\mu t, \sigma\sqrt{t})$ .

Lembramos que o preço de uma call-option na data de exercício é  $(y(t) - K)^+$ . Analisemos o caso geral. Suponhamos que exista uma taxa de juros sem risco (referente a

poupança ou a inflação). Neste caso temos que comparar os investimentos no mesmo instante de tempo. Sendo assim o preço  $p_c$  pago por uma call-option européia hoje valeria  $e^{rt} \cdot p_c$  na data de exercício e pelo teorema da arbitragem o preço justo da opção nesta data seria  $E[(y(t) - K)^+]$ . Daí:

$$p_c = e^{-rt} E[(y(t) - K)^+]$$

Para calcularmos  $p_c$  faremos algumas manipulações algébricas. Primeiro vamos escrever  $W$  como  $\left(r - \frac{\sigma^2}{2}\right)t + \sigma\sqrt{t}Z$ , onde  $Z$  é uma variável aleatória com distribuição normal padrão. Dessa forma temos que:

$$\begin{aligned} y(t) > K &\iff y(0) \exp\left(\left(r - \frac{\sigma^2}{2}\right)t + \sigma\sqrt{t}Z\right) > K \\ &\iff \exp\left(\left(r - \frac{\sigma^2}{2}\right)t + \sigma\sqrt{t}Z\right) > \frac{K}{y(0)} \\ &\iff \left(r - \frac{\sigma^2}{2}\right)t + \sigma\sqrt{t}Z > \ln\left(\frac{K}{y(0)}\right) \\ &\iff Z > \frac{\ln\left(\frac{K}{y(0)}\right) - \left(r - \frac{\sigma^2}{2}\right)t}{\sigma\sqrt{t}} \\ &\iff Z > \sigma\sqrt{t} - d \end{aligned}$$

onde

$$d = \frac{rt + \sigma^2 t/2 - \ln\left(\frac{K}{y(0)}\right)}{\sigma\sqrt{t}}$$

Definindo  $I$  como a função indicadora do uso da opção temos:

$$I = \begin{cases} 1 & \text{se } y(t) > K \\ 0 & \text{se } y(t) \leq K \end{cases}$$

Pelo que acabamos de ver:

$$\begin{aligned} E[I] &= P[y(t) > K] \\ &= P[Z > \sigma\sqrt{t} - d] \\ &= P[Z < d - \sigma\sqrt{t}] \\ &= N(d - \sigma\sqrt{t}), \end{aligned}$$

onde  $N(x)$  é a função de probabilidade acumulada de uma normal padrão.

Provemos agora que  $e^{-rt}E[Iy(t)] = y(0)N(d)$ . Para isso consideremos  $c = \sigma\sqrt{t} - d$  então:

$$\begin{aligned}
 E[Iy(t)] &= \int_c^\infty y(0)e^{\left(r-\frac{\sigma^2}{2}\right)t+\sigma\sqrt{t}x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
 &= \frac{1}{\sqrt{2\pi}} y(0)e^{rt} \int_c^\infty e^{-\frac{\sigma^2 t}{2} + \sigma\sqrt{t}x - \frac{x^2}{2}} dx \\
 &= y(0)e^{rt} \frac{1}{\sqrt{2\pi}} \int_c^\infty e^{-\frac{(x-\sigma\sqrt{t})^2}{2}} dx \\
 &= y(0)e^{rt} \frac{1}{\sqrt{2\pi}} \int_{-d}^\infty e^{-\frac{y^2}{2}} dy \quad \left(\text{fazendo } y = x - \sigma\sqrt{t}\right) \\
 &= y(0)e^{rt} P[Z > -d] \\
 &= y(0)e^{rt} N(d)
 \end{aligned}$$

