

3

Fundamentos Teóricos

3.1

O Mercado de Opções

O Mercado de Opções sobre Ações é o mercado onde são negociados direitos de compra ou venda de um lote de ações, com preços e prazos de exercício pré-estabelecidos.

O mercado de opções surgiu com a finalidade de criar um mecanismo de proteção aos investidores contra possíveis perdas no mercado, um seguro contra fortes oscilações futuras dos ativos. “No Brasil, o mercado organizado de opções existe desde 1979, implantado inicialmente pela Bolsa de Valores de São Paulo” (Bessada, Barbedo e Araujo, 2007). Considerando que os instrumentos financeiros têm seus preços e retornos expostos à flutuações imprevisíveis, as opções são usadas para adaptar o risco às expectativas e metas do investidor.

Um agricultor produtor de soja, que tem custos fixos no seu produto e não pode depender do humor do mercado e das oscilações dos preços do seu ativo até o momento da colheita e venda da produção, pode utilizar as opções como um seguro e pagar um prêmio na data do plantio para ter o direito de vender o seu produto a um preço que lhe seja vantajoso na hora da colheita, alguns meses à frente.

Com as opções o investidor tem também a possibilidade de “alavancar” sua posição, aumentando o potencial retorno de um investimento, sem aumentar a quantidade de capital investido, uma vez que o capital inicial investido para comprar uma opção é relativamente pequeno em comparação com o ganho.

Uma opção é o direito de comprar ou vender um ativo específico, por um preço, adquirido mediante o pagamento de um valor (o prêmio), para ser exercido em uma data pré-estabelecida (data de vencimento). Titular é o investidor que compra a opção e adquire os direitos (de comprar ou vender ações) a ela referentes. Lançador é o investidor que vende a opção e assume os compromissos a ela referentes.

O prêmio - ou preço da opção - é negociado entre o comprador e o vendedor no momento da operação em mercado, pago no momento da aquisição da opção e é influenciado por diversos fatores.

O preço de exercício, comumente chamado de strike, é o valor pactuado no contrato da opção no qual o titular terá o direito de pagar/receber pelo ativo na data de exercício.

3.1.1

Opção de Compra

Numa opção de compra (CALL), o comprador (titular) adquire o direito de comprar as ações-objeto a preço e prazo pré-determinados. O vendedor (lançador) assume a obrigação de vender as ações-objeto, caso o direito seja exercido pelo titular. Ele recebe, por isso, o valor do prêmio.

O resultado de uma CALL em seu vencimento consiste no máximo entre zero e o valor do ativo-objeto (S) subtraído do preço de exercício (K), Eq. (1).

$$CALL = \text{máx}(0, S - K) \quad \text{Eq. (1)}$$

Caso o investidor seja o titular subtrai-se o valor do prêmio pago ao resultado da opção, caso seja o lançador soma-se o valor do prêmio recebido. Podemos ver a seguir como se comportam os gráficos de retorno financeiro no vencimento de uma CALL, com a variação preço do ativo-objeto.

Na Figura 1 e Figura 2 podem-se ver exemplos de titular e lançador para uma Call com R\$60,00 de preço de exercício e R\$2,00 de prêmio. Na Figura 1 vemos que, no vencimento, o titular perde os R\$2,00 pagos no prêmio caso o ativo-objeto não fique acima de R\$60,00 (strike da opção) e começa a ter lucro somente quando o ativo ultrapassa R\$62,00 (strike + prêmio). Para o lançador na Figura 2 o resultado é inverso. O lançador embolsa todo prêmio de R\$2,00 recebido caso o ativo-objeto fique abaixo do strike e começa a perder a partir de R\$62,00. A negociação de uma opção é equivalente a um jogo de soma-zero.

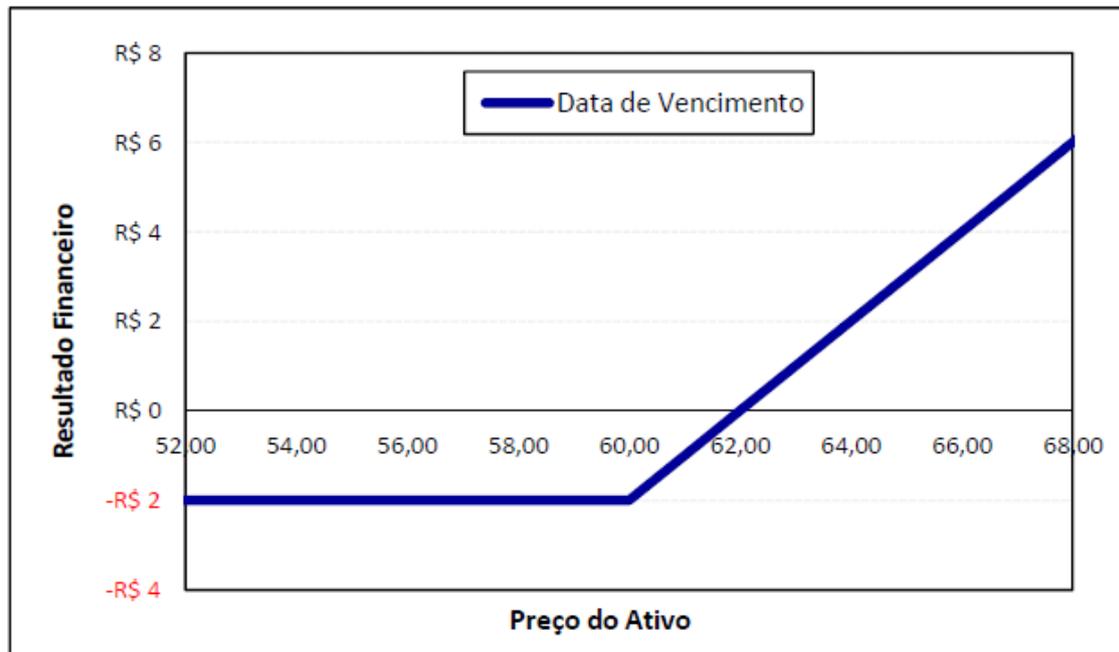


Figura 1 – Payoff para o titular de uma Call

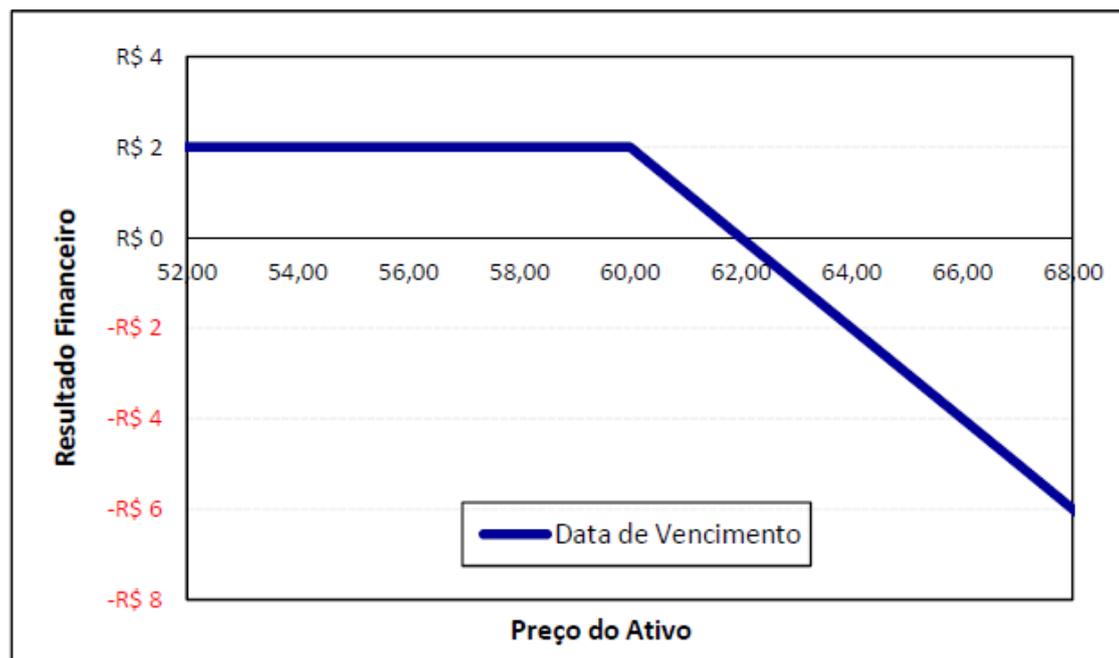


Figura 2 – Payoff para o lançador de uma Call

3.1.2

Opção de Venda

Em uma opção de venda (PUT), o titular compra o direito de vender as ações-objeto a determinado preço e prazo. O lançador assume a obrigação de comprar as ações-objeto no caso do titular exercer seu direito.

O resultado de uma PUT em seu vencimento é no máximo entre zero e o valor do preço de exercício (K) subtraído do valor do ativo-objeto (S), Eq. (2).

$$PUT = \text{máx}(0, K - S) \quad \text{Eq. (2)}$$

Caso o investidor seja o titular subtrai-se o valor do prêmio pago ao resultado da opção, caso seja o lançador soma-se o valor do prêmio recebido. Podemos ver a seguir como se comportam os gráficos de retorno financeiro no vencimento de uma Put, com a variação preço do ativo-objeto.

Na Figura 3 e Figura 4 pode-se ver exemplos de titular e lançador para uma Put com R\$60,00 de preço de exercício e R\$2,00 de prêmio. Na Figura 3 vemos que, no vencimento, o titular perde os R\$2,00 pagos no prêmio caso o ativo-objeto acabe acima de R\$60,00 (strike da opção) e começa a ter lucro somente quando o ativo cai abaixo de R\$58,00 (strike - prêmio). Para o lançador na Figura 4 o resultado é inverso. O lançador embolsa todo prêmio de R\$2,00 recebido caso o ativo-objeto acabe acima do strike e começa a perder a partir de R\$58,00.

