

## 3 Metodologia

Neste capítulo apresentaremos a metodologia utilizada ao longo do trabalho para construir um modelo capaz de ajudar o investidor a tomar decisões quanto à hora de comprar ou vender um determinado lote de ações. Para tanto, dividiremos a metodologia em três partes: tratamento dos dados iniciais, montagem dos gráficos de controle e escolha das estratégias.

Os procedimentos a serem apresentados foram idealizados para que se possa, com os recursos de pesquisa disponíveis, fazer uma comparação experimental de opções de estratégias para escolha de momentos favoráveis de compra e venda de ações. Basicamente, a metodologia aqui apresentada tratará dos procedimentos que tornaram possíveis a simulação, com os dados de julho de 1994 a abril de 2003, dos resultados financeiros que a adoção de determinada estratégia de compra e venda de ações poderia ter gerado a investidores interessados em aplicar modelos que englobam tanto uma análise de rentabilidade quanto um estudo dos níveis de risco associados.

### 3.1 TRATAMENTO DOS DADOS INICIAIS

#### 3.1.1 Coleta de dados

A coleta de dados precisava ser feita a partir de uma fonte reconhecida pelo mercado investidor. Pela facilidade de exportação de dados e por ser uma fonte acessível e bastante respeitada no mercado, optou-se pelo uso do programa **Economática** que apresenta uma gama de dados históricos sobre ações listadas nas bolsas de valores brasileiras, além de alguns índices de preços, cotações de moedas e índices de outras bolsas de valores do mundo.

Para garantir a significância e a abrangência das conclusões a serem tiradas, era preciso optar por um período relativamente extenso e por uma periodicidade para estudo. Diante das alternativas, considerou-se que o período a partir do plano Real seria suficiente para gerar um histórico de dados capaz de atender às necessidades deste trabalho, uma vez que garantiria uma quantidade de dados suficiente para um estudo diário, semanal ou até mesmo mensal, além de facilitar a manipulação dos dados, uma vez que torna dispensável o uso de

ferramentas (como por exemplo o uso de deflatores) para tornar os dados comparáveis. Assim, foram pesquisadas as cotações de julho de 1994 até a última informação disponível no momento da coleta (29/04/2003).

No entanto, tratando-se da proposta de uma nova metodologia para indicar pontos de compra e venda de ações, considerou-se que seria conveniente separar o conjunto dos dados coletados em duas amostras distintas: uma de investigação inicial e outra de controle. Na primeira estaríamos interessados em investigar e definir, com base nos dados históricos, se é possível identificar pontos de compra e venda de ações com alguma característica que coincida com períodos relativos às rentabilidades mais elevadas (por exemplo, perceber que investimentos iniciados sempre após uma queda superior ao LIC e finalizados após uma alta maior do que o LSC teriam gerado rentabilidades superiores às demais condições investigadas). E, na segunda, em aplicar as estratégias previamente definidas (simulando o seu uso em tempo real) para avaliar seu desempenho em outro período de tempo. Para constituir a primeira amostra separamos os dados referentes ao período de julho de 1994 a outubro de 2002; os demais, referentes ao último bimestre de 2002 e início de 2003, compuseram a amostra de controle.

Adicionalmente, para atender às necessidades do investidor de risco, optou-se pelo estudo das cotações diárias de ações. E, para atender aos propósitos desta dissertação e poder sugerir alguma generalização, foram coletados os dados referentes às dez maiores ações (segundo a sua participação no Ibovespa) listadas na Tabela 2.1, além dos valores divulgados tanto para o Ibovespa quanto para o Dow Jones, índice da bolsa de Nova York.

Mais uma decisão precisava ser tomada: qual cotação escolher? Estavam disponíveis dados de cotação de abertura, máxima, mínima, média e de fechamento. Inicialmente optou-se por armazenar todos estes valores, pois poderiam ser interessantes para a escolha das estratégias, mas decidiu-se trabalhar, na construção dos gráficos de controle, apenas com as cotações de fechamento, por ser esta uma prática de mercado universalmente adotada.

### **3.1.2 Análise Preliminar**

Como em qualquer estudo quantitativo, é conveniente realizarmos uma investigação preliminar dos dados a serem utilizados. “Só é possível monitorar um processo após conhecê-lo bem: a etapa inicial, que antecede a própria construção e a utilização dos gráficos de controle, é de aprendizagem: nela

procuram-se conhecer os fatores que afetam a característica de qualidade X.”<sup>1</sup> Como se trata de um estudo fundamentado em ferramentas estatísticas, optou-se, nesta fase, pela utilização dos gráficos mais utilizados para análise exploratória de dados. Os métodos gráficos são fortemente indicados na medida em que possuem forte apelo visual e permitem que o investigador melhore a sua capacidade de identificar a ocorrência de alguns fenômenos, tais como comportamentos cíclicos, trajetórias de crescimento, etc.

