

FUNDAÇÃO GETÚLIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

SHEYLA CRISTINA DOS SANTOS BRITO

**COMPORTAMENTO DE PARES DE AÇÕES NO MERCADO BRASILEIRO SOB A
ÓTICA DA COINTEGRAÇÃO, PARA PREÇOS INTRA-DIÁRIOS**

SÃO PAULO
2011

SHEYLA CRISTINA DOS SANTOS BRITO

**COMPORTAMENTO DE PARES DE AÇÕES NO MERCADO BRASILEIRO SOB A
ÓTICA DA COINTEGRAÇÃO, PARA PREÇOS INTRA-DIÁRIOS**

Dissertação apresentada à Escola de
Economia da Fundação Getúlio Vargas
(FGV/EESP) como requisito para
obtenção do título de Mestre em
Finanças e Economia Empresarial.

Orientador: Prof. Dr. Afonso de Campos
Pinto

SÃO PAULO
2011

SHEYLA CRISTINA DOS SANTOS BRITO

**COMPORTAMENTO DE PARES DE AÇÕES NO MERCADO BRASILEIRO SOB A
ÓTICA DA COINTEGRAÇÃO, PARA PREÇOS INTRA-DIÁRIOS**

Dissertação apresentada à Escola de
Economia da Fundação Getúlio Vargas
(FGV/EESP) como requisito para
obtenção do título de Mestre em
Finanças e Economia Empresarial.

Data da Aprovação: ____/____/____

Banca Examinadora:

Prof. Dr. Afonso de Campos Pinto
(Orientador)
FGV-EAESP

Prof. Dr. Rafael Felipe Schiozer
FGV-EAESP

Prof. Dr. Roberto Anis Calfat
INSPER-SP

Prof. Dr. Ricardo Ratner Rocha
FGV-EAESP

Brito, Sheyla C. S..

Comportamento de pares de ações no mercado brasileiro sob a ótica da cointegração, para preços intra-diários / Sheyla Cristina dos Santos Brito. - 2011.

47 f.

Orientador: Afonso de Campos Pinto

Dissertação (MPFE) - Escola de Economia de São Paulo.

1. Ações (Finanças) -- Brasil. 2. Operações com pares (Finanças). 3. Bolsa de valores -- Brasil. 4. Cointegração. I. Pinto, Afonso de Campos. II. Dissertação (MPFE) - Escola de Economia de São Paulo. III. Título.

CDU 336.763.21(81)

AGRADECIMENTOS

Agradeço enormemente a todas as pessoas que me ajudaram a completar mais este desafio, principalmente ao Prof. Dr. Afonso de Campos Pinto por fornecer toda a orientação que foi necessária, e por me ajudar a equilibrar ansiedade e direcionar esforços.

Dedico especial agradecimento ao especialista em mercado financeiro e amigo Alberto Luis de Souza Araújo, por dividir sua larga experiência e contribuir nas análises apresentadas nesta dissertação.

Agradeço à minha família e amigos, e especialmente ao Guilherme pela compreensão nos longos períodos em que me dediquei a este trabalho.

As últimas etapas do estudo tiveram ainda o imprescindível apoio do Prof. Dr. Ricardo Ratner Rocha, que contribuiu com sua revisão e acompanhamento, a quem agradeço fortemente.

RESUMO

Esta dissertação estuda o movimento do mercado acionário brasileiro com o objetivo de testar a trajetória de preços de pares de ações, aplicada à estratégia de *pair trading*. Os ativos estudados compreendem as ações que compõem o Ibovespa e a seleção dos pares é feita de forma unicamente estatística através da característica de cointegração entre ativos, sem análise fundamentalista na escolha.

A teoria aqui aplicada trata do movimento similar de preços de pares de ações que evoluem de forma a retornar para o equilíbrio. Esta evolução é medida pela diferença instantânea dos preços comparada à média histórica. A estratégia apresenta resultados positivos quando a reversão à média se efetiva, num intervalo de tempo pré-determinado.

Os dados utilizados englobam os anos de 2006 a 2010, com preços intra-diários para as ações do Ibovespa. As ferramentas utilizadas para seleção dos pares e simulação de operação no mercado foram MATLAB (seleção) e Streambase (operação). A seleção foi feita através do Teste de Dickey-Fuller aumentado aplicado no MATLAB para verificar a existência da raiz unitária dos resíduos da combinação linear entre os preços das ações que compõem cada par. A operação foi feita através de *back-testing* com os dados intra-diários mencionados.