Por fim temos que:

$$\begin{aligned}
 p_c &= e^{-rt}E[y(t) - K]^+ \\
 &= e^{-rt}E[I \cdot (y(t) - K)] \\
 &= e^{-rt}E[Iy(t) - IK] \\
 &= e^{-rt}E[Iy(t)] - Ke^{-rt}E[I] \\
 &= y(0)N(d) - Ke^{-rt}N(d - \sigma\sqrt{t})
 \end{aligned}$$

Chegamos então à famosa fórmula “Black Scholes” para opções do tipo Call européia.

$$p_c(y(0), t, K, \sigma, r) = y(0)N(d) - Ke^{-rt}N(d - \sigma\sqrt{t}) \quad (5.1)$$

onde

$$d = \frac{rt + \sigma^2 t/2 - \ln\left(\frac{K}{y(0)}\right)}{\sigma\sqrt{t}}$$



---

---

# CAPÍTULO 6

---

## Testes computacionais e conclusões

Neste capítulo apresentaremos os resultados computacionais que obtivemos com os algoritmos por nós propostos na seção 3.5 para minimização da  $C$ - $VaR$  e da  $VaR$ . Afim de compararmos nossos resultados com os publicados em [2], [6] e [14], apresentaremos também uma análise quantitativa de nossas implementações para tais algoritmos. Além disso, vamos expor algumas conclusões que observamos em tais testes.

Os experimentos foram realizados em um computador com processador Pentium 4 2.4 GHz e 1 GB de memória RAM. Os testes relativos aos problemas não lineares foram feitos através do *software* *Algencan*, disponível em [www.ime.unicamp.br/martinez/](http://www.ime.unicamp.br/martinez/). Já os experimentos relativos aos problemas lineares foram feitos através do GLPK (simplex disponível em [www.gnu.org/software/glpk/](http://www.gnu.org/software/glpk/)) e do PCX (algoritmo de pontos interiores disponível em [www-fp.mcs.anl.gov/OTC/Tools/PCx/](http://www-fp.mcs.anl.gov/OTC/Tools/PCx/)).

Os testes computacionais foram feitos tanto com problemas artificiais, provenientes de dados aleatórios, quanto de problemas reais. Estes últimos foram realizados com dados históricos de 25/07/03 a 29/06/07 das vinte empresas com maior peso no IBOVESPA que disponibilizavam os valores de suas ações por todo o período analisado.

Para os problemas não lineares implementados no *Algencan* fornecemos suas deriva-

das analíticas sempre que estas existiam. Quando as primeiras derivadas não podiam ser calculadas ajustávamos o parâmetro “gtype = 1” e quando as segundas derivadas não existiam ajustávamos “hptype = 4”, usando assim diferenças finitas.

A seguir apresentaremos alguns resultados obtidos, representando cada grupo de testes feitos. Os algoritmos serão indicados por:

- SD → Algoritmo sem a suavização, trabalhamos diretamente com  $\max\{0, x\}$ .
- Pol2 → Suavização utilizando o polinômio de grau 2.
- Pol4 → Suavização utilizando o polinômio de grau 4.
- Mod → Suavização utilizando a aproximação do módulo.
- PCX → Algoritmo de pontos interiores para resolução de problemas de programação linear.
- GLPK → Algoritmo do método Simplex para resolução de problemas de programação linear.
- RDOVO → Algoritmo apresentado em [6].

Nos primeiros testes, de 1 a 11, resolvemos problemas de minimizar a CVaR. Os testes de 1 a 3 abordavam problemas artificiais de pequeno porte e tinham como objetivo validar as implementações realizadas.