Normalmente o uso de métodos gráficos é interessante para buscar padrões e relações, confirmar expectativas definidas a priori (por exemplo, que dias de alta de dólar geralmente implicam em dias de queda das cotações da Bolsa de Valores), descobrir novos fenômenos, confirmar suposições que se tenha, mas que ainda não possuam um embasamento teórico ou apresentar resultados de modo mais rápido e fácil. No caso desta dissertação, estaremos interessados em todos estes aspectos, de modo que iremos nortear a análise inicial de acordo com estes objetivos.

Para atingir o alvo pretendido, começaremos o trabalho com as seguintes ferramentas: plotagem da série de cotações em seqüência cronológica; histograma dos retornos com a divisão dos dados em 10 classes com mesma amplitude, com e sem *outliers* e montagem do *Box-plot* (também dos retornos).

De acordo com a teoria sobre retorno de ações, a razão entre cotações  $\frac{CA_i}{CA_{i-\tau}}$ <sup>2</sup>, constitui uma distribuição log-normal e representa adequadamente a evolução de ganhos e perdas de ativos de risco. Desta forma, seria conveniente aplicarmos logaritmo neperiano aos dados para transformá-los numa série de observações cujos valores são normalmente distribuídos, apesar de Wheeler & Chambers argumentarem que, não é condição necessária haver normalidade dos dados para que o gráfico de controle funcione adequadamente. Vale destacar que, segundo os autores, ao contrário do que se pode pensar, este bom funcionamento não é resultado do Teorema Central do Limite<sup>3</sup>, uma vez os gráficos de controle funcionam mesmo para dados individuais e amostras bastante reduzidas.

<sup>1</sup> COSTA, A.F.; EPPRECHT, E. & CARPINETTI, L.C., op. cit., pp.30, 31.

<sup>2</sup> Relembrando que, conforme definido na seção 2.2,  $CA_i$  = cotação atual do ativo i e  $CA_{i-\tau}$  = cotação do ativo i na data base t (no caso de retorno diário,  $\tau = 1$ )

<sup>3</sup> In Larson, H.; Introduction to Probability Theory and Statistical Inference, 1982, p. 288: “Theorem 5.6.1. Let  $X_1, X_2, \dots, X_n$  be a sequence of independent identically distributed random variables, each with mean  $\mu$ , variance  $\sigma^2$ , and moment generating function  $m(t)$ .

## 3.2 MONTAGEM DOS GRÁFICOS DE CONTROLE

### 3.2.1 Obtenção de modelo de ruídos e sem a influência de sazonalidade

Depois de escolher, armazenar e analisar os dados, a proposta inicial era trabalhá-los de modo a se obter um modelo de partida que apresentasse apenas resíduos normalmente distribuídos, independentes e com média nula. Este procedimento foi adotado em função dos argumentos apresentados na subseção 2.1.2<sup>4</sup>.

Para obtenção dos resíduos, partiu-se do pressuposto que o ln RD é um processo estocástico (vide definição no Anexo) que pode ser estimado. Assim, o resíduo seria a diferença entre os valores previsto e observado em cada período. Logo, torna-se fundamental estudar o comportamento dos logaritmos dos retornos. Para tanto, foram utilizadas duas metodologias para identificação de modelos: alisamento exponencial (com e sem tendência) e Box & Jenkins. O primeiro procedimento consiste em estimar o peso das observações passadas no dado presente de acordo com médias móveis históricas. Já o segundo, procura, através de um processo de minimização de erros, encontrar uma equação que melhor se ajuste aos dados. Com o intuito de facilitar os cálculos, foi escolhido o software Forecast Pro, tendo os seus resultados sido verificados no Excel.

Outro teste que parecia interessante era averiguar a presença de sazonalidade em algum dia da semana. É comum encontrarmos histórias que qualificam a quinta-feira como o “dia do boato” ou que afirmam que os resultados de sexta-feira são influenciados pela proximidade do final de semana. Portanto, parece relevante adotar algum procedimento que permitisse confirmar ou rejeitar estas hipóteses. Neste sentido, foi adotado um procedimento de dessazonalização dos dados, trabalhando-se com a hipótese de um modelo aditivo. O modelo multiplicativo foi descartado por duas razões:

- i) os dados para cálculo já representavam logaritmos de modo que, se os dados originais apresentassem comportamento sujeito a fatores sazonais multiplicativos, o tratamento inicial tê-los-ia

---

If a new sequence of standardized partial sums is defined by  $Z_n = (\sum_{i=1}^n X_i - n\mu) / (\sigma \sqrt{n})$ ,  $n=1, 2, 3, \dots$  then the probability law for  $Z_n$  converges to the standard normal as  $n \rightarrow \infty$

<sup>4</sup> Mais adiante estará sendo apresentada a metodologia para testar a necessidade deste procedimento (seção 3.3.2).

transformado num conjunto de valores onde a sazonalidade é aditiva;

- ii) por se tratar da estimação de um modelo de ruídos com média esperada em torno de zero, uma carga sazonal aditiva parece fazer mais sentido.