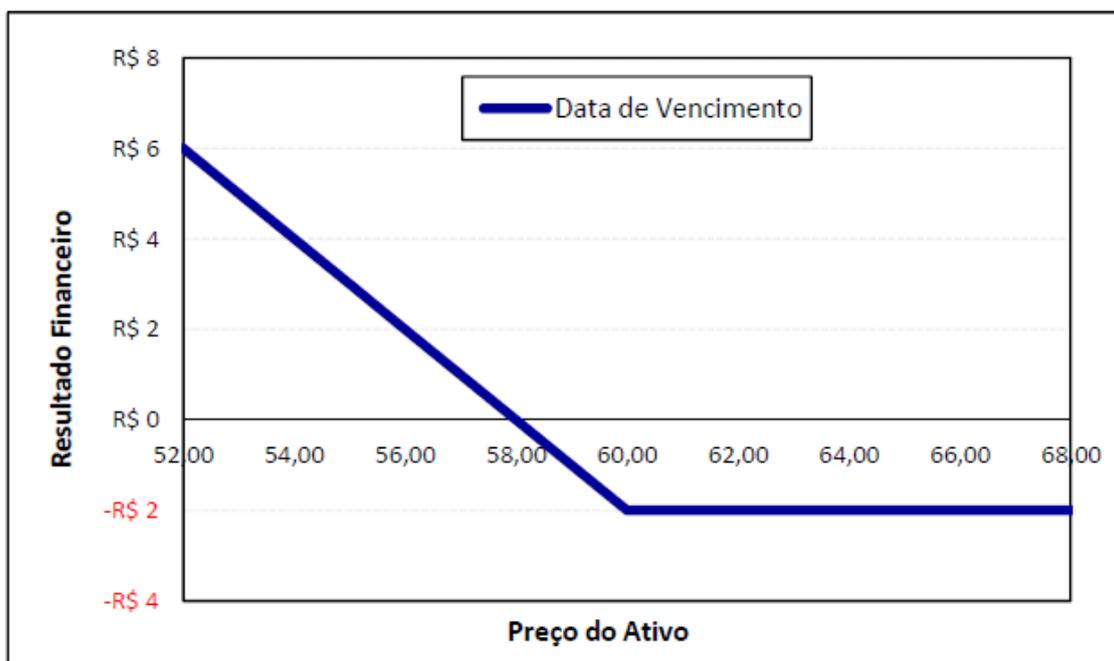


Figura 3 – Payoff para o titular de uma Put

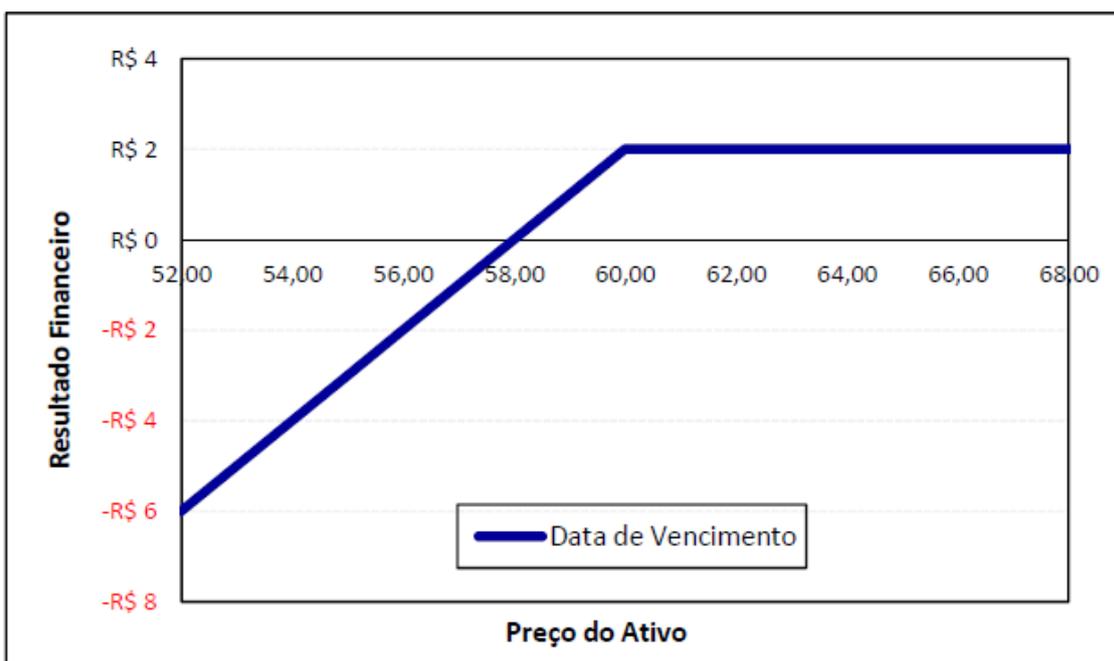


Figura 4 – Payoff para o Lançador de uma PUT

3.1.3

Opções Européias e Americanas

As opções são classificadas em Americanas e Européias de acordo com o seu tipo de exercício. Com as opções americanas o titular pode exercer o seu direito a qualquer momento enquanto que no estilo europeu o titular só pode exercer o seu direito na data exata de vencimento da opção.

3.2

Precificação de Opções

Um modelo de precificação de opções é um algoritmo matemático ou fórmula, por onde se estima o preço teórico ou justo de uma opção. Vários fatores influenciam o preço de uma opção, que é formado por dois componentes principais: o valor intrínseco e o valor extrínseco/temporal.

Duas variáveis que obviamente influenciam o preço de uma opção são o preço de exercício (strike) dessa opção e o preço do ativo-objeto. O valor intrínseco de uma call é zero quando o preço do ativo está abaixo do strike da opção e é igual à diferença entre o valor do ativo e o strike ($S - K$), quando o ativo está valendo mais que o strike (para uma put a relação é invertida). O valor intrínseco não é o único afetado por essas duas variáveis. Opções no dinheiro ou próximas do dinheiro (Apêndice 2) costumam ter maior valor extrínseco (valor temporal) comparado com opções muito dentro do dinheiro ou muito fora do dinheiro (Apêndice 2). Quanto mais próximo estiver o preço do ativo do strike da opção, maior as chances de que este ultrapassará o strike, dando assim valor extrínseco à opção antes do vencimento, e assim ganhando mais valor temporal, valor de risco.

A volatilidade do preço do ativo utilizada como entrada em um modelo de precificação, por não ser facilmente verificada, é o fator mais crítico para se determinar o preço de uma opção. Volatilidade é uma medida da variabilidade do preço de um ativo por unidade de tempo, o que varia muito de ativo para ativo. Certos ativos como ações de novas empresas e empresas pré-operacionais, costumam ter volatilidades muito altas, na casa dos 60% a.a, enquanto que ações

de empresas do setor de energia, por exemplo, são muito menos voláteis, apresentando volatilidade anualizada na casa dos 15% a.a.

A volatilidade calculada utilizando o histórico de preços do ativo é chamada de volatilidade histórica. O valor temporal da opção é altamente influenciado pela volatilidade do ativo, pois quão maior a velocidade de movimento do ativo, maior a probabilidade da opção ser exercida.

O valor temporal da opção também é influenciado pelo prazo para vencimento. Quanto maior o tempo para o vencimento, maior é o componente do valor temporal no preço da opção, pois mais tempo resta para o ativo variar de preço, logo maior a probabilidade, para ele alcançar e ultrapassar o exercício. O valor temporal é a variável especulativa da opção. Com o passar do tempo há um decaimento desse valor, valendo zero no vencimento da opção, pois não resta mais tempo para especulação.

O valor da taxa de juros livre de risco e dos dividendos esperados também têm alguma, porém pouca influência no preço das opções. Uma call se valoriza com um aumento nos juros, enquanto que uma put se desvaloriza. Essa relação pode ser entendida da seguinte forma: suponha um arbitrador⁵ que fica vendido⁶ no papel e compra calls dentro do dinheiro para se proteger de possíveis perdas financeiras devido a oscilações adversas que podem ocorrer. O financeiro que ele recebe dessa operação pode ser aplicado à taxa livre de risco recebendo juros, portanto quanto maior a taxa livre de risco, mais ele estará disposto a pagar na call para fazer essa troca e aplicar o dinheiro. A taxa de dividendos a receber da ação tem um efeito contrário no preço da opção. Esse mesmo arbitrador com essa mesma posição fica responsável em pagar os dividendos que a ação tiver direito quando o mesmo entregar o ativo que ele vendeu, portanto quanto mais dividendos essa ação receber, maior os custos para o arbitrador, que pagará então menos pela call para montar a operação. As taxas de Juros e os dividendos influenciam no custo de carregamento de um investimento e impactam nos preços das opções.

⁵ O arbitrador atua no mercado através de operações de arbitragem, que têm como finalidade tirar proveito do desequilíbrio do preço justo entre dois ativos, dois mercados ou de possíveis mudanças nessas diferenças no futuro.

⁶ Um investidor que fica vendido em um ativo está apostando na queda do preço deste ativo. Ele vende esse ativo, sem realmente possuí-lo, para recomprar em algum momento mais a frente.

O skew e a curtose da distribuição dos retornos do ativo também influenciam no preço das opções, são chamados de terceiro e quarto momento e definem certos aspectos do formato da curva de distribuição de retornos. Ao contrário da distribuição normal, uma distribuição com skew negativo tem uma cauda esquerda maior e uma cauda direita menor com o topo inclinado para a direita (exatamente o inverso para skew positivo) (Figura 5). Uma ação com skew negativo seria aquela que demonstra quedas bruscas ocasionais contra frequentes, mas pequenas, altas. Portanto o skew da distribuição dos retornos da ação afeta a expectativa de retorno final de uma operação estruturada⁷ com opções, sendo então relevante ao preço da opção.



Figura 5 – Distribuição Normal e Distribuições com Skew.

A curtose é uma medida de dispersão que caracteriza o "achatamento" da curva da função de distribuição. Uma distribuição platicúrtica, com curtose negativa, é mais achatada que a distribuição normal, sendo mais larga no centro e apresentando caudas mais curtas. Uma distribuição leptocúrtica, com curtose positiva, é mais alta, fina e afunilada que a distribuição normal, tendo também caudas mais longas (Figura 6).



Figura 6 – Distribuições com Curtose.