Dentro do intervalo testado, a estratégia mostrou-se rentável para os anos de 2006, 2007 e 2010 (com retornos acima da Selic). Os parâmetros calibrados para o primeiro mês de 2006 puderam ser aplicados com sucesso para o restante do intervalo (retorno de Selic + 5,8% no ano de 2006), para 2007, onde o retorno foi bastante próximo da Selic e para 2010, com retorno de Selic + 10,8%. Nos anos de maior volatilidade (2008 e 2009), os testes com os mesmos parâmetros de 2006 apresentaram perdas, mostrando que a estratégia é fortemente impactada pela volatilidade dos retornos dos preços das ações. Este comportamento sugere que, numa operação real, os parâmetros devem ser calibrados periodicamente, com o objetivo de adaptá-los aos cenários mais voláteis.

Palavras-chave: *pair trading*, *spread*, *trading* quantitativo, arbitragem estatística, reversão à média, cointegração, simulação histórica, *back-testing*, estacionaridade, MATLAB, Streambase, mercado acionário brasileiro, intra-diário.

ABSTRAT

This dissertation is focused on studying the Brazilian stock market behavior, more specifically related to a pair trading strategy. The assets included in here come from listed stocks of Brazilian Stock Exchange Index (Ibovespa) and the pair selection is exclusively based on a statistic characteristics, known as cointegration, without fundamentalist analysis.

The applied theory treats similar movement of stock prices between pairs which tends to revert to an equilibrium mean of price differences. The strategy will present positive returns when reversion occurs in a pre-defined time.

Back-testing data comprises intraday prices from 2006 until 2010 of Ibovespa stocks. The tools in which pair selection and trading rules are coded are MATLAB (selection) and Streambase (trading). Selection is processed through Dickey-Fuller augmented test into MATLAB to check the existence of a unit root on an error time series of a linear combination of stock prices, for each pair. Operation is simulated through intraday *back-testing* data as mentioned, input into Streambase tool.

Within back-testing period, the strategy results are profitable in 2006, 2007 and 2010. Parameters, to enter and stop the operation, were adjusted for the first month of 2006 and could be successfully applied for the whole year of 2006 (yield of Selic + 5.8% for 2006), for 2007, where yield were close to Selic and for 2010, with yield of Selic + 10.8%. In periods of high volatility (2008 and 2009), tests with the same parameters of the ones adjusted for 2006 generated losses, showing the strategy is highly impacted per volatility returns of stock prices. This behavior suggests that, in actual operations, parameters should be constantly reevaluated in order to adapt them to volatile scenarios.

Keywords: pairs trading, spread, quantitative trading, statistical arbitrage, mean reversion, cointegration, historical simulation, back-testing, stationarity, MATLAB, Streambase, Brazilian stock market, intraday.

SUMÁRIO

1. Introdução	1
2. Revisão bibliográfica	4
3. Metodologia	8
3.1. Dados utilizados e tratamento	9
3.2. Formação do pares	11
3.3. Seleção dos pares.....	12
3.3.1. Cointegração	12
3.3.2. Modelagem de Dickey-Fuller	13
3.4. Parâmetros da regressão linear entre as séries de preços - OLS.....	14
3.4.1. Utilização dos parâmetros encontrados, no período de trading.....	14
3.5. Fluxograma de operação	16
3.6. Metodologia para apuração dos resultados	19
3.7. Custos operacionais.....	24
4. Apresentação dos resultados	25
5. Conclusão e sugestão de pesquisas futuras	32
6. Bibliografia	34

1. INTRODUÇÃO

O objetivo deste trabalho é o de testar a trajetória de preços de pares de ações no mercado brasileiro, identificados de forma estatística e controlados através de ferramenta de *trading* automático. Este sistema simula a compra e venda de ativos à medida que alguns parâmetros, previamente calculados com relação aos preços intra-diários dos mesmos ativos, são atingidos. A eficiência desta estratégia será avaliada em comparação ao CDI. As ações que compõem o Ibovespa representam o universo estudado e a seleção dos pares é feita através da característica de cointegração entre ativos.

Este modelo está associado à literatura principalmente no contexto de arbitragem, um dos conceitos fundamentais em finanças. Arbitragem é definida como “a compra e venda simultânea do mesmo ativo, ou ativos similares, em dois mercados distintos, para obter diferenças vantajosas de preços” (Sharpe e Alexander, 1990). A contribuição que este trabalho pretende adicionar à literatura está na utilização de dados intra-diários com frequência de 15 em 15 minutos, no mercado brasileiro. Trabalhos similares (Kabashima, 2005 e Longo, 2008) tratam do mesmo assunto, porém utilizam-se de dados de fechamento dos preços das ações.