Assim, em se tratando de uma modelagem com suposição de sazonalidade aditiva, estaremos estimando equações do tipo:

$$z_t = \phi(t) + \sum_{i=1}^5 D_i S_i + \varepsilon_t$$

Onde:

$S_i$  = valor da carga sazonal do  $i$ -ésimo dia da semana;

$D_i = 1$  se  $i$  corresponde ao mesmo dia da semana de  $t$  ou 0, caso contrário;

$1 \leq i \leq 5$ ;

$\phi(t)$  representa um modelo qualquer (conhecido).

Como os anos em questão apresentavam feriados móveis, um passo inicial imprescindível era reconstruir a série sem que houvesse “*missing values*”. O procedimento empregado no sentido de se substituir o logaritmo dos retornos dos dias sem pregão foi o seguinte:

Passo 1: identificar os feriados;

Passo 2: para cada seqüência de dois dias úteis seguidos (considerou-se como seqüência de dois dias úteis seguidos o conjunto formado por dois dias consecutivos com negócios de pregão depois de excluídos sábado e domingo, ou seja, sem feriados entre um determinado pregão e o seguinte), calcular o logaritmo neperiano dos retornos acumulados para dois dias e montar uma série histórica com estes valores e com a sua decomposição pelos dois dias consecutivos (por exemplo; se, para o período de 2 dias foi encontrado um retorno de 2,51%, é preciso identificar que parcela deste refere-se ao primeiro dia da seqüência e que parcela refere-se ao segundo dia, de modo que o resultado acumulado não seja alterado, no caso, 2% e 0,5%)

Passo 3: para cada feriado, calcular o valor do logaritmo neperiano do retorno do dia útil imediatamente posterior (digamos que tenha sido encontrado o valor 2,52%);

Passo 4: procurar na série construída no Passo 2 o dia em que o logaritmo neperiano do retorno para dois dias for mais próximo ou igual ao encontrado no Passo 3 (no caso de não existir um valor igual na série histórica, procurar o

maior valor que esteja abaixo do encontrado no Passo 3), identificando sua decomposição quanto aos retornos para o primeiro e segundo dias (continuando com o exemplo, imagine que 2,51% é o menor valor mais próximo de 2,52%);

Passo 5: atribuir, ao feriado, o valor identificado no Passo 4 para o primeiro dia (neste caso, o feriado receberia, para efeito de cálculo, o valor de 2% de retorno);

Passo 6: trocar o retorno do primeiro dia útil posterior ao feriado pela diferença entre o valor original deste e o encontrado no Passo 5, ou seja, a soma dos logaritmos dos retornos atribuídos à seqüência de dois dias (feriado e dia útil seguinte) será exatamente igual ao valor original do logaritmo do retorno apurado entre o dia útil imediatamente posterior e o imediatamente anterior ao feriado (no nosso exemplo, 0,5098% para ser atribuído ao dia que, originalmente, apresentava retorno de 2,52%);

Passo 7: no caso de o feriado referir-se a uma seqüência de dias maior do que dois (por exemplo, Carnaval ou Semana Santa), aplicar o procedimento dos Passos 2 a 6 apenas ao último dia não útil e, logo depois, repetir os Passos 2 a 6 para o penúltimo dia não útil e, depois, para o ante-penúltimo dia não útil, se for o caso, de modo não reste nenhuma seqüência de segunda a sexta-feira sem um valor de retorno associado.<sup>6</sup>

Com este procedimento, montou-se uma base de dados com aproximadamente 260 observações para cada ano, totalizando 2.301 observações para o período total, representando um acréscimo de 119 dados, ou 5,4% da amostra inicial. Com estes valores foi adotada uma metodologia para estimação dos efeitos sazonais aditivos, envolvendo as seguintes etapas:

Etapla 1: calcular a média móvel para cada cinco observações:

$$MM_i = \sum_{t=i-4}^i x_t / 5;$$

Etapla 2: calcular uma estimativa pontual para a carga sazonal segundo a seguinte fórmula:

$$\hat{S}_i = x_i - MM_i, \text{ para } i = 5, 6, 7, \dots, 780;$$

<sup>5</sup> onde 1 = segunda-feira, 2 = terça-feira, 3=quarta-feira, 4 = quinta-feira e 5 = sexta-feira.