	GLPK	PCX	SD	Pol2	Pol4	Mod
it	27	9	50	2	2	2
tempo	0,00	0,03	0,41	0,05	0,02	0,00
erro	0	0	0	0	0	0

Tabela 6.1: Teste 1- 4 ativos, 20 cenários. Resultados para  $\epsilon = 1$ .  $erro \equiv \frac{\|X_{GLPK} - X_i\|}{\|X_{GLPK}\|}$

	GLPK	PCX	SD	Pol2	Pol4	Mod
it	1095	29	50	4	2	2
tempo	0,80	0,62	25,34	0,13	0,17	0,16
erro	0	0	0	0	0	0
CVaR	-4.539	-4.539	-4.539	-4.539	-4.539	-4.539

Tabela 6.2: Teste 2- 15 ativos, 1000 cenários. Resultados para  $\epsilon = 1$ .  $erro \equiv \frac{\|X_{GLPK} - X_i\|}{\|X_{GLPK}\|}$

	GLPK	PCX	PCX*	SD	Pol2	Pol4	Mod
it	2347	31	39	50	2	2	2
tempo	4,8	0,97	1,63	70,08	0,06	0,13	0,13
erro	0	0,14	0	0	0	0	0
CVaR	-5.102	-5.019	-5.102	-5.102	-5.102	-5.102	-5.102

Tabela 6.3: Teste 3- 15 ativos, 2200 cenários. Resultados para  $\epsilon = 1$ .  $erro \equiv \frac{\|X_{GLPK} - X_i\|}{\|X_{GLPK}\|}$

Podemos ver que as implementações estão coerentes, uma vez que as soluções encontradas são, substancialmente, as mesmas para todos os algoritmos utilizados. Entretanto, ressaltamos aqui a necessidade de modificar o parâmetro de centragem no PCX para obter a convergência do algoritmo para o problema com 15 variáveis e 2.200 cenários. Esse resultado repitiu-se para problemas maiores e pode representar uma possível dificuldade de algoritmos de pontos interiores ao tratarem problemas com essa estrutura.

Nos testes 4 e 5 foram utilizados um vetor de custo aleatório uniformemente distribuído entre 1 e 10, e uma matriz , representando as funções de perda lineares, com entradas entre -8 e 2, também uniformemente distribuída. Consideraremos aqui vinte ativos, dez mil e cinquenta mil cenários. Os objetivos nestes testes eram averiguar o comportamento dos algoritmos suavizados em relação a variação de  $\epsilon$  e comparar a eficiência destes com a de um PL.

	GLPK	PCX	SD	Pol2	Pol4	Mod	Pol2	Pol4	Mod
it	10.628	31	50	50	50	50	50	2	2
tempo	262,7	5,84	2.456	27,64	349	388	138	3,08	1,92
eps	-	-	-	0.1	0.1	0.1	1	1	1
CVaR	-6.541	-75	-6.540	-6.540	-6.540	-6.540	-6.540	-6.540	-6.540

Tabela 6.4: Teste 4- 20 ativos, 10.000 cenários. Resultados para  $\epsilon \leq 0,1$ .

	Pol2	Pol4	Mod	Pol2	Pol4	Mod	Pol2	Pol4	Mod
it	50	2	2	4	2	2	10	3	2
tempo	31,72	1,2	0,92	2,03	0,44	0,55	5,14	0,45	0,53
eps	10	10	10	$10^2$	$10^2$	$10^2$	$10^3$	$10^3$	$10^3$
CVaR	-6.540	-6.540	-6.540	-6.541	-6.541	-6.540	-6.541	-6.541	-6.541

Tabela 6.5: Teste 4- 20 ativos, 10.000 cenários. Resultados para  $10 \leq \epsilon \leq 1.000$ .

	Pol2	Pol4	Mod	Pol2	Pol4	Mod	Pol2	Pol4	Mod
it	6	3	3	4	4	3	10	3	3
tempo	2,02	1,2	0,33	1,98	0,48	0,39	7,64	0,41	0,63
eps	$10^4$	$10^4$	$10^4$	$10^5$	$10^5$	$10^5$	$10^6$	$10^6$	$10^6$
CVaR	-6.035	-6.185	-6.541	2.185	2.185	-6.540	2.296	2.274	-6.508

Tabela 6.6: Teste 4- 20 ativos, 10.000 cenários. Resultados para  $\epsilon \geq 10.000$ .