Oportunidades para estratégias de arbitragem sem risco, no sentido mais estrito, que se aproveitam dessas diferenças de preços em mercados ou tempos distintos, tem tornado-se cada vez mais escassas, de acordo com analistas de mercado. Com a automação incorporada às operações de Bolsas de valores ao redor do mundo (com os *automated trading systems*), as barreiras regionais desapareceram e os agentes tem acesso a ativos que antes estavam restritos aos investidores locais. A velocidade em que as ordens de compra e venda são disparadas por esses sistemas rapidamente elimina as diferenças de preços. Consequentemente, as oportunidades de retornos são reduzidas, e em muitos casos, geram prejuízos aos agentes que precisam acumular grandes volumes de operações, por conta das características do mercado aqui descritas (Dewan e Mendelson, 1998).

Assim surgem as variantes da arbitragem pura, que a teoria classifica como livre de risco. E as estratégias que se desenvolveram a partir deste momento embutem um grau de risco ao investidor, mas tentam estabelecer algum controle sobre o comportamento do *portfolio* (Vidyamurthy, 2004).

Neste contexto, o *trading* estatístico-quantitativo aparece como alternativa para a implementação de estratégias que incorporam cálculos sobre o comportamento passado dos ativos, e disparam ordens de compra e venda sem a intervenção humana, dada uma regra pré-estabelecida pelos *traders* (Ernest Chan, 2009).

O *Pair Trading* representa uma das possíveis variantes da arbitragem, intitulada de estratégia quantitativa neutra, sempre retomando a idéia de que não se trata de um *portfolio* sem risco, mas teoricamente neutro ao mercado. Segundo Vidyamurthy (2004), este *portfolio* neutro é construído utilizando-se dois ativos, com posições compradas e vendidas, numa razão predeterminada. Analisando-se a evolução deste *portfolio*, a medida mais comum para quantificar esta posição instantânea é o *spread*. Essa diferença é calculada pelos preços dos dois ativos num determinado momento, levando-se em conta a razão média entre eles, e formando uma série de tempo. O *spread* é relacionado ao componente de retorno residual. *Pair trading* envolve a abertura de posições quando o *spread* está substancialmente longe da média, com a expectativa que este irá reverter ao valor previamente calculado. Nesse momento, a posição é revertida na convergência.

Paralelamente, arbitragem estatística é baseada na idéia de preços relativos. A premissa está no fato de que ações com características semelhantes deverão ter preços com comportamentos similares. O *spread* pode ser entendido como o grau de erro na precificação. Ou seja, quanto maior o *spread*, maior a magnitude em que os preços estão deslocados da sua média e, portanto, maior o potencial de lucro. Esta estratégia, como mencionado, abre posições compradas e vendidas quando o *spread* se desloca da média, com a expectativa que os preços tenderão a se corrigir.

Analisando-se os estudos já efetuados sobre *Pair trading*, Nunzio Tartaglia é citado como precursor desta estratégia. De acordo com Vidyamurthy (2004), em meados de 1980, o analista quantitativo de Wall Street organizou um time de físicos, matemáticos, e engenheiros de computação para trabalhar sobre uma oportunidade de arbitragem nos mercados de ações. O grupo de Tartaglia utilizou métodos estatísticos sofisticados para desenvolver programas que operavam no mercado financeiro. À época, os programas já utilizavam a tecnologia de sistemas que rodam automaticamente, e que eliminam a intuição e as habilidades dos *traders*, substituindo-as por disciplina e filtros com regras consistentes. Os programas de Tartaglia identificaram pares de ações cujos preços tendiam a mover-se de forma similar. O

grupo operou com sucesso com esses pares em 1987, ano em que o grupo apurou lucros da ordem de 50 milhões de dólares para o Morgan Stanley Group. Apesar de o time ter se desfeito após alguns anos, *pair trading* tornou-se uma estratégia bastante popular, conhecida como neutra ao mercado, utilizada por *traders* independentes e de grandes instituições, assim como *hedge funds*.

O presente trabalho aplica a seleção dos pares cointegrados através do teste de Dickey-Fuller aumentado e utiliza-se de dados intra-diários, em intervalos de 15 em 15 minutos para testar a estratégia de *pair trading*, abrindo posições compradas e vendidas à medida que o *spread* se desloca da média. É então testada a tendência teórica de correção dos preços, apurando lucro à estratégia.