De acordo com Natenberg (1994), a distribuição de retornos comumente observada em ações é leptocúrtica, apresentando mais retornos extremos, tanto positivos quanto negativos, do que esperados por uma distribuição normal e

⁷ Operações Estruturadas com opções são estratégias que incluem uma combinação de dois ou mais contratos de opções.

também mais variações pequenas, onde os preços variam muito pouco, do que variações moderadas. O efeito dessa característica no preço das opções pode ser observada no denominado sorriso de volatilidade (Figura 7). O sorriso da volatilidade mostra que opções bem fora do dinheiro e bem dentro do dinheiro acabam sendo negociadas a preços acima do esperado por modelos de precificação que consideram uma distribuição normal, logo apresentando uma maior volatilidade implícita.



Figura 7 – “Sorriso” da volatilidade

O primeiro momento da distribuição de retornos do ativo, a média, também é importante na precificação de uma opção. Uma média diferente de zero pode ser vista como uma tendência de longo prazo do ativo. Numa tendência positiva, calls podem valer mais e puts menos, devido à expectativa de se manter o movimento direcional do ativo, apreciando o valor da call e depreciando o valor da put. O inverso é válido para uma tendência negativa.

Ciclos sazonais e eventos também têm grande impacto na precificação de opções. A volatilidade das opções tende a aumentar quando há a expectativa de algum anúncio ou notícia específica que possa influenciar o ativo. Em momentos como anúncio de balanços e resultados a volatilidade pode aumentar e como isso acontece trimestralmente, pode ser considerado um ciclo sazonal na volatilidade de ações. Muitos outros ciclos podem ser verificados, principalmente em

“commodities”. Rumores e expectativas de anúncio de importantes notícias como fusões, aquisições, novas descobertas, lançamentos de produtos, decisões judiciais entre outras podem gerar fortes picos na volatilidade, influenciando também nos preços das opções.

Um grande desafio para os participantes do mercado de opções é encontrar o preço justo para se pagar em uma opção que está sendo negociada no mercado. O artigo de Black e Scholes (1973) apresentou um modelo matemático para calcular o valor justo para o prêmio de uma opção europeia sem dividendos. De acordo com Black e Scholes (1973) o prêmio de uma opção de compra ou venda é função do preço do ativo-objeto, do preço de exercício, do tempo até o exercício, da taxa de juros livre de risco e da volatilidade do ativo-objeto até o exercício e se dá através da Eq. (3) para uma call e da Eq. (4) para uma put :

$$c = SN(d_1) - Xe^{-rt}N(d_2) \quad \text{Eq. (3)}$$

$$p = Xe^{-rt}N(-d_2) - SN(-d_1) \quad \text{Eq. (4)}$$

$$d_1 = \frac{\ln(S/X) + \left(r + \frac{s^2}{2}\right)T}{s\sqrt{T}} \quad \text{Eq. (5)}$$

$$d_2 = d_1 - s\sqrt{T} \quad \text{Eq. (6)}$$

Onde:

c é o preço justo da call;

p é o preço justo da put;

S é o preço atual do ativo-objeto;

X é o preço de exercício da opção;

T é o período de tempo até a data de vencimento da opção;

r é a taxa de juros livre de risco no período;

s é a volatilidade do ativo-objeto no período;

N(d) é a função de densidade acumulada da distribuição Normal padronizada.

Este modelo segue uma série de premissas que não necessariamente se verificam na prática: as opções são europeias; o preço da ação segue um movimento geométrico browniano; os retornos do ativo-objeto seguem uma distribuição normal; vendas a descoberto são permitidas; não há pagamentos de

dividendos durante a existência do derivativo; a taxa de juros livre de risco é constante e a mesma para todos os prazos de maturação; a volatilidade do ativo-objeto é um valor conhecido e constante ao longo do tempo; assumindo-se ausência de arbitragens obtemos o preço justo dos derivativos.

Merton (1973) e Amin (1993) relaxam a premissa de normalidade dos retornos do ativo-objeto. Neste trabalho os autores consideram que o retorno do ativo-objeto segue um processo de saltos em conjunto com um processo de normalidade.

Hull e White (1987), Amin (1993) e Heston (1993) relaxam a premissa de volatilidade constante até o prazo de vencimento da opção. O comportamento da volatilidade é modelado por um processo estocástico independente do preço do ativo-objeto. Os resultados encontrados indicam que o modelo de Black e Scholes (1973) avalia bem opções no dinheiro e subestima o preço de opções dentro do dinheiro e fora do dinheiro.

Apesar da grande quantidade de modelos, nenhum deles consegue descrever completamente a complexidade do apreçamento de opções, diante disso o modelo de Black e Scholes (1973), por sua simplicidade, ainda continua a ser o mais utilizado no apreçamento de opções no mercado brasileiro e é o utilizado neste trabalho.

Em um mercado eficiente e racional podemos considerar todas as opções como européias, por ser desvantajoso exercer uma opção antes do seu vencimento. O motivo que a torna desvantajosa é que um titular de uma Call não desembolsaria antecipadamente o financeiro para exercer seu direito de compra do ativo quando ele pode deixar o dinheiro rendendo a taxa livre de risco até a data do vencimento. No caso em que o investidor decida se desfazer dessa “posição” antecipadamente, ele pode vender a sua Call no mercado secundário por um preço superior ao da Eq. (1), devido ao valor extrínseco/temporal da opção, o que é mais vantajoso e menos custoso que exercer o direito de compra e vender o ativo em seguida. Todas as Puts são, por convenção, européias no mercado brasileiro.

No mercado brasileiro pode-se considerar que não há pagamentos de dividendos durante a existência da opção devido a um mecanismo operacional da BOVESPA que diminui o valor do preço de exercício da opção pelo valor do dividendo no momento em que este é anunciado (Bessada, Barbedo e Araujo, 2007).

3.3

Volatilidade

No Mercado existem várias visões e tipos diferentes de volatilidade: Volatilidade futura, volatilidade histórica, volatilidade prevista, volatilidade implícita entre outras.

A volatilidade histórica é a realizada por um ativo financeiro ao longo de um determinado período de tempo. Esta medida é muito simples de calcular, basta tomarmos o desvio padrão dos retornos históricos do ativo, normalmente dos retornos logarítmicos.

A volatilidade futura é o que todo investidor gostaria de saber, é a volatilidade que melhor descreve a distribuição futura dos preços do ativo-objeto. É esse valor que se procura fornecer como entrada em um modelo de precificação de opções. Conhecendo-se a volatilidade futura, sabe-se a probabilidade do ativo alcançar certos níveis de preço no vencimento.

Mesmo não conhecendo o futuro, se um investidor pretende chegar a um preço justo para comprar/vender uma opção, ele precisará encontrar um valor para tal volatilidade futura. No mercado financeiro, como em outras áreas de conhecimento, um bom ponto de partida é olhar para os dados históricos e verificar a volatilidade num determinado período passado. Se nos últimos anos a volatilidade de um ativo nunca foi maior que 30% ou menor que 15%, não seria inteligente esperar uma volatilidade de 40% ou de 8%, o que não quer dizer que seria impossível de acontecer.

Da mesma forma que existem sistemas que tentam prever movimentos futuros no preço de um ativo, existem sistemas que procuram prever a volatilidade futura de ativos. Essa previsão costuma ser feita para um intervalo de tempo proporcional ao tempo restante para o vencimento da opção analisada. A previsão da volatilidade futura é um dos objetivos desse trabalho.

As volatilidades futuras, históricas e previstas, estão associadas ao ativo-objeto, porém existe outra interpretação de volatilidade que está associada a uma certa opção, a volatilidade implícita. Suponha-se que certo ativo está valendo R\$98,50 com taxa de juros de 8%a.a. e que a expectativa de volatilidade futura para a call de preço de exercício R\$105,00, com três meses para o vencimento, seja de 16%a.a. Para se saber o preço teórico dessa opção, entra-se com os valores

indicados num modelo de precificação, no caso o Black & Scholes, e encontra-se o valor de R\$0,96, enquanto que essa opção está sendo negociada a R\$1,34 a mercado.

Para explicar tal discrepância no preço, considerando que todos no mercado estão usando o mesmo modelo de precificação, analisa-se as variáveis de entrada, verificando que a única que pode divergir é a volatilidade, por não ser facilmente verificada, enquanto que as outras variáveis são triviais. Usando-se um método matemático como Newton-Raphson (Apêndice 1), pode-se verificar que para precificar a opção em R\$1,34 o mercado está considerando uma volatilidade de 18,5%, chamada de volatilidade implícita da opção.

3.4

Modelo EWMA

No cálculo da volatilidade pela média histórica, são atribuídos os mesmos pesos para todos os retornos da série histórica que compõem a amostra, dificultando a detecção, com rapidez, de mudanças de tendência de comportamento da volatilidade, uma vez que atribui a mesma importância tanto para os retornos mais recentes quanto para os mais antigos.

O modelo EWMA (*exponentially weighted moving average*) é um método de suavização exponencial que utiliza médias móveis (Wikipedia, Moving Average). Com ele detecta-se com mais rapidez as mudanças nas condições do mercado. Em vez de aplicar o mesmo peso para todos os dados observados da amostra, são atribuídos pesos maiores às observações mais recentes. O peso alocado para cada um dos dados da amostra é uma função do fator de decaimento ou grau de suavização (λ).

A expressão para o cálculo da volatilidade para o instante t , utilizando-se o EWMA, pode ser representada pela Eq. (7):

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) r_{t-1}^2, \quad 0 \leq \lambda \leq 1 \quad \text{Eq. (7)}$$

onde:

σ^2 é variância dos retornos no instante t

r_{t-1}^2 é o quadrado do retorno observado no instante $t-1$

O fator λ determina a taxa em que os pesos das observações passadas decaem à medida que se tornam mais distantes. Quanto menor o λ mais importantes são as novas observações e mais descontadas são as observações mais antigas.

Pode-se observar na equação acima que o cálculo da volatilidade com base no EWMA se dá através da combinação de dois elementos: o primeiro é a estimativa da variância do dia anterior, que recebe o peso igual a λ ; e o segundo é o quadrado do retorno observado no dia anterior, que recebe o peso igual a $(1-\lambda)$. A volatilidade que desejamos obter é o desvio padrão, ou seja, a raiz quadrada da variância.