<sup>6</sup> Pode-se imaginar que, para estes casos, dever-se-ia ter repetido o processo começando, no Passo 2, não com uma seqüência de dois dias, mas com as seqüências com o mesmo número de feriados mais um, no entanto, há apenas duas ocorrências deste tipo por ano, ou seja, usar a metodologia aqui proposta poupa recursos computacionais, mas sem gerar prejuízos relevantes para o que se pretende estudar.

Etapa 3: calcular a média dos fatores sazonais para cada um dos cinco dias da semana (2ª a 6ª feira), de modo que se obtenham apenas 5 cargas sazonais ( $\hat{S}_j$ ), uma para cada dia da semana;

Etapa 4: ajustar as cargas sazonais para que a sua soma seja zero, segundo a seguinte correção:  $S_j^* = \hat{S}_j - \sum \hat{S}_j / 5$ ;

Etapa 5: dessazonalizar a série com o seguinte cálculo:  $x_{id} = x_i - S_j^*$ .

Sobre as estimativas das cargas sazonais foram aplicados uma série de testes (para verificar a significância dos estimadores; em outras palavras, se a série exiba ou não sazonalidade):

Teste 1: verificar se as cargas sazonais médias (das ações, excluindo os índices) para cada dia da semana são estatisticamente diferentes de zero, ou seja, fazer um teste de média onde  $H_0: \mu = 0$  contra  $H_1: \mu \neq 0$ , a idéia é verificar se há sazonalidade ou se os valores encontrados diferem de zero apenas por efeito aleatório, caso se conclua que há sazonalidade, ela poderá ser incorporada ao modelo e, além disso, pode fornecer informações de que compras ou vendas em um determinado dia da semana devem ser evitadas, por exemplo;

Teste 2: averiguar se há homogeneidade de variâncias para as cargas sazonais entre as ações, ou seja, testar a hipótese  $H_0: \sigma_i = \sigma_j$  contra  $H_1: \sigma_i \neq \sigma_j$ , onde  $i$  e  $j$  representam cada uma das ações e  $i \neq j$ , assim, se for detectado que determinada ação possui variância menor em relação às demais, pode ser um indicativo de que ela é preferível às demais porque deve apresentar um risco menor;

Teste 3: investigar se há homogeneidade de variâncias das cargas sazonais dos dias da semana, ou seja, testar a hipótese  $H_0: \sigma_i = \sigma_j$  contra  $H_1: \sigma_i \neq \sigma_j$ , onde  $i$  e  $j$  representam cada um dos dias da semana e  $i \neq j$ , ou seja, averiguar se os pressupostos para uma análise de variância das cargas sazonais, considerando o dia da semana como fator (Teste 5, adiante), são satisfeitas;

Teste 4: apurar se os valores encontrados para as cargas sazonais médias das ações para cada um dos dias em questão estão normalmente distribuídos, através da construção de um teste de normalidade das cargas sazonais para cada dia da semana, com a montagem de cinco gráficos de plotagem de probabilidade normal ou *p-plot*. O objetivo deste teste é verificar se as diferenças entre cargas sazonais obtidas para um mesmo dia e diferentes ações

pode ser considerada aleatória e, neste caso, pode-se pensar em usar um único fator sazonal para cada dia da semana, para todas as ações;

Teste 5: analisar a variância para todas as cargas sazonais conjuntamente, com a aplicação de um teste F para averiguar se as médias das cargas sazonais de cada dia da semana são todas iguais, ou seja, testar a hipótese  $H_0: \mu_i = \mu$  contra  $H_1: \exists \mu_i \neq \mu$ , onde  $i$  e  $j$  representam cada um dos dias da semana e  $i \neq j$ . A idéia é verificar se há influência de, pelo menos, um fator sazonal, respondendo à mesma questão levantada no Teste 1 (se houver evidências para rejeitar  $H_0$  no Teste 5, então algum dia da semana deverá ter carga sazonal diferente de zero e também apresentará evidências para se rejeitar a hipótese nula do Teste 1);

Teste 6: observar o comportamento do modelo gerado sob a hipótese inicial de cargas sazonais nulas em comparação com o modelo obtido com o uso dos dados dessazonalizados.

Este conjunto de procedimentos é necessário para confirmar a existência ou não de influências de sazonalidade.