No texto que se segue, serão apresentados: a revisão bibliográfica, no capítulo 2, onde o trabalho é contextualizado com relação à arbitragem, *pair trading*, teoria de sistemas cointegrados, *trading* de alta frequência, assim como a técnica utilizada para obtenção dos dados. O capítulo 3 trata da Metodologia detalhada para seleção de dados e forma de estruturação para obter-se os resultados. Na sequência, o capítulo 4 apresenta os resultados obtidos, assim como os retornos que a técnica proposta estima, para o período de *back-testing* analisado. Finalmente, no capítulo 5, são apresentadas as conclusões, assim como as oportunidades para trabalhos futuros neste assunto, respectivamente.

2. REVISÃO BIBLIOGRÁFICA

Neste capítulo, o trabalho é contextualizado com relação ao histórico de *pair trading* e arbitragem na literatura, são apresentadas a teoria de sistemas cointegrados e uma abordagem sobre *trading* de alta frequência, assim como a técnica de Dickey-Fuller, utilizada na aplicação de seleção dos dados.

Na literatura sobre *Pair trading*, são encontradas diversas metodologias com o objetivo de identificar pares com comportamento similares, seja nos retornos, seja nos preços. O objetivo é idêntico: aproveitar os deslocamentos do que seria o equilíbrio e sua reversão.

Estudos fundamentalistas sobre as características das empresas são largamente utilizados, na estratégia em questão, para a seleção dos pares, mais frequentemente nas operações com modelos de médio e longo prazo.

Ehrman (2006) detalha a metodologia de seleção de pares através da modelagem dos retornos baseada em cinco categorias de parâmetros: lucratividade, preço, liquidez, alavancagem e eficiência. Alguns possíveis parâmetros seriam: margem de lucro líquido, retorno sobre patrimônio (ROE), ativos líquidos totais divididos pelo número total de ações no mercado, dívida (passivos sobre ativos), entre outros. Processa-se então a regressão linear do histórico dos retornos de um determinado ativo versus os parâmetros fundamentalistas e, com base nos resultados, estima-se os retornos futuros. Faz-se então a ordenação dos ativos e, a partir desse momento, é possível montar os pares. As posições compradas são abertas para os ativos de maiores retornos esperados, e as posições vendidas, para os ativos de menores retornos esperados.

Neste trabalho, é utilizada metodologia de seleção de pares utilizando métodos unicamente quantitativos. Nesta linha, Engle e Granger (1987) observaram o seguinte comportamento: assumimos que duas series temporais, aqui se tratando de duas series de preços de ações, sejam não estacionárias; teoria bastante aceita na literatura. Apesar de individualmente não estacionárias, é possível que uma combinação linear específica de duas séries seja estacionária, ou seja, duas séries mover-se-iam juntas num intervalo de tempo. Engle e Granger usaram o termo cointegração e propuseram tal idéia em seu artigo em 1987, cujo reconhecimento aconteceu em 2003, ao ganharem o Prêmio Nobel de Economia, por esta e outras idéias.

Stock e Watson (1988) abordam também o tema “cointegração”, porém através do modelo de tendências comuns. Em seu *paper* “*Testing for Common Trends*”, os autores detalham de que forma a combinação linear anula o efeito de tendência embutido em duas séries de tempo, desde que esse efeito seja idêntico nas duas séries. A modelagem proposta estabelece que cada série de preço de ação é composta por um componente de tendência e outro estacionário. Em sistemas cointegrados, o componente de tendência é mesmo, como mencionado anteriormente.

Vidyamurthy (2004) discute extensivamente a aplicação de cointegração e apresenta também as duas formas tradicionais de identificação de pares para posterior montagem de estratégia de *trading*: a fundamentalista e a estatística. A abordagem fundamentalista detalhada por Vidyamurthy (2004) baseia a seleção de pares na teoria de precificação por arbitragem (APT); assim, o retorno esperado num determinado período tende a ser similar para ativos expostos aos mesmos fatores de risco.

O mesmo autor apresenta modelos estatísticos para identificação de pares. Primeiramente, segue a linha de pre-seleção através de medidas de correlação. Dada esta lista reduzida de pares, segue-se para verificar se o sistema é cointegrado. Os passos para esta checagem compreendem primeiramente estimar-se a relação linear entre duas séries de preços e medir-se o grau de reversão à média. Posteriormente testa-se a estacionariedade do resíduo histórico. Contudo, o autor restringe-se a somente citar a teoria de reversão à média para identificação dos pares cointegrados, e estabelece a frequência de “*zero-crossing*” como uma medida intuitiva para selecionar um determinado par. Não são apresentadas métricas para um filtro claro entre pares cointegrados e não cointegrados. De acordo com Vidyamurthy (2004), a frequência em que a série dos resíduos cruza o zero, ou seja, quando ocorre a reversão completa à média, é utilizada como parâmetro para identificar se um determinado par de ações poderia compor uma estratégia de *trading*.