Segundo JP Morgan (1994), no modelo EWMA, a volatilidade estimada reage mais rapidamente aos movimentos extremos dos mercados, uma vez que as observações mais recentes recebem maior peso. Após a ocorrência desses movimentos, a volatilidade diminui gradualmente à medida que os pesos atribuídos a esses movimentos decaem. JP Morgan (1994), após testes empíricos com diversas classes de ativos, recomenda a utilização de $\lambda=0,94$.

3.5

Modelo ARCH

Os modelos mais simples não consideravam o fato de a volatilidade variar com o tempo. Engle (1982) desenvolveu um modelo denominado ARCH (*Autoregressive Conditional Heterocedasticity*) que considera ser a variância heterocedástica, ou seja, não é constante ao longo do tempo. Neste modelo, a variância condicional é uma função linear do quadrado das inovações passadas. Assim sendo, o modelo ARCH (q) pode ser representado pela Eq. (8) e tem as restrições da Eq. (9):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2, \quad \text{Eq. (8)}$$

Onde:

$$R_t = c + \varepsilon_t$$

$$E(\varepsilon_t) = 0, E(\varepsilon_t^2) = 1, \quad \text{Eq. (9)}$$

$$E(\varepsilon_t^2 \setminus I_{t-1}) = \sigma_t^2, I_{t-1} \text{ (informações disponíveis em } t-1),$$

R_t é o retorno em t .

c é uma constante.

ε_t é o erro residual do modelo em t .

α são restrições paramétricas.

Para esse modelo ser bem definido e a variância condicional ser positiva, as restrições paramétricas devem satisfazer $\alpha_0 > 0$ e $\alpha_i > 0, i = 1, 2, \dots, p$

3.6

Modelo GARCH

Uma importante extensão do modelo ARCH é a sua versão generalizada proposta por Bollerslev (1986), denominada GARCH (*Generalized Autoregressive Conditional Heterocedasticity*). Neste modelo, a função linear da variância condicional inclui também variâncias passadas. Assim sendo, a volatilidade dos retornos depende dos quadrados dos erros anteriores e também de sua própria variância em períodos anteriores. A variância é dada pela Eq. (10):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad \text{Eq. (10)}$$

Onde:

ε_{t-i} são os erros anteriores.

σ_{t-j}^2 são as variâncias em períodos anteriores

β_j é uma restrição paramétrica

onde as restrições são dadas por: $\alpha_i > 0, i = 1, 2, \dots, q$; $\beta_j > 0, j = 1, 2, \dots, p$ e $\alpha_i + \beta_j < 1$. Assim sendo, σ_t^2 segue um modelo GARCH (p,q), onde p representa ordem do componente GARCH e q a ordem do componente ARCH.

De acordo com Enders (2004), o modelo GARCH (1,1) é a versão mais simples e mais utilizada em séries financeiras. Supondo-se que os erros são normalmente distribuídos, a variância é dada pela Eq. (11).

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad \text{Eq. (11)}$$

O coeficiente α_1 mede a extensão em que um choque no retorno hoje afeta a volatilidade do retorno de amanhã. A soma $\alpha_1 + \beta_1$ revela a medida de persistência da volatilidade, ou seja, a taxa que reflete como o impacto de um choque no retorno hoje se propaga ao longo do tempo sobre a volatilidade dos retornos futuros. Isso mostra que a alta persistência do choque enfraquecerá gradualmente.

Dado que a volatilidade é uma variável que tende a reverter à média de longo prazo, períodos de volatilidades mais altas tendem a ser seguidos de períodos com volatilidades mais baixas e vice-versa. O GARCH(1,1) simplesmente adiciona uma variância média de longo prazo, para a qual esperamos que ela reverta, na equação do EWMA. Para simplificar podemos escrever a Eq. (11) da forma a seguir na Eq. (12):

$$\sigma_t^2 = \gamma V + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \quad \text{Eq. (12)}$$

Onde

$\gamma + \alpha + \beta = 1$ e V é a variância de longo prazo. Se $\gamma=0$, $\beta=\lambda$ e $\alpha=1-\lambda$, retornamos ao modelo EWMA.

3.7

Redes Neurais Artificiais

Redes Neurais Artificiais (RNAs) são uma forma de computação não-algorítmica que relembra a estrutura do cérebro humano e por não ser baseada em

regras ou programas se constitui em uma excelente alternativa a computação algorítmica convencional, (Haykin 2000).

As RNAs são capazes de reconhecer padrões, extrair regularidades e detectar relações subjacentes em um conjunto de dados aparentemente desconexos, apresentando como uma de suas características a previsão de sistemas não lineares.

RNAs são sistemas paralelos distribuídos e compostos por unidades de processamento (neurônios) que calculam determinadas funções matemáticas. Tais unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões (sinapses). Estas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede.

O funcionamento destas redes é inspirado na estrutura do cérebro humano, onde a experiência é a fonte do conhecimento adquirido e esse conhecimento é armazenado nas sinapses.

A habilidade das RNAs em realizar mapeamentos não-lineares entre suas entradas e saídas têm dado destaque aos problemas de reconhecimento de padrões e na modelagem de sistemas complexos.

Algumas áreas de aplicação são: robótica, controle de processos, classificação de dados, reconhecimento de padrões, previsão de séries temporais, análise de imagens, avaliação de riscos de financiamento, filtro de ruídos eletrônicos, identificação de reservas de petróleo, etc.

Três conceitos básicos caracterizam os diferentes tipos de RNAs: O modelo do neurônio artificial, sua estrutura de interconexão (topologia) e a regra de aprendizado. A seguir apresentaremos esses conceitos.

3.7.1

O Neurônio Artificial

A rede neural artificial é formada por unidades elementares de processamento, denominadas neurônios artificiais. Uma grande rede neural artificial pode ter centenas ou milhares de unidades de processamento; já o cérebro de um mamífero pode ter muitos bilhões de neurônios (Kovács, 2006).

O neurônio artificial, inspirado no biológico (Figura), possui um conjunto de entradas w_m (dendritos) e uma saída y_k (axônio). "As entradas são ponderadas por pesos sinápticos w_{km} (sinapses), que determinam o efeito da entrada x_m sobre o processador k . Estas entradas ponderadas são somadas fornecendo o potencial interno do processador v_k (na Figura 8). A saída ou estado de ativação y_k , do elemento processador k é finalmente calculada através de uma função de ativação φ " (Velasco, 2007).

O estado de ativação pode então ser definido pela Eq. (13):

$$y_k = \varphi \left(\sum_{j=1}^m x_j * w_{kj} + b_k \right) \quad \text{Eq. (13)}$$

Onde m é o número de entradas do neurônio k e b_k é um termo de polarização do neurônio (viés).

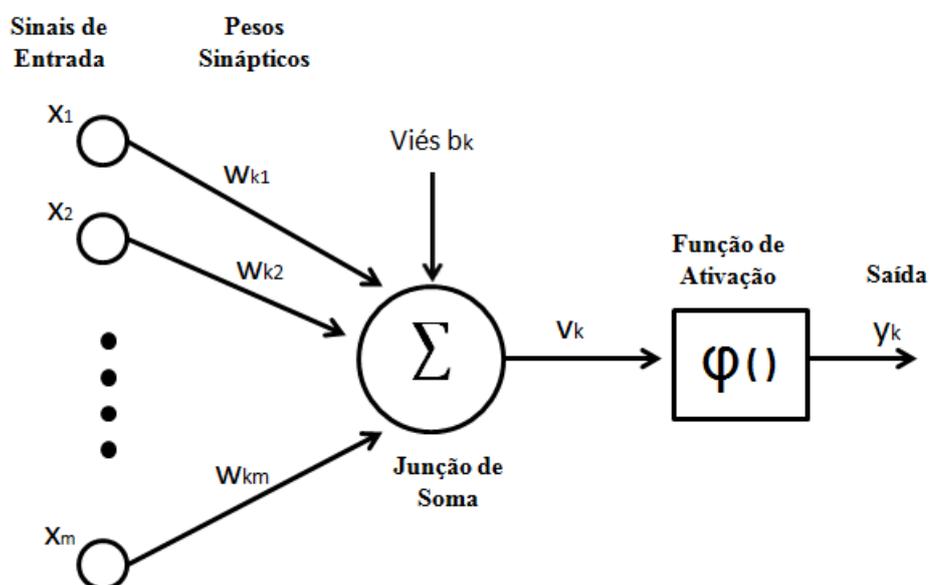


Figura 8 – Esquema de um neurônio artificial

3.7.2

Funções de Ativação

A função de ativação ϕ deve simular as características não lineares do neurônio biológico, ela processa o sinal v_k (ver Figura 8) e produz a saída final do neurônio, y_k . As funções mais utilizadas são:

- Função Linear (Figura 9), Eq. (14):

$$F(x) = a * x \quad \text{Eq. (14)}$$

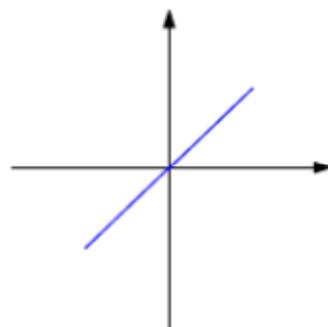


Figura 9 – Função Linear

- Função Degrau (Figura 10):

É uma equação utilizada para valores binários e é na forma da Eq. (15):

$$F(x) = 1, \text{ se } x > 0 \quad \text{Eq. (15)}$$

$$F(x) = 0, \text{ se } x \leq 0$$

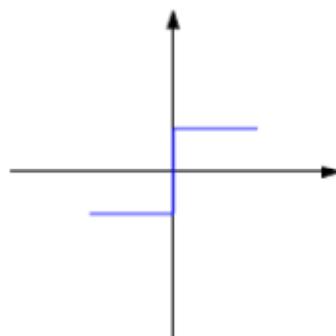


Figura 10 – Função Degrau

- Função Sigmóide/Logística (Figura 11):

É uma função contínua que permite a transição gradual entre os dois estados, variando entre 0 e 1, Eq. (16).