### 3.2.2 Gráficos de Controle

Conforme apontado no capítulo anterior (Seção 2.1.2), a construção de gráficos de Shewhart para uma série de dados autocorrelacionados pode aumentar sensivelmente a probabilidade de alarmes falsos. Assim, é prudente utilizar a modelagem de séries temporais para calcular resíduos, com os quais é possível construir gráficos de controle sem correr o risco de aumentar a incidência de alarmes falsos. O princípio em que se baseia esse procedimento é que, se o modelo escolhido for adequado, então os resíduos devem ser i.i.d. e, portanto, satisfazem os pressupostos (condições de validade) para a aplicação dos gráficos. Mais ainda, se houver alguma alteração na dinâmica ou nos parâmetros do processo devido a alguma causa especial, ocorrerá um desvio entre o modelo e a série, que provocará um aumento nos resíduos, que poderão ser sinalizados pelos gráficos de controle. Um resíduo positivo excepcionalmente grande (acima do LSC) indicaria uma tendência de aumento na variável (pois o modelo estará subestimando os valores desta), e, simetricamente, um grande resíduo negativo (abaixo do LIC), indicaria uma provável redução no valor da variável considerada, em relação aos valores previstos pelo modelo.

Antes da montagem dos gráficos de controle, com base nos resultados obtidos com o programa FPW na estimação de modelos do tipo alisamento exponencial e Box&Jenkins, pretende-se escolher um único tipo de modelo

(dentre todos os modelos obtidos e testados por ambas as metodologias, optar, por exemplo, pelo modelo AR(1), que, segundo a “prática de mercado”<sup>7</sup>, acredita-se representar adequadamente o comportamento de uma série de retornos em bolsa) que deverá ser utilizado para estimação dos ativos em estudo. A escolha de apenas um único tipo de modelagem é imprescindível para facilitar a operacionalização de uma metodologia para o dia-a-dia do investidor que esteja interessado em aplicá-la. Escolhido o modelo, serão gerados seus resíduos e, com estes, será montada a base de dados primária para construção dos gráficos de controle.

A montagem dos gráficos foi feita na planilha eletrônica Excel, ambiente eletrônico comum aos possíveis interessados em aplicar esta metodologia. Como optou-se por estudar dados de fechamento, o gráfico mais apropriado é o de observações individuais, o gráfico de X. Adicionalmente, como volatilidade é uma preocupação de quem trabalha com ativos de risco, construiu-se também o gráfico de mR, que poderá ser útil na definição futura das estratégias.

Segundo Wheeler & Chambers, para esses gráficos temos que considerar os seguintes limites de controle:

GRÁFICO	LIC	LC	LSC
X	$LC - 2,66LC_{mR}$	$\bar{X} = \sum_{i=1}^n X_i / n$	$LC + 2,66LC_{mR}$
mR	–	$\sum_{i=2}^n  X_i - X_{i-1}  / (n-1)$	$3,268 * LC_{mR}$

Onde:

$$X_i = \ln RD_i = \ln (CA_i / CA_{i-1});$$

$$|X_i - X_{i-1}| = |\ln RD_i - \ln RD_{i-1}| = VP \text{ (variação entre os dias).}$$

Optou-se por não excluir do cálculo dos limites os valores de possíveis *outliers*. Esta decisão baseou-se na idéia de tornar mais simples, na prática, a aplicação da ferramenta, e na consideração de que qualquer efeito dos *outliers* sobre a escolha das estratégias deverá ser compensado pelo fato de que

---

<sup>7</sup> Entende-se por “prática de mercado” a modelagem que analistas financeiros utilizam para precificar ativos, baseada nas teorias desenvolvidas por estudiosos da área.

estaremos experimentando o uso não apenas dos limites de 3-sigma, mas também frações destes.

Uma questão importante após a construção dos gráficos é observar padrões indicativos de “processo fora de controle”, com seqüências de valores que não correspondem a um padrão aleatório ou pontos fora dos limites de controle. Esses podem representar um importante sinal, especialmente em se tratando de investimentos em bolsa de valores, quando buscamos nos prevenir de retornos negativos, mas gostaríamos de antecipar retornos positivos. Os pontos fora de controle são fundamentais para se detectar, de modo rápido, que o processo em estudo não se encontra aleatoriamente distribuído em torno de sua média histórica e, portanto, pode estar seguindo uma nova trajetória. O modo mais simples de identificação é o visual.

Mas nem sempre esta caracterização é um trabalho simples e rápido. Confiar na investigação visual pode implicar em sérios riscos de cometer erros. Isto porque sabemos que o simples fato de encontrarmos vários pontos acima ou abaixo da média pode não ser devido a mudanças no comportamento do fenômeno em questão, já que, estatisticamente, sempre existe uma pequena chance de um processo mesmo sob controle e com distribuição simétrica apresentar valores consecutivos de um mesmo lado da linha média.