Com maior rigor estatístico, uma das formas de testar a estacionariedade das séries é através do método de Dickey-Fuller (Enders, 2004), que verifica a existência da raiz unitária da série de resíduos da combinação linear do logaritmo dos preços.

David A. Dickey e Wayne A. Fuller (1981) construíram a função da razão de probabilidade para a hipótese nula onde o modelo verdadeiro é um passeio aleatório com média zero, para um processo estacionário ARMA (*auto-regressive moving*

average). Sendo assim, num determinado teste, ao rejeitar-se a hipótese nula, admite-se que o modelo segue um passeio aleatório de média zero. Dickey & Fuller aplicaram a distribuição de Fisher para encontrar os valores críticos e os calcularam para diferentes tamanhos de amostras e probabilidades. Esta foi a base teórica selecionada para os testes apresentados neste trabalho, com seus valores detalhados nas seções seguintes.

No *paper* de Said e Dickey (1984), “*Testing for unit roots in autoregressive-moving average models of unknown order*”, os autores generalizam as hipóteses estudadas por Dickey & Fuller (1979) provando que o teste continua válido mesmo nos cenários em que o grau de auto-regressão seja desconhecido. O trabalho assume que estes parâmetros de defasagem do modelo, chamados de p e q , são iguais a 1 e concluem que os valores críticos encontrados por Dickey & Fuller (1979), para valores de p e q conhecidos, continuam válidos.

Tratando-se dos aspectos mais operacionais para montagem da estratégia e execução do *trading*, Vidyamurthy (2004) propõe uma banda definida a partir de uma série *white noise* Gaussiana para se determinar os pontos de entrada e saída. Segundo modelo ótimo de análise, são apresentados os valores de desvio Δ do *spread* em relação à sua média, para os quais existe maior probabilidade de reversão. Através da maximização da função em relação a Δ , o valor ótimo obtido é de $0,75 \sigma$ (ou seja, 0,75 vezes o desvio calculado da série de tempo do *spread*).

Estratégias de *pair trading* tem sido extensivamente testadas nos diferentes mercados ao redor do mundo, incluindo o Mercado Brasileiro. Kabashima (2005) desenvolveu trabalho no Mercado Brasileiro para o período de 1993 a 2005, no universo das ações do Índice Bovespa. Sua tese aplica a metodologia de Vidyamurthy (2004) quase que em sua totalidade, utiliza correlação para pré-seleção dos pares, assim como teste de Dickey-Fuller para seleção final, verificando a cointegração da série de preços em pares de ações.

Com relação ao estudo de Kabashima (2005), o presente trabalho acrescenta a abordagem de alta frequência, aplicando a metodologia de cointegração ao *pair trading* intra-diário.

Nesse sentido, Ernest Chan (2009) aborda o *trading* de alta frequência. Para a estratégia de pares, são apresentadas técnicas já mencionadas anteriormente nesta

revisão, como correlação e cointegração. Ernest aborda a importância de se incorporar custos de transação nas simulações de *back-testing* para operação de alta frequência, tratada como grande diferencial na contabilização de resultados, quando comparada a operações de mais longo prazo. Este tratamento é mencionado como crucial para que prejuízos não sejam subestimados por cenários sem custo, inviáveis nos mercados reais.

A segunda contribuição de Chan (2009) está no detalhamento da metodologia a ser aplicada, principalmente na identificação de pares. São apresentadas rotinas estruturadas na linguagem MATLAB, com a teoria de Dickey-Fuller desenvolvida.

O presente trabalho estende as idéias citadas, em dois aspectos. Em primeiro lugar, as simulações e resultados apresentados a seguir trabalham com a identificação de pares de forma apenas estatística, ou seja, não são incorporados aspectos fundamentalistas na formação dos pares que alimentarão a estratégia, como se verifica em alguns dos estudos pesquisados. É utilizada teoria de cointegração para a classificação e seleção dos pares, de acordo com os preceitos propostos por Ernest P. Chan (*Quantitative Trading*, 2009). O segundo aspecto refere-se à periodicidade dos dados. Enquanto a maioria dos estudos que analisa a aplicação de cointegração para *pair trading* trabalha com dados de fechamento diário, o presente estudo trata de dados intra-diários, no intervalo de 15 em 15 minutos, aproximando a análise das aplicações do *High-frequency trading*.