$$f(x) = \frac{1}{1 + e^{-\lambda x}} \quad \text{Eq. (16)}$$

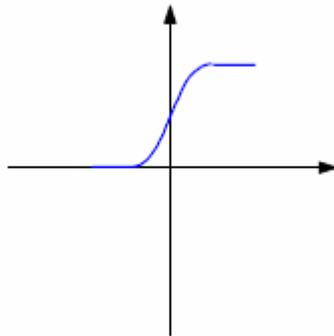


Figura 11 – Função Sigmoide

- Função Tangente Hiperbólica (Figura 12)

É uma função sigmoide que varia entre -1 e 1, Eq. (17).

$$\tanh(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}} \quad \text{Eq. (17)}$$

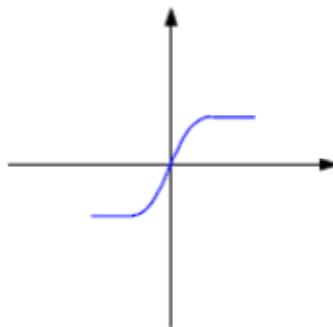


Figura 12 – Função Tangente Hiperbólica

3.7.2

Arquiteturas de Rede

Em uma RNA Feedforward os neurônios são organizados em forma de camadas, existindo sempre uma camada de entrada e uma camada de saída da rede, sendo chamada de perceptron, podendo também existir redes com camadas “escondidas” as multi layer perceptron (Figura 13). Nessas redes o sentido de propagação do sinal é único, sendo sempre da entrada para a saída da rede, não

havendo nem realimentação das camadas seguintes para as anteriores, nem também ligações entre neurônios de uma mesma camada.

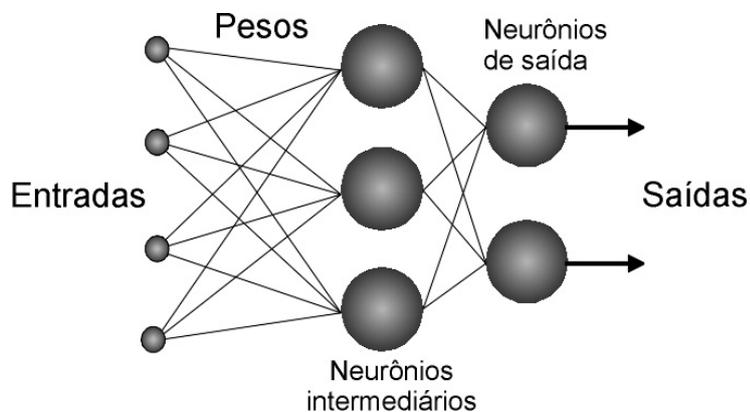


Figura 13 – Arquitetura de um multi layer perceptron

De acordo com Kovács (2006), as redes recorrentes chamadas de redes com memória (Figura 14), não possuem estrutura rígida e seus neurônios têm liberdade de se ligar com qualquer outro neurônio, mesmo pertencente a uma camada anterior, ou até a si próprio. Devido à realimentação suas saídas são determinadas pelas entradas atuais e pelas saídas anteriores.

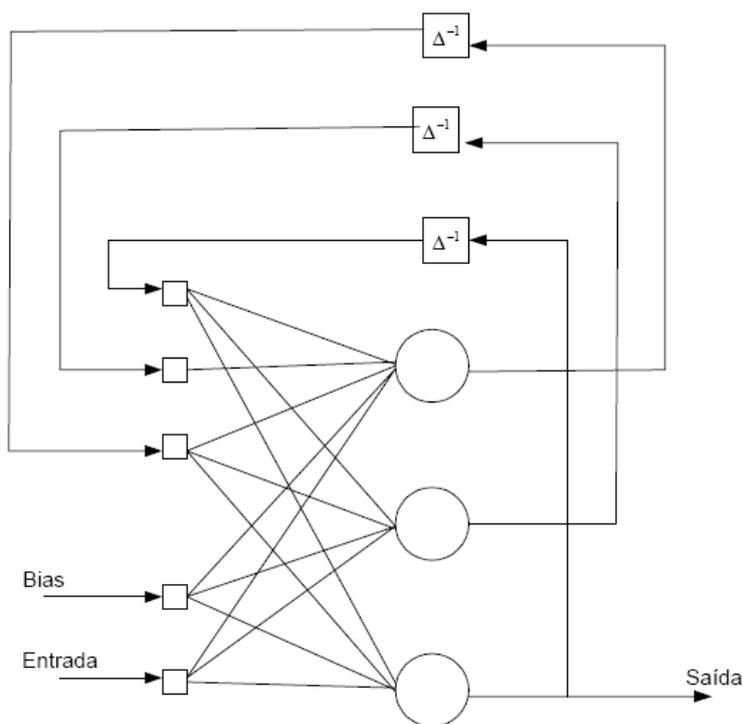


Figura 14 – Exemplo de arquitetura de uma rede recorrente

3.7.3

Aprendizagem

A propriedade mais importante das redes neurais é a habilidade de aprender de seu ambiente e com isso melhorar seu desempenho. Isso é feito através de um processo interativo de ajustes aplicado a seus pesos, o treinamento. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas. Há dois tipos de treinamento: supervisionado e não-supervisionado.

As redes neurais são treinadas para aprender a partir dos dados de entrada. Assim como o cérebro humano, elas aprendem a partir de experiências e não através de uma programação pré-concebida. O conjunto de treinamento deve ser formado a partir de dados históricos. Deve também ser separado um conjunto de testes que não é apresentado à rede durante a fase de treinamento, para que a capacidade de generalização dessa rede possa ser avaliada.

O treinamento supervisionado é baseado em um conjunto de padrões de pares (entrada, saída) que são apresentados a rede. A partir da entrada, a rede realiza o processamento e a saída obtida é comparada a saída alvo. Caso não sejam iguais é aplicado um processo de ajuste dos pesos sinápticos até obter um resultado satisfatório.

Como já mencionado, a rede deve ser capaz de generalizar, devendo-se tomar cuidado para que ela não seja super treinada (*overfitting*) e acabe memorizando os exemplos apresentados, perdendo assim a capacidade de reconhecer padrões (generalização).

"Backpropagation é o algoritmo para treinamento de Redes Multi Camadas mais difundido" (Braga, Carvalho, Ludermir, 1999). Baseia-se no Aprendizado Supervisionado por Correção de Erros, e é formado de duas partes principais: 1° - Propagação: Depois de apresentado o padrão de entrada, a resposta de uma unidade é propagada como entrada para as unidades na camada seguinte até a camada de saída, onde é obtida a resposta da rede e o erro é calculado. 2° - Retropropagação: Depois de calculado o erro, desde a camada de saída até a camada de entrada, são feitas alterações nos pesos sinápticos a fim de que esse erro seja minimizado na próxima iteração.

Resumo do Algoritmo:

1-Inicialização: Inicializam-se os pesos sinápticos e os viés⁸ (*bias*) aleatoriamente, com valores no intervalo [-1;1];

2-Apresentação dos Exemplos de Treinamento: Treinamento "on-line": Para cada exemplo do conjunto de treinamento, efetua-se os passos 3 e 4. Treinamento "em batch": Para cada "época" do conjunto de treinamento, efetua-se os passos 3 e 4.

3-Propagação: Depois de apresentado o exemplo do conjunto de treinamento $T=\{(x(n),d(n))\}$, sendo $x(n)$ a entrada apresentada à rede e $d(n)$ a saída desejada, calcula-se o valor da ativação v_k e a saída y_k para cada unidade da rede, da forma da Eq. (18):

$$y_k = \varphi \left(\sum_{j=1}^m x_j * w_{kj} + b_k \right) \quad \text{Eq. (18)}$$

Onde (ver Figura 8):

x são as entradas do neurônio

w são os pesos sinápticos do neurônio k

b é a entrada de viés do neurônio k

para o cálculo do valor da ativação e $\varphi(v_k)$ para o cálculo da saída y_k da unidade/neurônio k , onde φ é a função de ativação desejada. Utiliza-se a saída das unidades de uma camada como entrada para a seguinte, até a última camada. A saída das unidades da última camada será a resposta da rede.

4-Cálculo do Sinal de Erro: Sendo a saída $y_j = O_j(n)$, $O_j(n)$ é a saída obtida como resposta final da rede para a amostra j ; calcula-se o sinal de erro através da Eq. (19):

$$e_j(n) = d_j(n) - O_j(n) \quad \text{Eq. (19)}$$

⁸ Vies (*bias*) e uma entrada ajustavel do neuronio que 'e somado ao valor acumulado das outras entradas. Serve para aumentar os graus de liberdade, permitindo uma melhor adaptação, por parte da rede neural, ao conhecimento à ela fornecido.

onde $d_j(n)$ é a saída desejada com resposta para cada unidade na interação (n) . Este sinal de erro é utilizado para computar os valores dos erros das camadas anteriores e fazer as correções necessárias nos pesos sinápticos.

5-Retropropagação: Calcula-se os erros locais, δ , para cada unidade, desde a camada de saída até a de entrada. O gradiente local é definido pela Eq. (20):

$$\delta(n) = e_j(n)O_j(n)(1 - O_j(n)) \quad \text{Eq. (20)}$$

para a unidade da camada de saída ou como na Eq. (21):

$$\delta(n) = O_j(n)(1 - O_j(n)) \sum \delta_k w_{jk} \quad \text{Eq. (21)}$$

para as unidades das demais camadas.

Onde:

$O_j(1-O_j)$ é a função de ativação diferenciada em função do argumento, valor de ativação.

δ_k é o erro das unidades da camada anterior conectadas a unidade j ;

w_{jk} são os pesos das conexões com a camada anterior.

Após o cálculo dos erros de cada unidade, calcula-se o ajuste dos pesos de cada conexão segundo a regra delta generalizada na Eq. (22), e atualizam-se os pesos sinápticos, Eq. (23):

$$\Delta w_{kj}(n + 1) = \alpha w_{kj}(n) + \eta \delta_j y_j \quad \text{Eq. (22)}$$

$$w(n + 1) = w(n) + \Delta w_{kj}(n) \quad \text{Eq. (23)}$$

onde:

α é a constante de momentum, quando $\alpha = 0$, esta função funciona como a regra delta comum;

η é a taxa de aprendizado;

δ_j é o erro da unidade;

y_j é a saída produzida pela unidade j ;

6-Interação: Refaz-se os itens 3, 4 e 5 referentes à propagação, cálculo do erro e retropropagação, apresentando-se outros estímulos de entrada, até que sejam satisfeitas as condições de treinamento, as quais podem ser: O erro da rede está baixo, sendo pouco alterado durante o treinamento ou o número máximo de ciclos de treinamento foi alcançado.