	GLPK	PCX	SD	Pol2	Pol4	Mod	Pol2	Pol4	Mod
it	52.160	31	50	50	2	3	5	7	2
tempo	35.570	61,56	9.694	93,28	9,97	7,91	13,38	16,61	3,25
eps	-	-	-	1	1	1	10	10	10
CVaR	-6.605	-32	-6.605	-6.605	-6.005	-6.605	-6.605	-6.605	-6.605

Tabela 6.7: Teste 5- 20 ativos, 50.000 cenários. Resultados para  $\epsilon \leq 10$ .

---

	Pol2	Pol4	Mod	Pol2	Pol4	Mod	Pol2	Pol4	Mod
it	50	3	2	7	3	9	50	4	50
tempo	586	2,13	5,66	25,16	7,61	16,25	177,17	4,09	108,56
eps	$10^2$	$10^2$	$10^2$	$10^3$	$10^3$	$10^3$	$10^4$	$10^4$	$10^4$
CVaR	-6.605	-6.605	-6.605	-6.605	-6.603	-6.605	-6.113	-6.268	-6.604

Tabela 6.8: Teste 5- 20 ativos, 50.000 cenários. Resultados para  $100 \leq \epsilon \leq 10.000$ .

	Pol2	Pol4	Mod	Pol2	Pol4	Mod
it	6	5	3	5	4	3
tempo	12,31	649	1,73	19,80	10,92	1,70
eps	$10^5$	$10^5$	$10^5$	$10^6$	$10^6$	$10^6$
CVaR	2.010	2.010	-6.602	2.040	2.035	-6.567

Tabela 6.9: Teste 5- 20 ativos, 50.000 cenários. Resultados para  $\epsilon \geq 100.000$ .

Observamos que para valores de  $\epsilon$  muito pequenos os algoritmos encontram a solução do problema mas não atingem o critério de parada indicando otimalidade. Acreditamos que isso aconteça devido ao fato dos problemas estarem numericamente muito próximos de problemas não diferenciáveis, logo suas condições KKT podem, numericamente, não existir. O mesmo acontece com o problema não diferenciável, o qual pode ser visto como limite dos anteriores. Para valores de  $\epsilon$  muito altos os problemas apresentavam convergência, entretanto, o ponto encontrado não é solução do problema original. Isso deve-se ao fato de estarmos modificando demasiadamente o problema. Quanto a comparação com os algoritmos que resolvem a formulação linear obtemos uma grande vantagem. Quanto a robustez o PCX mostrou-se ineficiente em todos os testes feitos. Já em relação a velocidade vemos uma grande superioridade dos algoritmos não lineares em relação ao GLPK.

O teste 6 refere-se apenas a uma mudança de escala no teste 5. Nele  $x_i$  não representa a quantidade de ações adquiridas mas sim a quantidade comprada de lotes de mil ações. Portanto o valor do vetor de custo e da matriz com os coeficientes das funções de perda foram multiplicados por mil. Tomamos o cuidado de dividir o ponto inicial pelo mesmo fator.

	Pol2	Pol4	Mod	Pol2	Pol4	Mod	Pol2	Pol4	Mod
it	50	50	50	50	50	5	50	50	50
tempo	6.164	72,73	892	2.655	69,59	10,75	3.477	7,19	78,36
eps	1	1	1	10	10	10	$10^2$	$10^2$	$10^2$
CVaR	-6.605	-32	-6.605	-6.605	-6.605	-6.605	-6.605	-6.605	-6.605

Tabela 6.10: Teste 6- Mudança de escala do teste 5. Resultados para  $\epsilon \leq 1$ .