O capítulo a seguir trata da metodologia aplicada neste trabalho, desde a seleção e tratamento dos dados, formação e seleção de pares, assim como as regras aplicadas no período de *trading*, todos simulados através de dados de back-testing. Será apresentada ainda a forma de cálculo dos retornos para os diferentes períodos analisados.

3. METODOLOGIA

Neste capítulo, são apresentadas as metodologias de: seleção e tratamento dos dados, formação e seleção de pares, assim como as regras aplicadas no período de *trading*, todos simulados através de dados de *back-testing*. Será apresentada ainda a forma de cálculo dos retornos para os diferentes períodos analisados.

O modelo utilizado neste trabalho tem o objetivo de selecionar e testar pares de ações, utilizando a estratégia que mantém quantidade comprada de 1 das ações de cada par e vendida na outra, tomando-se as ações listadas no Ibovespa. A estratégia tem resultado positivo quando a diferença dos preços das ações que compõem o par reverte à média, calculada no período anterior.

O primeiro passo compreende a definição dos períodos de análise, seguidos dos períodos de operação no mercado, de maneira similar a uma média móvel. O período completo é composto de 5 anos (2006 a 2010). Para cada ano, são formados sub-períodos semanais. A estratégia então analisa o comportamento de todas as combinações de pares nas primeiras 10 semanas (50 dias úteis, de acordo com Ernest Chan, 2009) e, com base no teste de Dickey-Fuller, seleciona um sub-conjunto de pares cointegrados. Estes são os pares que entram na estratégia de operação, na semana subsequente. Os parâmetros de *trading*: “média do *spread*”, “desvio padrão” e “razão entre os preços” são obtidos neste mesmo período de 10 semanas de análise e são utilizados na semana de *trading*.

A estratégia, que será detalhada nos itens seguintes deste capítulo, evolui simulando compras e vendas à medida que os preços oscilam. A cada 15 minutos, o algoritmo compara a diferença de preços de cada par com a média definida no intervalo anterior de 10 semanas e dispara ordens de compras e vendas de acordo com a distância desta diferença de preços com relação à média.

A análise dos resultados é feita separadamente para cada um dos 5 anos (de 2006 a 2010). Em cada ano, é calculado o resultado de 10 períodos compostos por 4 semanas cada um. Não foram utilizadas subdivisões de meses, para que o número de dias úteis fosse o mais próximo possível entre os sub-períodos.

Os retornos são calculados considerando o máximo de capital utilizado neste sub-período de 4 semanas (total da posição vendida, que será detalhada no parágrafo

3.6 deste capítulo). Ou seja, se houve um dia de pico em que muitos pares estavam abertos simultaneamente, este é o capital utilizado para o cálculo de retorno.

Os tópicos a seguir detalham cada um dos passos citados nesta introdução.

3.1. Dados utilizados e tratamento

Com o intuito de filtrar o universo de possibilidades e privilegiar a maior liquidez, a montagem do conjunto de dados compreendeu os papéis que compõem o Ibovespa em Março de 2011, totalizando 69 ações. O intervalo contemplado nas análises foi de 5 anos, indo de 2006 a 2010, com preços intra-diários e intervalo de 15 em 15 minutos (fonte Reuters). O Anexo 1 apresenta todos os papéis incluídos no estudo.

O tratamento dos preços para os eventos corporativos (dividendos, juros sobre capital, bonificações, *splits* e *inplits*) foi feito individualmente para cada ano. Dados já tratados, para a fonte utilizada, estavam disponíveis apenas para preços de fechamento, não estando disponíveis para dados intra-diários. Os preços foram então corrigidos da seguinte forma: ao ocorrer uma mudança de capital da empresa ou pagamento de dividendo, os preços anteriores ao evento são multiplicados por um fator (para o caso de *splits*, *inplits* ou bonificação) ou reduzidos de uma fração do preço (para o caso de dividendos ou juros sobre capital):

Splits e inplits:

$$Preço\ ajustado = Preço * Fator \quad (3.1)$$

Dividendos e juros sobre capital:

$$Preço\ ajustado = Preço - Valor \quad (3.2)$$

Sendo:

Fator: fator multiplicador que considera a mudança de capital

Preço: preço cheio, ou seja, anterior ao ex-date (data a partir da qual a ação é operada no mercado sem o direito ao evento).