3.8

Métodos de Otimização

Otimizar é o processo de ajuste das entradas às características de um problema, sistema ou processo matemático, a fim de achar uma saída/resultado máximo ou mínimo. As entradas são as variáveis do problema, o sistema é uma função-objetivo ou função-de-custo e a saída é o custo ou objetivo.

Matematicamente o processo de busca por raízes de uma função procura por zeros da função, enquanto que o processo de otimização procura por zeros da derivada da função, incluindo assim um passo a mais ao processo. Muitas vezes essa derivada não existe ou é muito difícil de ser encontrada, o que torna o problema muito complicado. Outra dificuldade da otimização é definir se um mínimo encontrado é o “menor mínimo” (mínimo global) ou um ponto sub-ótimo, um mínimo local. Encontrar um mínimo de uma função é um processo complicado. Para simplificar, muitas vezes considera-se funções não-lineares como lineares numa pequena região, ou restringe-se a busca/otimização em uma região delimitada somente.

“Tentativa e erro” é o método mais simples de otimização, onde normalmente se conhece as variáveis que afetam o problema, mas não se entende muito do processo e de como essas variáveis são tratadas para se atingir a saída. “A descoberta da Penicilina (antibiótico) surgiu desse método de otimização” (Haupt 2004).

Se a otimização tiver somente uma variável, ela é tida como um problema unidimensional; com mais de uma variável se torna um problema multidimensional. Quanto mais dimensões, mais difícil é a otimização. “Muitas otimizações multidimensionais são solucionadas por diversas otimizações unidimensionais” (Rao 1996).

Uma otimização é dinâmica quando a sua saída é função do tempo, enquanto que uma otimização estática independe do tempo. Para exemplificar isso pode-se pensar no problema de achar a rota mais rápida para ir de casa ao trabalho e otimizar minimizando o tempo do percurso. Pode-se considerar esse problema como estático e achar a solução somente com um mapa na mão ou, considerá-lo como dinâmico, quando então seriam necessárias variáveis como horário, condições do trânsito e acidentes, condições do tempo entre outras.

Uma otimização pode ser discreta ou contínua, dependendo de suas variáveis de entrada. Variáveis discretas têm um número finito de valores possíveis, enquanto que variáveis contínuas têm infinitos valores. Otimização com variáveis discretas são conhecidas como otimizações combinatórias, pois a solução ótima consiste na combinação dos possíveis valores (finitos) das variáveis de entrada. Se estiver tentando achar o mínimo de uma função $f(x)$ que percorre uma linha, por exemplo, tem-se um problema contínuo.

Variáveis muitas vezes têm limites e restrições. A otimização com restrições incorpora as igualdades e as inequações que restringem os valores das variáveis na função-custo. Algoritmos de otimização trabalham melhor com funções sem restrição e para isso, em alguns casos, fazem transformações nas variáveis de entrada para transformá-las em variáveis sem-restrição. O problema de minimizar a função $f(x)$ sobre o intervalo $-1 \leq x \leq 1$, por exemplo, pode ser transformado em um problema de minimizar $f(\sin(u))$ onde $x = \sin(u)$ para qualquer valor de u . Uma otimização com restrições lineares é chamada de Programação Linear.

Alguns algoritmos como os “*minimum-seekers*” tentam minimizar a função-custo, iniciando de algum ponto escolhido da superfície de custo e seguindo certos passos matemáticos pré-definidos, para achar a direção por onde “caminhar” para chegar a um ponto mais baixo da superfície. “Esses algoritmos tendem a ser rápidos, mas ficam presos muito facilmente em mínimos locais” (Fletcher, 2000). Por outro lado os “*random-methods*” utilizam cálculos probabilísticos para encontrar conjuntos de variáveis. Eles são mais vagarosos e normalmente exigem maior potência computacional, mas “são melhores para resolver problemas com superfícies de custo irregulares ou com primeiras derivadas descontínuas” (White 1971).

3.9

Algoritmos Genéticos

Algoritmos Genéticos (GAs) são métodos de busca e otimização, que têm sua inspiração nos conceitos da teoria de seleção natural das espécies proposta por Darwin e na genética (Sivanandam, Deepa 2008). Os sistemas desenvolvidos a partir desse princípio “são utilizados para procurar soluções de problemas complexos ou com espaço de busca muito grande, o que os tornam problemas de difícil modelagem e solução quando se aplicam métodos de otimização convencionais” (Lazo, 2000).

“GAs têm sido aplicados a diversos problemas de otimização, tais como: Otimização de Funções Matemáticas, Otimização Combinatorial, Otimização de Planejamento, Problema do Caixeiro Viajante, Problema de Otimização de Rota de Veículos, Otimização de Layout de Circuitos, Otimização de Distribuição, Otimização em Negócios e Síntese de Circuitos Eletrônicos.” (Velasco, 2007)

Estes algoritmos são baseados nos processos genéticos de organismos biológicos para procurar soluções ótimas. Primeiro codifica-se as possíveis soluções de um problema em uma estrutura chamada de cromossoma, sendo este composto por uma cadeia de bits ou caracteres. Estes cromossomas representam indivíduos que são evoluídos por várias gerações, de acordo com o princípio da seleção natural e sobrevivência do mais apto.

Durante o processo evolucionário os cromossomas são submetidos à avaliação, seleção, cruzamento e mutação, sendo que após vários ciclos, a população terá indivíduos evoluídos, mais aptos e cada um destes indivíduos representam soluções ao problema proposto.

Um Algoritmo Genético pode ser resumido pelas seguintes etapas (Figura 15):

- 1) INICIALIZAÇÃO: Gera aleatoriamente uma população de N cromossomas (soluções convenientes para o problema)
- 2) AVALIAÇÃO: Avalia cada cromossoma através da função objetivo (grau de adaptação)
- 3) NOVA POPULAÇÃO: Cria uma nova população com os seguintes passos:

- SELEÇÃO: Seleciona cromossomas (pais) dentre a população de acordo com seu grau de adaptação, privilegiando o mais apto.
 - CRUZAMENTO: Cruza os pais para formar novos descendentes (filhos).
 - MUTAÇÃO: Altera alguns cromossomas da nova geração.
 - ACEITAÇÃO: Introduz-se a nova geração na atual população
- 4) SUBSTITUIÇÃO: Utiliza-se a nova população gerada para outro funcionamento do algoritmo.
 - 5) TESTE: Se o critério de fim é alcançado, termina-se. Caso contrário volta-se ao passo dois.

Os critérios de fim podem ser o descobrimento de uma boa solução, um número definido de gerações G ou um tempo limite t .

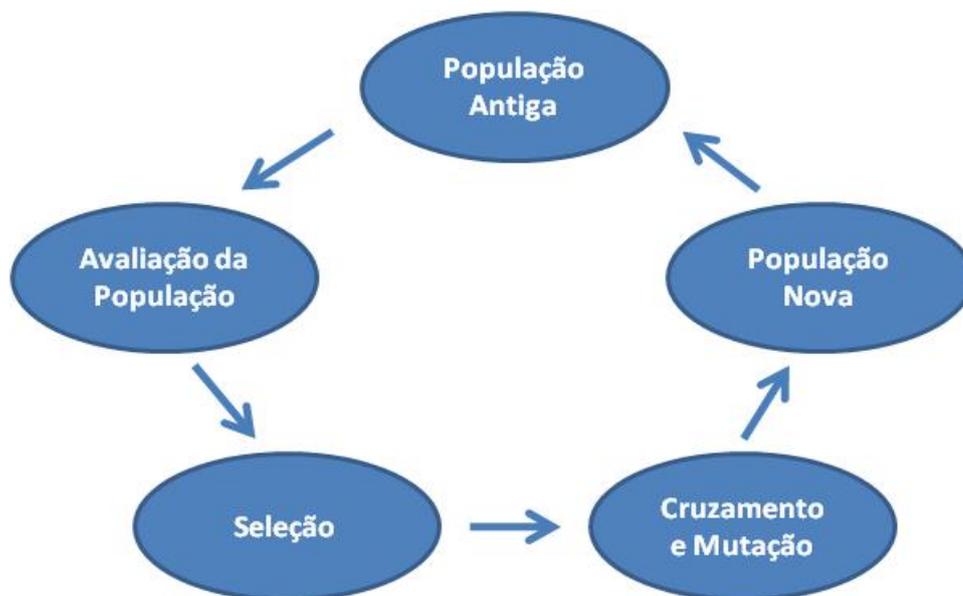


Figura 15 – Etapas de um Algoritmo Genético

O objetivo dos operadores genéticos é transformar a população através de sucessivas gerações até chegar a um resultado satisfatório. Eles são necessários para que a população se diversifique. O cruzamento e a mutação são os principais operadores genéticos.

Indivíduos selecionados (e reproduzidos na população seguinte) são recombinados através do operador de cruzamento (com uma certa probabilidade). Pares de genitores são escolhidos aleatoriamente da população, baseado na aptidão, e novos indivíduos são criados a partir da troca do material genético. Os

descendentes serão diferentes de seus pais, mas com características genéticas de ambos os genitores (Figura 16).

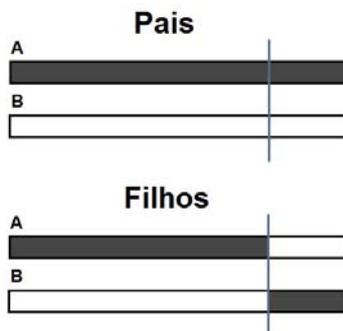


Figura 16 – Cruzamento Genético

A mutação é uma alteração aleatória e ocasional do valor de uma posição qualquer do *cromossoma*. Esta alteração ocorre de acordo com uma probabilidade prefixada e, por exemplo, no caso de um cromossoma binário, poderia significar a mudança de "1" para "0" ou de "0" para "1" (Figura 17).