	Pol2	Pol4	Mod	Pol2	Pol4	Mod	Pol2	Pol4	Mod
it	50	3	3	50	4	3	5	4	3
tempo	2.675	6,06	6,06	564	5,16	7,42	19,8	2,55	6,19
eps	$10^3$	$10^3$	$10^3$	$10^4$	$10^4$	$10^4$	$10^6$	$10^6$	$10^6$
CVaR	-6.602	-6.603	-6.605	-6.113	-6.268	-6.004	2.040	2.350	-6.567

Tabela 6.11: Teste 6- Mudança de escala do teste 5. Resultados para  $\epsilon \geq 1.000$ .

Como esperado, a solução encontrada continua a mesma, a não ser pela mudança de escala. Entretanto, quanto à eficiência e à robustez dos métodos tivemos diferenças consideráveis. Esse fato não era previsto pois a função  $\phi$  age em  $(f_j(x) - t)$  e esse valor deveria continuar como nos testes anteriores, pois os valores de  $x$  são dividido por mil pra satisfazer a restrição  $\sum_{i=1}^n y_i^0 x_i \leq M$ . Podemos perceber que o número de algoritmos onde obtemos convergência é bem menor nesse caso.

No teste 7 realizamos, de certa forma, outra mudança de escala. Aumentamos a quantidade de dinheiro disponível multiplicando-a por um fator de mil. Dessa forma o argumento  $(f_j(x) - t)$ , a ser utilizado em  $\phi$ , também é multiplicado por mil, pois o vetor  $x$  assim deve ser. Novamente tomamos o cuidado de ajustar o ponto inicial, dessa vez, multiplicando-o pelo mesmo fator.

	Pol2	Pol4	Mod	Pol2	Pol4	Mod	Pol2	Pol4	Mod
it	50	50	50	50	5	11	22	9	9
tempo	151	2.993	3.304	92,28	6,19	6,73	67,41	4,92	6,14
eps	0,1	0,1	0,1	1	1	1	10	10	10
CVaR	-6.605	-6.605	-6.605	-6.602	-6.603	-6.567	-6.113	-6.268	-6.264

Tabela 6.12: Teste 7- Mudança de escala do teste 5, aumentando o dinheiro disponível.

Podemos perceber claramente que os resultados mais encorajadores foram obtidos para valores de  $\epsilon$  bem menores do que os anteriores, entretanto esse comportamento não foi linear. Mais uma vez pudemos ver uma mudança na faixa dos valores razoáveis.

Nos testes de 8 a 10 utilizamos dados históricos reais e construímos nossos cenários basicamente como exposto no capítulo 4. Tivemos apenas que fazer alguns ajustes pois haviam cenários demasiadamente pessimistas o que, muitas vezes, fazia que a decisão ótima fosse não investir o dinheiro.

	GLPK	Pol2	Pol4	Mod
it	9.731	50	5	11
tempo	237	180	2,61	3,75
eps	-	10	10	10
CVaR	-184	-184	-184	-184

Tabela 6.13: Teste 8- Dados reais 20 ativos, 10.000 cenários.

	GLPK	Pol2	Pol4	Mod
it	48.087	50	2	2
tempo	33.216	196,91	2,66	3,77
eps	-	10	10	10
CVaR	-184	-184	-184	-184

Tabela 6.14: Teste 9- Dados reais 20 ativos, 50.000 cenários.

	GLPK	Pol2	Pol4	Mod
it	94.886	50	3	3
tempo	252.120	298	22,84	3,39
eps	-	100	100	100
CVaR	-215	-215	-215	-215

Tabela 6.15: Teste 10- Dados reais 20 ativos, 100.000 cenários.

Para valores reais, portanto os mais expressivos, novamente pudemos perceber uma grande vantagem, na questão de tempo, em relação à formulação de programação linear.



No teste 11 modificamos a matriz relativa às funções de perda do problema introduzindo um resíduo uniformemente distribuído entre 0 e 5%. Com isso simulamos erros na previsão dos cenários para ver quanto a solução obtida anteriormente distancia-se da nova solução.