Preço ajustado: preço reduzido de acordo com o pagamento do evento.

Valor: valor, em moeda nacional, correspondente ao montante pago por ação para dividendos e juros sobre capital.

Assim, esta nova série de preços ajustados garante que o modelo não tenha saltos que não representariam uma oscilação de mercado.

O tratamento para janelas de preços, ou seja, intervalos nos quais não houve operação para uma ou as duas ações que compõem um determinado par, é feito eliminando-se o registro da amostra. Assim, um registro no intervalo de 15 minutos é considerado somente se as duas ações do par tiverem preço disponível, não havendo interpolação para janelas de preços.

A disposição dos dados é feita através de janelas móveis de 10 semanas para o *back-testing* e 1 semana para a operação de *trading*. Ou seja, dadas as 52 semanas de cada 1 ano analisado, e retiradas as primeiras 10 semanas que serão utilizadas unicamente para *back-testing*, tem-se 42 semanas subseqüentes de *trading* para cada ano.

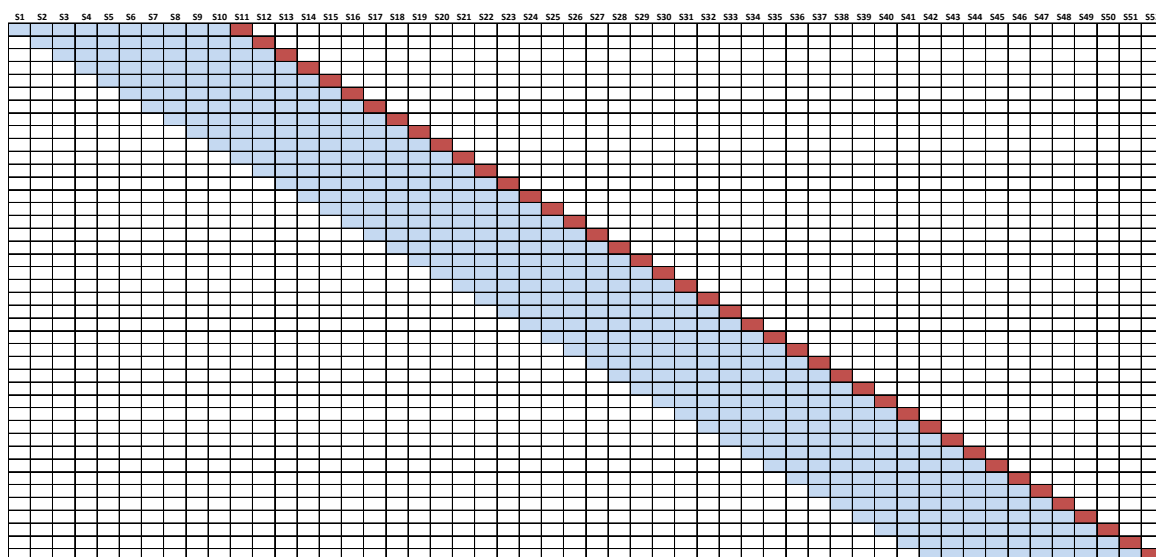


Figura 3-1 - Períodos móveis de seleção de pares e operação

De acordo com a figura 3-1, as semanas marcadas em azul representam os períodos de *back-testing* formados por 10 semanas de dados, enquanto que cada semana subsequente em vermelho representa o período correspondente de *trading*. Para melhor entendimento, uma ampliação do esquema nas primeiras semanas, apresentaria:

S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
0														
-1	0													
-2	-1	0												
-3	-2	-1	0											

Figura 3-2 - Detalhe da Figura 3-1

Finalmente, para cada ano analisado, serão descartadas as 2 últimas semanas para que possam ser compostos 10 grandes intervalos de 4 semanas de *trading*, sendo que, sempre a cada mudança de semana, novos parâmetros de *trading* serão calculados. Esta padronização de 4 semanas foi adotada para comparação dos resultados entre períodos. Não foi adotado um único período contínuo de 42 semanas por facilidade operacional nos softwares escolhidos para a execução da estratégia. A simulação no mercado ocorre dentro de cada um dos 10 intervalos anuais, sendo que todas as posições abertas são encerrados ao final de cada 4 semanas, sendo reabertas no intervalo seguinte.

3.2. Formação do pares

As ações utilizadas nas simulações compreendem o Ibovespa de Março de 2011 para todos os períodos. Ou seja, para cada ano, não necessariamente todas as ações estarão presentes. Para efeitos de simplificação, o conjunto de ações não foi revisto de acordo com o Ibovespa de cada ano.