Figura 17 – Mutação Genética

As técnicas de busca e otimização tradicionais iniciam-se com um único candidato que, iterativamente, é manipulado utilizando algumas heurísticas (estáticas) diretamente associadas ao problema a ser solucionado. Por outro lado, “as técnicas de computação evolucionária operam sobre uma população de candidatos em paralelo” (Holland, 1992). Assim, elas podem fazer a busca em diferentes áreas do espaço de solução, alocando um número de membros apropriado para a busca em várias regiões.

Os Algoritmos Genéticos (AGs) diferem dos métodos tradicionais de busca e otimização, principalmente em quatro aspectos:

- 1) AGs trabalham com uma codificação do conjunto de parâmetros e não com os próprios parâmetros.
- 2) AGs trabalham com uma população e não com um único ponto.
- 3) AGs utilizam informações de custo ou recompensa e não derivadas ou outro conhecimento auxiliar.

- 4) AGs utilizam regras de transição probabilísticas e não determinísticas.

3.10

Avaliação de Carteiras e Medidas de Risco x Retorno

A seguir apresenta-se as principais teorias de análise de Risco x Retorno e avaliação de carteiras de investimento que serviram de base para as funções de avaliação e otimização de estratégias com opções financeiras desenvolvidas nesse trabalho.

Uma carteira de investimentos é um conjunto de ativos e derivativos financeiros pertencentes a um investidor em um determinado momento, tendo como objetivo principal obter o maior rendimento correndo o menor risco possível. Para que o investidor possa auferir um maior retorno deverá aceitar um nível maior de risco, isso se deve ao fato de que investimentos mais arriscados possuem um maior prêmio de risco (correlação fortemente positiva), cabendo então ao investidor montar e gerenciar sua carteira de forma eficiente.

Markowitz (2001) e Sharpe (1966) contribuíram imensamente com o processo de seleção de carteiras de investimento, desenvolvendo metodologias de avaliação e compensação de risco através da diversificação de investimentos e suas teorias continuam a ser utilizadas nos dias de hoje.

3.10.1

Critério de Média-Variância de Markowitz

Segundo Damodaran (2002) o risco pode ser conceitualmente dividido em dois riscos básicos: o risco diversificável (não sistemático) que refere-se a uma empresa ou a um grupo pequeno de empresas e o risco não diversificável que afeta todo o mercado.

Markowitz, ganhador do Prêmio Nobel de economia em 1991, desenvolveu um processo de alocação de ativos por diversificação, para eliminar o risco não sistemático, que envolve a combinação de ativos correlacionados no sentido de formar portfólios eficientes.

Ele considerou o retorno futuro como uma variável aleatória com distribuição normal, descrita pelos seus dois primeiros momentos: a média (expectativa de retorno) e a variância (indicador de risco). O retorno esperado da carteira é a soma ponderada das médias de cada ativo e a variância da carteira foi registrada como a soma das variâncias individuais de cada ação multiplicada pelas covariâncias entre pares de ações da carteira, de acordo com o peso de cada ação. Markowitz comenta que deve haver uma carteira de ações que maximize o retorno para uma dada variância ou que minimize a variância para um dado retorno, e esta deve ser a carteira recomendada para um investidor.

O modelo de Markowitz é dado pela Eq. (24), Eq. (25) e pela Eq. (26):

$$E = \sum_{i=1}^n X_i \mu_i \quad \text{Eq. (24)}$$

$$V = \sum_{i=1}^n \sum_{j=1}^n X_i X_j \sigma_{ij} \quad \text{Eq. (25)}$$

$$\begin{aligned} \sum_{i=1}^n X_i &= 1 \\ X_i &\geq 0 \end{aligned} \quad \text{Eq. (26)}$$

Onde:

E é o retorno esperado da carteira;

V é a variância da carteira;

X_i é a participação de cada ativo;

μ_i é o retorno esperado de cada ativo;

σ_{ij} é a covariância entre o par de ativos se (i) diferente (j) e variância se (i) igual a (j);

Teríamos então um problema de otimização que busca o menor risco (V) para dado patamar de retorno, ou o maior retorno (E) para dado nível de risco.

Segundo Bernstein (1996), o modelo mostra que o retorno de uma carteira diversificada equivale à média ponderada dos retornos de seus componentes individuais, já a sua volatilidade é inferior à volatilidade média de seus componentes.

Markowitz assume uma distribuição normal e, portanto, as características de risco-retorno de uma carteira podem ser descritas inteiramente pelo primeiro e segundo momento da distribuição. Essa pode ser uma boa aproximação para ativos individuais, mas não para carteiras de ativos. A teoria de Markowitz também leva em consideração somente ativos financeiros, não podendo ser calculada para derivativos e opções financeiras pela incapacidade de se calcular as correlações e a distribuição de retorno desses derivativos.

3.10.2

Métricas de Avaliação de Desempenho de Carteiras e Fundos

3.10.2.1

Índice de Sharpe

O Índice de Sharpe, criado por William Sharpe, em 1966, é uma das métricas mais utilizadas na avaliação de carteiras e fundos de investimento. Ela expressa a relação retorno/risco: “É a razão do retorno anualizado que excede a taxa livre de risco dividido pela volatilidade” (Sharpe, 1966), Eq. (27).

$$IS = \frac{\mu - r_f}{\sigma} \quad \text{Eq. (27)}$$

Onde:

IS é o Índice de Sharpe.

μ representa a média dos retornos passados.

r_f é a taxa livre de risco ou qualquer retorno mínimo aceitável. No Brasil toma-se como taxa sem risco a taxa de títulos públicos do Governo Federal (Selic), ou a taxa diária do CDI, que se refere a títulos privados de alta qualidade de crédito, que é muito próximo a Selic e mais facilmente obtida.

σ representa o desvio padrão dos retornos.

O índice de Sharpe por ser uma métrica bem intuitiva, fácil de entender e de se utilizar ganhou muito espaço no mercado. A teoria de Markowitz determina a

composição da carteira ótima como a que contem a melhor relação risco/retorno e as carteiras com maior IS são exatamente essas carteiras ótimas.

Diversos cuidados devem ser tomados ao se aplicar o IS na seleção ou classificação de investimentos. O primeiro deles vem do fato de o cálculo do IS não incorporar informação sobre a correlação entre os ativos. Portanto, este indicador não nos auxilia quando queremos adicionar um novo ativo com risco a uma carteira que já tenha ativos arriscados.

O IS é baseado em valores históricos e isso gera alguns problemas na sua avaliação. Primeiro que só se pode utilizá-lo para avaliar carteiras com um razoável histórico de rentabilidade, não permitindo uma análise de uma nova carteira. Outro problema é que não significa que uma estratégia que nunca tenha apresentado prejuízo no passado e tenha um alto IS tenha pouco risco. Várias estratégias simples com opções se comportariam dessa forma se avaliadas pelo IS, pois é fácil conseguir um IS alto vendendo opções bem fora do dinheiro. Durante vários dias a estratégia ganhará dinheiro, de pouco em pouco, mas um dia poderá sofrer uma perda tão grande que pode quebrar o investidor. Portanto qualquer medida de risco baseada única e exclusivamente no histórico, só mostrará o que ocorreu e não o que pode ocorrer. “Retornos passados parecem demonstrar previsões muito fracas e inconsistentes a respeito dos retornos futuros” (Capocci, 2007).

Outro cuidado que deve-se levar em conta quando se utiliza o IS é que ele considera que os retornos históricos da carteira ou fundo analisado seguem uma distribuição normal e que portanto pode-se avaliar seu risco/retorno somente pelo primeiro e o segundo momento da distribuição (média e desvio padrão), o que pode induzir a conclusões equivocadas. Para ilustrar tal efeito a Figura 18 mostra duas distribuições de retorno D_1 e D_2 com a mesma média e desvio padrão, que receberiam avaliação idêntica pelo IS. Obviamente um investidor não seria indiferente entre essas duas estratégias.

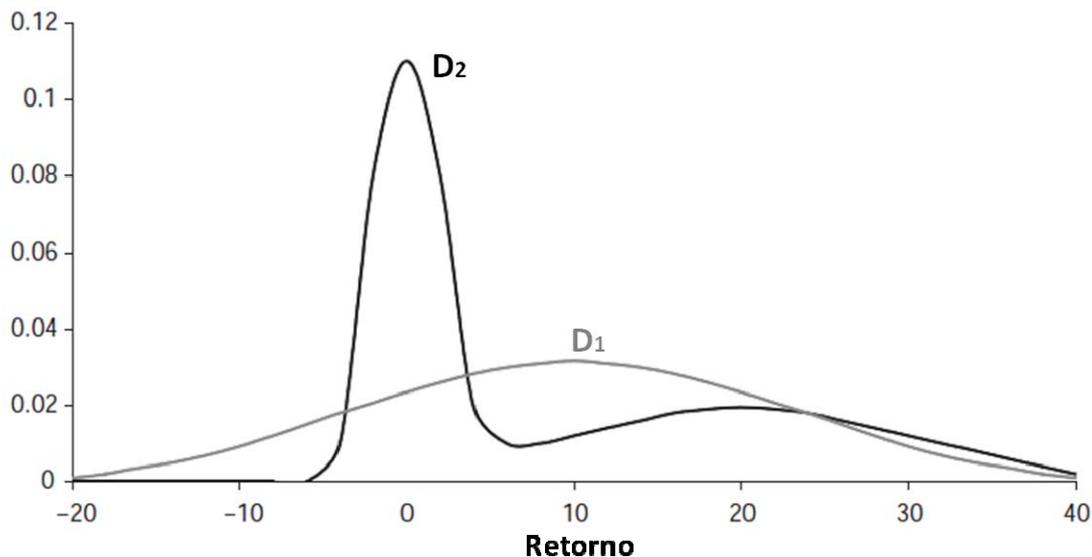


Figura 18 – Distribuições de Retorno com mesma média e desvio padrão.