Portanto, é indispensável somar às impressões visuais alguns testes estatísticos formais. No caso dos gráficos de controle, existe, na literatura e nos programas computacionais apropriados, uma grande quantidade de testes descritos para identificar o descontrole de um processo e a presença de influência de causas especiais. No entanto, aplicar todos ao mesmo tempo também traz problemas na medida em que cada um contém um risco de erro do tipo I, o que pode implicar no aumento exagerado do nível  $\alpha$  de significância conjunta dos testes.

Na medida em que se opta por aplicar uma maior quantidade de testes, isto implica em incrementar as chances de se detectar um problema onde ele não existe (denominado “alarme falso”). Para efeito de exemplo, suponha um teste de hipótese com probabilidade de erro do tipo I = 1%. Se aplicarmos 10 testes com o mesmo grau de certeza, estaremos diminuindo a certeza conjunta dos testes, ou seja, a chance de nenhum deles apresentar alarmes falsos, vai para 90,44%, supondo independência dos resultados dos diversos testes, o que pode ser preocupante, já que é bastante superior à probabilidade individual de alarmes falsos de cada teste individual. Ou seja, mesmo que se trabalhe com uma probabilidade baixa para o erro do tipo I de cada teste, dependendo da

quantidade de testes que se aplique, é possível que a probabilidade conjunta de se gerar pelo menos um alarme falso seja bastante superior ao valor aceitável. Para contornar este problema, Wheeler & Chambers recomendam a realização de apenas quatro testes conjuntos sobre a amostra, a saber:

1º teste: observar pelo menos um ponto fora do limite 3DP (3 sigma);

2º teste: encontrar pelo menos 2 pontos, numa seqüência de 3, com uma distância igual ou maior do que 2DP (2 sigma) da média;

3º teste: reconhecer pelo menos 4 pontos, numa seqüência de 5, com uma distância igual ou superior a 1DP (1 sigma) da média;

4º teste: observar pelo menos 8 pontos seguidos de um mesmo lado (acima ou abaixo) da média.

Um processo será caracterizado como estando fora de controle sempre que for observado um desses quatro padrões.

### 3.3 ESCOLHA DAS ESTRATÉGIAS

#### 3.3.1 Avaliação das estratégias na amostra inicial

Após a montagem dos gráficos de controle dos resíduos encontrados, serão realizadas as etapas de construção e teste das estratégias. Visando produzir uma análise mais abrangente, serão considerados como possíveis limites não apenas os limites de 3-sigma, mas também frações destes, múltiplas de 0,5 sigma, inclusive 0 sigma (o que equivale a ter como limite a própria média) e o valor zero (que representa um limite que é ultrapassado quando a diferença entre o valor observado e o valor estimado muda de sinal). Assim, os dias cujos resíduos corresponderem a uma distância, em relação à linha central (LC), maior do que um certo valor (0,0; 0,5; 1,0; 1,5; 2,0; 2,5 ou 3-sigma) ou forem diferentes de zero, serão considerados possíveis pontos de compra (bons momentos para compra) ou venda (bons momentos para venda).

Formalmente, estarão sendo consideradas as seguintes regras:

→ **Para Compra:**  $LnRD_t - Ln\hat{R}D_t < LI$  ou erro de previsão do modelo  $< LI$ ;

→ **Para Venda:**  $LnRD_t - Ln\hat{R}D_t > LS$  ou erro de previsão do modelo  $> LS$ ;

Onde:

LI = limite inferior (distância da LC, especificada de acordo com a estratégia de interesse);

LS = limite superior (distância da LC, especificada de acordo com a estratégia de interesse);

$$\text{Ln}\hat{RD}_t = \text{constante} + \rho \text{Ln} RD_{t-1}.$$

Partindo dos dados relativos à amostra inicial (julho/94 a outubro/02), as estratégias serão testadas através da simulação de operações de compra sempre que o limite inferior for ultrapassado (com um retardo de um dia útil, ou seja, as compras são realizadas com um dia de atraso em relação à sua detecção – D+1), e de venda quando o limite superior também for ultrapassado (com o mesmo atraso para efetivação da operação). Para o cálculo dos resultados, utilizar-se-á como cotação de compra ou venda a cotação média do dia da operação (D+1).

Inicialmente serão avaliados os casos onde os limites de compra são inferiores ou no máximo iguais aos limites de venda (RESTR1), tratando-se de um conjunto de 105 estratégias para teste (combinações 2 a 2 de todas as distâncias anteriormente citadas, que satisfazem RESTR1). Num segundo momento, no entanto, expandir-se-á a análise para os demais casos, ou seja, incluindo os casos onde o limite de compra é superior ao de venda (RESTR2). Apesar de, a princípio, este não parecer um conjunto de estratégias relevantes, também não parece razoável excluí-lo da investigação.