	Pol2	Pol4	Mod
it	50	50	3
tempo	315	172	5,05
eps	100	100	100
$\frac{\ X_{T10}-X_{T11}\ }{\ X_{T10}\ }$	0,00	0,00	0,00
CVaR	-218	-218	-218

Tabela 6.16: Teste 11- Teste de sensibilidade acrescentando resíduos no teste 10.

A proximidade entre as soluções aponta para a estabilidade dessa abordagem de risco.

Os próximos testes são relativos a minimização da VaR. Os maiores problemas que resolvemos apresentavam cinco ativos e trinta cenários. O teste 12 é referente à um problema artificial. Já os problemas 13 e 14 são relativos a dados reais. Os valores encontrados nas tabelas referem-se a solução de melhor valor para a função objetivo.

	RDOVO	Pol2	Pol4	Mod	SD
it	6	10	9	7	10
tempo	0,05	0,05	0,02	0,00	0,02
eps	-	1	1	1	-
VaR	-29,69	-25,97	-12,41	-25,75	-25,97

Tabela 6.17: Teste 12- 5 ativos, 30 cenários.

	RDOVO	Pol2	Pol4	Mod
it	8	7	11	10
tempo	0,14	0,02	0,03	0,03
eps	-	10	100	10
VaR	-32,70	-69,00	-56,04	-62,17

Tabela 6.18: Teste 13- 5 ativos, 30 cenários.

	RDOVO	Pol2	Pol4	Mod
it	9	10	12	8
tempo	0,22	0,02	0,05	0,03
eps	-	1000	10	10
VaR	-29	-11	-47	-40

Tabela 6.19: Teste 14- 5 ativos, 30 cenários.

Podemos notar uma melhora substancial em relação ao tempo se compararmos percentualmente o algoritmo por nós proposto e o descrito em [6]. Entretanto, como só foi possível resolver com sucesso problemas pequenos, a diferença de tempo absoluta é muito pequena o que impede afirmar a superioridade de nosso algoritmo. Quanto a qualidade das soluções não pudemos tirar grandes conclusões, pois, a quantidade de minimizadores locais encontrados, nos dois casos, dificulta nossa análise.

Por fim, no teste 15 minimizamos a C-VaR relativa a perdas não lineares onde era possível comprar  $x_1$  unidades de um ativo base e  $x_2$  unidades da *Call Option*, com preço de exercício  $x_3$ , a ele relacionada. O objetivo neste teste era mostrar a aplicabilidade das suavizações para funções de perda não lineares. Obtivemos a convergência do Algencan.

	Pol2	Pol4	Mod	SD
it	2	2	2	2
tempo	0,98	0,32	0,48	1,64
eps	1	1	1	-
VaR	-130	-129	-129	-130

Tabela 6.20: Teste 15- Funções de perda não lineares.

---

## 6.1 Conclusões finais e perspectivas futuras

---

Com este trabalho pudemos estudar vários modelos matemáticos com aplicações em economia. A parte central do texto, voltada à otimização, foi realizada com sucesso. Conseguimos reproduzir e sugerir algoritmos para minimização da VaR e da C-VaR. Nos testes computacionais conseguimos ligar aspectos teóricos a situações reais. Percebemos em tais testes uma deficiência do PCX em resolver os problemas de minimizar a C-VaR com a estrutura que o formulamos. Constatamos a maior eficiência das formulações não lineares em relação ao GLPK, entretanto averiguamos a sensibilidade dos algoritmos suavizados em relação a  $\epsilon$ . Notamos a suavização do módulo mais robusta do que as demais. Acreditamos que a robustez obtida deve-se ao fato desta ser a função mais suave dentre as quais trabalhamos. Quanto à eficiência, observamos que a suavização utilizando o polinômio de grau quatro está no mesmo patamar da suavização do módulo, isto pode ser explicado por sua maior semelhança à função original  $(x)^+$ .