A Figura 19 mostra duas estratégias que foram avaliadas pelo IS. A estratégia 1 (E_1) teve $IS=7,4$ enquanto que a estratégia 2 (E_2) teve $IS=4,7$, o que não faz muito sentido quando se pode argumentar qualitativamente que a E_2 é uma estratégia claramente superior.

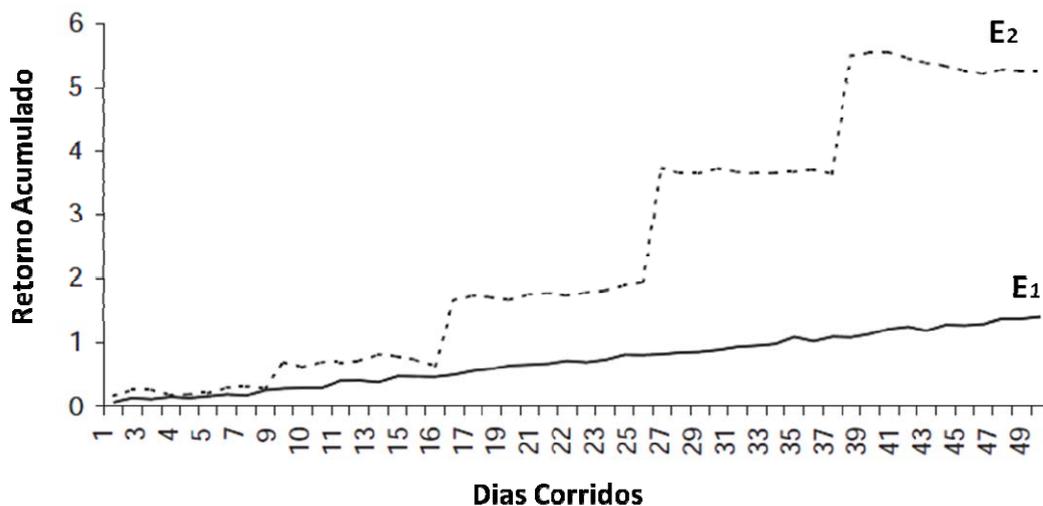


Figura 19 – Gráficos Comparativos entre duas estratégias

A idéia geral por trás do IS é computar uma medida boa (o retorno) sobre uma medida ruim (o risco). Mas a volatilidade (desvio padrão dos retornos) não é uma medida necessariamente ruim, a volatilidade causada na estratégia por grandes retornos positivos é boa (exemplo da E_2 na Figura 19), somente as causadas por perdas são ruins. Alguns pesquisadores com essa questão em mente

alteraram o denominador (medida de risco) da função de avaliação mantendo a mesma idéia geral.

3.10.2.2

Índice de Sortino

Sortino (1994) utilizou o desvio negativo (*downside deviation*), que considera no cálculo do desvio padrão apenas as perdas financeiras. Sua métrica de avaliação segue a forma da Eq. (28):

$$ISor = \frac{\mu - r}{\sigma_d} \quad \text{Eq. (28)}$$

Onde:

μ representa a média dos retornos passados.

r é um retorno mínimo aceitável.

σ_d é o desvio padrão negativo (*downside deviation*).

3.10.2.3

Índice de Calmar

Outra medida derivada do IS é o Índice de Calmar definido como o excesso do retorno contra um benchmark (retorno mínimo aceitável) sobre a perda máxima do histórico dos retornos, demonstrado na Eq. (29). Na literatura um Índice Calmar de 1,0 é considerado bom. “Se você está em busca de um x% de retorno deve estar preparado para um x% de perda.” (Young, Terry, 1991).

$$ICalmar = \frac{\mu - r}{\text{perda máx}} \quad \text{Eq. (29)}$$

3.10.2.4

Índice de Sterling

O Índice de Sterling é muito parecido com o ICalmar e é definido como o excesso do retorno sobre a média das maiores perdas. Normalmente utiliza-se a

média das 5 (cinco) maiores perdas para diminuir o efeito que *outliers* poderiam ter sobre o ICalmar, Eq. (30).

$$ISterling = \frac{\mu - r}{\text{m\u00e9dia das maiores perdas}} \quad \text{Eq. (30)}$$

“A melhor estrat\u00e9gia n\u00e3o \u00e9 necessariamente a que demonstra maiores ganhos quando as coisas v\u00e3o bem; ela pode ser a que demonstra menores perdas quando as coisas v\u00e3o mal” (Natenberg, 1994).

3.10.2.5

Fun\u00e7\u00e3o Omega

Essas varia\u00e7\u00f5es do \u00cdndice de Sharpe tentam melhorar o fato de que a vari\u00e2ncia n\u00e3o consegue capturar bem o risco da carteira. Uma outra forma de abordar esse problema \u00e9 considerando mais do que somente o segundo momento da distribui\u00e7\u00e3o de retornos, usando a distribui\u00e7\u00e3o inteira, pois “os primeiros dois momentos n\u00e3o descrevem todo o comportamento hist\u00f3rico que se est\u00e1 tentando analisar” (Kazemi, Gupta, 2003), como se pode ver na Figura 18. Pensando nisso Keating and Shadwick (2002) desenvolveram a Fun\u00e7\u00e3o Omega que t\u00eam o formato da Eq. (31):

$$\Omega(r) = \frac{\int_r^b (1 - F(x)) dx}{\int_a^l F(x) dx} \quad \text{Eq. (31)}$$

Onde:

[a, b] \u00e9 o intervalo dos retornos.

F \u00e9 a fun\u00e7\u00e3o de densidade acumulada da distribui\u00e7\u00e3o dos retornos.

r \u00e9 o retorno m\u00ednimo aceit\u00e1vel. Retornos abaixo de r s\u00e3o como se fossem negativos.

“Omega tamb\u00e9m leva em conta um n\u00edvel de retorno chamado de limite (L) definido exogenamente, o qual \u00e9 a fronteira entre o que se considera como ganho e como perda. Mesmo em distribui\u00e7\u00f5es normais, dependendo do valor do L, a medida Omega fornece informa\u00e7\u00f5es adicionais que s\u00f3 a m\u00e9dia e a vari\u00e2ncia n\u00e3o

conseguiriam. Isso levaria a obter diferentes resultados em otimização de carteiras, se comparado com a otimização clássica de Markowitz” (Castro, 2008).

Omega é o somatório do valor do retorno multiplicado pela probabilidade desse retorno acontecer, para os retornos acima do limite “l”, divididos pelo somatório do valor do retorno multiplicado pela probabilidade desse retorno acontecer, para retornos abaixo do limite “l”. Quanto maior o Omega melhor a carteira/fundo/estratégia.

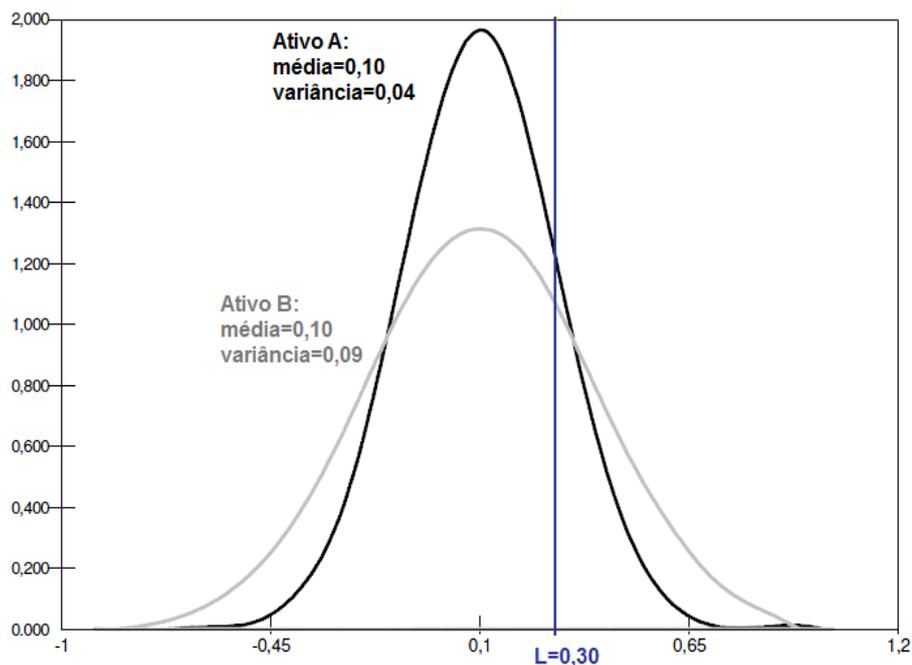


Figura 20 – Distribuição de Retornos dos ativos A e B com limite $L=0,30$

Para exemplificar a diferença entre o IS e a função Omega, temos a Figura 20. Ambos os ativos são normalmente distribuídos e têm mesma média. Eles diferem na variância que são 0,04 e 0,09 respectivamente para o ativo A e B.

Claramente pelo IS deve-se preferir o ativo A, que possui uma menor variância e minimiza tanto as perdas potenciais quanto os ganhos. Se considerarmos que uma perda é definida por um retorno abaixo de $L=0,3$ podemos calcular as probabilidades acima e abaixo desse valor:

Ativo A:

- probabilidade acima de $L = 0,1581$
- probabilidade abaixo de $L = 0,8419$
- probabilidade lucro / probabilidade prejuízo = $0,1581/0,8419 = 0,1878$

Ativo B:

- probabilidade acima de $L = 0,2522$
- probabilidade abaixo de $L = 0,7178$
- probabilidade lucro / probabilidade prejuízo = $0,2522/0,7178 = 0,3373$

Calculando-se a probabilidade de lucro dividido pela probabilidade de prejuízo vemos que o ativo B resulta-se mais conveniente do que o ativo A, algo que o IS não consegue identificar.

Aparentemente o Omega é muito bom conceitualmente, mas extrapolar valores extremos, que podem ter ocorrido no passado, para o futuro, pode não ser uma decisão muito racional na visão de um investidor, e a presença (ou não) desses valores extremos alteram de forma significativa o valor do Omega (o que também é válido para as outras medidas).

Não existe função perfeita para avaliar os retornos de uma estratégia, sendo uma boa prática a utilização de diversas funções diferentes, separadamente, ou uma combinação linear entre algumas delas para se chegar a uma melhor decisão de investimento.