Por fim, serão testadas as estratégias com os mesmos limites anteriormente descritos, mas com a introdução de uma restrição: a venda só poderá ser efetuada quando a cotação de venda superar a de compra (RESTR3 e RETR4, respectivamente).

Para a apuração final dos lucros ou prejuízos, apenas uma exceção será considerada: quando houver operações em aberto no último dia útil do período de análise. Neste caso, se houver uma ação “encarteirada” no penúltimo dia útil de apuração, será efetuada uma operação de venda, de modo que cada estratégia leve sempre a uma posição “zerada” ao final da amostra. Similarmente, mesmo que a ferramenta aponte um momento favorável para compra no penúltimo dia do período em avaliação, este sinal não será considerado.

Conforme indicado anteriormente, a cada compra estará associada uma cotação média do papel; da mesma forma, para cada venda também haverá uma cotação média. O resultado de cada operação corresponderá à razão entre

estas duas cotações menos uma unidade, ou seja,  $\frac{CV}{CC} - 1$ , onde:

CV = cotação média de venda;

CC = cotação média de compra.

O resultado final de cada estratégia corresponderá ao produtório das razões de cada operação simulada menos uma unidade (ou seja, apesar de o resultado ser calculado conforme a fórmula  $\frac{CV}{CC} - 1$ , o resultado final será calculado como o produto de todas as razões  $\frac{CV}{CC}$  e este produto é que é reduzido de uma unidade).

Para cada uma das 105 estratégias iniciais e as 287 adicionais, será gerado um conjunto com o resultado final referente a dez ações e dois índices. No entanto, comparar os resultados encontrados sem algum tratamento pode ser precipitado, uma vez que este conjunto, aparentemente, ainda não representará um grupo comparável, já que englobará diferentes ações com diferentes riscos.

No sentido de tornar os dados comparáveis, será apurado o retorno escalonado de cada papel. Este corresponde à divisão do resultado final, de cada estratégia e para cada papel, pelo desvio-padrão do seu retorno diário. O retorno diário, por sua vez será calculado, como se cada operação de compra durasse, no máximo, um dia útil. Assim, sempre que uma ação permanecer “comprada” por mais de um dia útil, ou seja,  $n$  dias úteis, para efeito de cálculo de desvio-padrão do resultado diário, serão consideradas  $n$  pequenas operações de compra e venda (efetuadas pelas cotações médias) com duração de um dia.

Por fim, para cada estratégia teremos gerado uma amostra com 10 observações (correspondente às dez ações sob análise) de retornos escalonados. Para cada uma destas amostras serão calculados uma média e um desvio-padrão dos retornos escalonados. Com base nestas estatísticas, poder-se-á eliminar estratégias não adequadas segundo o critério de média e variância.

Neste cálculo não serão considerados os índices de preços, pois não correspondem a uma opção de investimento real no mercado à vista, não devendo fazer parte da montagem da amostra proposta.

O critério de média e variância, que vale para investidores com comportamento de aversão a risco (comportamento racionalmente esperado) e quando a distribuição de probabilidades dos retornos for normal, assegura que a esperança de utilidade de um ativo é superior à de outro ativo, e, portanto, o primeiro é preferível ao segundo, sempre que a esperança do retorno do primeiro for não inferior à do segundo e a variância do primeiro for menor do que a do segundo.

Assim, com a aplicação desse critério, garante-se que, para cada nível de risco (desvio-padrão), serão descartadas as estratégias com os menores retornos e, para cada nível de retorno, as estratégias com maiores níveis de risco. Ou seja, dado um nível de risco  $\sigma_0$ , mantém-se apenas a estratégia de maior retorno,  $Ret^*$ . Adicionalmente, descartam-se, também, as estratégias com  $\sigma > \sigma_0$  e  $Ret \leq Ret^*$ , conjuntamente. Este procedimento deverá resultar numa fronteira eficiente, com um conjunto bastante reduzido de estratégias.

### 3.3.2 Avaliação das estratégias na amostra de controle

Para a avaliação das estratégias na amostra de controle, serão estudados os dados referentes ao período de novembro/2002 a abril/2003. Com um conjunto das melhores estratégias, obtido segundo a metodologia da seção anterior, poder-se-iam fazer dois tipos de análise na amostra de controle (últimos seis meses): uma referente ao desempenho apenas das estratégias já apontadas como tendo sido as melhores, ou outra que comparasse todo o conjunto das 105 estratégias originais com seu resultado na amostra de controle.