O algoritmo proposto para minimizar a VaR mostrou-se comparável com o publicado em [6]. Deixamos como sugestão de trabalhos futuros a utilização de técnicas de globalização para o problema da VaR, para que assim possamos tirar conclusões mais efetivas. Outra sugestão importante que deixamos para trabalhos futuros é incorporar as restrições na função objetivo e obter expressões para os multiplicadores de Lagrange, assim como feito em [6]. Expressões para os multiplicadores são muito importantes pois podem combater problemas de voracidade que encontramos em alguns de nossos testes. Em nossos algoritmos contornamos o problema limitando inferiormente a variável a ser minimizada usando estimativas baseadas no problema de minimizar a CVaR. Por fim, ainda seguindo

a linha apresentada em [6] , sugerimos uma análise na relação entre mínimos locais e pontos KKT do problema reformulado e do problema original.

---

# BIBLIOGRAFIA

- [1] C. Acerbi, D. Tasche, Expected Shortfall: a natural coherent alternative to Value at Risk. *Economic Notes*, 31 2 (2002), 379-388.
- [2] S. Alexander, T. Coleman and Y. Li, Minimizing CVaR and VaR for a Portfolio of Derivatives *Journal of Banking and Finance* 30 (2006) 583–605
- [3] S. Alexander, T. F. Coleman, and Y. Li. Derivative portfolio hedging based on CVaR. In G. Szego, editor, *Risk Measures for the 21st Century*, Wiley, (2003) 339–363.
- [4] R. Andreani, E. G. Birgin, J. M. Martinez and M. L. Schuverdt. Augmented Lagrangian methods under the Constant Positive Linear Dependence constraint qualification. To appear in *Mathematical Programming*.
- [5] R. Andreani, E. G. Birgin, J. M. Martinez and M. L. Schuverdt. On Augmented Lagrangian methods with general lower-level constraints. Technical Report, Department of Applied Mathematics, Unicamp, 2005.
- [6] R. Andreani, C. Dunder and J. M. Martinez, Nonlinear-Programming Reformulation of the Order-Value Optimization problem *Math. Methods Oper. Res.* 61 (2005), no. 3, 365–384.
- [7] R. Andreani, C. Dunder and J. M. Martinez, Order-Value Optimization: Formulation and Solution by means of a Primal Cauchy Method *Math. Methods Oper. Res.* 58 (2003), no. 3, 387–399.

- [8] R. Andreani, J. M. Martinez, M. Salvatierra and F. Yano, Global Order-Value Optimization by means of a multistart harmonic oscillator tunneling strategy *Nonconvex Optim. Appl.*, 84, Springer, New York, 2006.
- [9] R. Andreani, J. M. Martinez, M. Salvatierra and F. Yano, Quasi-Newton methods for Order-Value Optimization and Value-at-Risk calculations *Pacific Journal of Optimization*, Vol. 2, pp. 11-33 (2006)
- [10] P. Artzner, F. Delbaen, J. Eber, e D. Heath Thinking coherent, *Risk* , **10** (1997), 68–71.
- [11] P. Artzner, F. Delbaen, J. Eber, e D. Heath Coherent Measures of Risk. *Mathematical Finance*, **9** (1999), 2003-228.
- [12] F. Black and M. Scholes, The price of Options and Corporate Liabilities *Journal of Political Economy* 81: 637-59
- [13] B. James, “Probabilidade: Um Curso em Nível Intermediário”, IMPA - Projeto Euclides, RJ 2002
- [14] R. T. Rockafellar and S. Uryasev, Optimization of Conditional Value-at-Risk *Journal of Risk 2* Pages: 21-41 January 2000
- [15] S. Ross, “An elementary Introduction to Mathematical Finance”, Cambridge University Press 2003
- [16] S. Uryasev, Conditional Value-at-Risk: Optimization Algorithms and Applications *Financial Engineering News* January 2000