No primeiro caso, o procedimento a ser adotado consiste em estudar, durante o período de novembro/02 a abril/03, o retorno escalonado médio e o desvio-padrão das estratégias que compõem a fronteira eficiente calculada com base na amostra inicial (de julho/94 a outubro/02).

No segundo caso, a idéia consiste em averiguar se a aplicação das estratégias tanto na amostra inicial quanto na amostra de controle leva aos mesmos resultados. Assim, para operacionalizar esta comparação, serão apresentados dois gráficos de dispersão, um para retorno escalonado médio e outro para o desvio-padrão de cada estratégia, ambos contendo tanto os valores calculados a partir da amostra inicial (eixo X) e quanto os valores obtidos com o uso da amostra de controle (eixo Y). Se, depois de ordenar as estratégias segundo o seu desempenho (em cada período), a posição de cada estratégia, nesta ordenação, não variar muito da amostra inicial para a amostra de controle, ou ainda, se o coeficiente de correlação entre os dois períodos (para cada um dos gráficos) for elevado, isto irá sugerir que o comportamento relativo das estratégias não sofreu grandes alterações entre os dois períodos.

Cumprida esta análise, poderemos separar as estratégias que apresentaram o melhor desempenho em ambos os períodos, e fazer uma comparação dos possíveis ganhos que estas poderiam ter proporcionado para um investidor, em comparação a outras opções de mercado.

### 3.4 ANÁLISE DOS RESULTADOS ENCONTRADOS

Uma primeira hipótese a ser averiguada é a respeito da necessidade de se aplicar os gráficos de controle a uma série de resíduos. Aplicar a metodologia do gráfico de controle de Shewhart a uma série de retornos parece bem mais atraente do que passar pelas etapas de cálculo do melhor modelo ARIMA (p,d,q) ou de alisamento exponencial. Para apurar a necessidade de implementação de toda a metodologia apresentada nas seções anteriores, vale a pena comparar seus resultados com os gerados pela simples aplicação das 105 estratégias sobre a série original de ln RD. Para tanto, todo o procedimento de cálculo de retorno escalonado e escolha das estratégias mais eficientes será replicado a um gráfico de controle de Shewhart (plotado com os dados de ln RD). Assim, para cada estratégia, poderemos identificar o número de ocorrências de resultados eficientes (situados na fronteira eficiente obtida como descrito em 3.3.1).

Mas, mesmo que o uso de técnicas de construção de gráficos de controle para dados autocorrelacionados mostre-se mais eficiente quanto à identificação de possíveis estratégias de escolha de momentos de compra e venda de ações, ainda assim não é possível concluir sobre a importância dos resultados alcançados se não compararmos estes às demais opções disponíveis ao investidor.

Por se tratar de uma metodologia onde não foi considerada a montagem de carteiras de ações, seria precipitado comparar os resultados encontrados com outras metodologias como a de delimitação de fronteira eficiente proposta por Markowitz. Esta deve ser uma questão a ser estudada por pesquisadores interessados no assunto, mas não fará parte do escopo deste trabalho.

Assim, acredita-se que é possível, neste momento, avaliar a eficiência da metodologia aqui proposta com uma comparação entre os seus resultados e os obtidos por outros fundos de ações. Para tanto, serão pesquisados, junto à Comissão de Valores Mobiliários (CVM), os valores das cotas dos fundos de ações dos cinco maiores bancos privados do país: Bradesco, Itaú, Unibanco, Real e Santander. O ideal seria realizar uma comparação para os dois períodos amostrais, mas a CVM só dispõe de dados para estes fundos a partir de março de 2002. Desta forma, a comparação só será possível para o período de nov/02 a abr/03.

A metodologia, neste caso, consistirá em apurar, junto aos bancos apontados, quais fundos de ações estariam disponíveis para o investidor aplicar

seu capital. Com este conjunto, serão levantados, junto à CVM, os valores de suas cotas diárias para o período destacado. Com estes dados montar-se-á uma base de dados, sobre a qual serão adotados os mesmos procedimentos envolvidos no processo de escolha das estratégias, ou seja, serão calculados os dados de retorno escalonado (considerando a rentabilidade total para o período e o desvio-padrão dos retornos diários) para cada fundo. Assim acredita-se ser capaz de gerar um conjunto de informações comparáveis às levantadas para as estratégias.

Por fim, serão apresentadas as rentabilidades bruta (conforme a fórmula de resultado final apresentada neste capítulo) e líquida (rentabilidade bruta descontada dos custos de investimento, a serem detalhados na seção 4.4.3) de cada opção (investimento em bolsa com o uso das estratégias ou aplicação em fundos de investimento) para efetiva avaliação dos ganhos possíveis em cada um dos casos.