Dadas as 69 ações do Ibovespa no período selecionado, o número de pares para cada período de 11 semanas (10 semanas de *back-testing*, mais 1 semana de *trading*), será no máximo:

$$\text{Número máximo de pares} = \frac{69 * 68}{2!} = 2346 \text{ pares}$$

Assim, para cada intervalo, serão analisados, no máximo, 2346 pares.

3.3. Seleção dos pares

A modelagem para seleção de pares trabalhou prioritariamente com rotinas em MATLAB. O processo se inicia com a formação de todos os pares possíveis, seguindo a combinação dos ativos dois a dois, para cada período. O grupo de pares alimenta a rotina de cálculo dos parâmetros da modelagem de Dickey-Fuller, filtra aqueles cujo parâmetro apresenta significância estatística para o teste de Fisher, com 99% de confiabilidade. Para efeitos operacionais, são selecionados os 100 primeiros pares, ordenados de acordo com parâmetro de cointegração, desde que estejam acima de 99% de confiabilidade. Ou seja, se o teste Dickey-Fuller selecionar menos de 100 pares, todos serão utilizados no período de *trading*. Se mais de 100 pares forem selecionados, somente os 100 primeiros serão operados.

3.3.1. Cointegração

Como já mencionado neste trabalho, Engle e Granger foram os primeiros a usar o termo cointegração, propondo este conceito em seu artigo em 1987. Esta literatura é base para pesquisas subsequentes que utilizam a idéia de que séries não estacionárias podem ser transformadas em estacionárias através da diferença entre duas séries, desde que estas sejam cointegradas (Vidyamurthy, 2004).

Tomemos duas séries de tempo não estacionárias x_t e y_t . Se para algum valor de γ , $y_t - \gamma x_t$ for estacionária, então as séries podem ser consideradas cointegradas.

A explicação para a dinâmica de cointegração é dada pelo conceito de “correção do erro”. A idéia por trás deste conceito é que sistemas cointegrados tem um equilíbrio de longo prazo, que seria a média da combinação linear das duas séries. Assim, se existe um desvio da média, uma ou ambas as séries se ajustarão de forma a retornar ao equilíbrio.

A correção do erro pode ser representação da seguinte forma:

$$y_t - y_{t-1} = \alpha_y(y_{t-1} - \gamma x_{t-1}) + \varepsilon_{y_t} \quad (3.3)$$

$$x_t - x_{t-1} = \alpha_x(y_{t-1} - \gamma x_{t-1}) + \varepsilon_{x_t} \quad (3.4)$$

Onde ε_{x_t} e ε_{y_t} são os processos de *white noise* das séries x_t e y_t , respectivamente. O lado esquerdo da equação representa o incremento da série de

tempo a cada passo. O lado direito é composto pela parcela de white noise, como mencionado, e pela parcela de correção do erro. A parcela $y_{t-1} - \gamma x_{t-1}$ representa o quanto a série desviou da média no passo $t - 1$, sendo que o coeficiente de cointegração é dado por γ e α_y representa a velocidade de reversão à média.

O próximo passo trata de verificar se duas séries são cointegradas, ponto crucial na identificação dos pares que comporão nossa estratégia de *trading*.

3.3.2. Modelagem de Dickey-Fuller

O presente trabalho aplicará a modelagem de Dickey-Fuller para testar se a série de resíduos da combinação linear dos preços descrita na sessão anterior é estacionária. Esta relação é dada pela equação:

$$z_t = \alpha + \beta z_{t-1} + \varepsilon_t \quad (3.5)$$

O objetivo do teste é verificar se β é diferente de 1, o que significa que a série de resíduos é estacionária. Porém, essa verificação será feita indiretamente. Primeiramente subtrairemos y_{t-1} dos dois lados da equação:

$$z_t - z_{t-1} = \alpha + \beta z_{t-1} - z_{t-1} + \varepsilon_t \quad (3.6)$$

Chamaremos $(\beta - 1)$ de ρ . A equação será então reescrita como:

$$\Delta_t = \alpha + \rho z_{t-1} + \varepsilon_t \quad (3.7)$$

Desta forma, é possível testar-se a Hipótese nula (H_0), ou seja, ρ igual a zero. Sendo $\rho = 0$, $\beta = 1$. Portando, ao rejeitar-se a Hipótese nula, chega-se a situação de estacionariedade da série de resíduos formada pela combinação linear dos preços dos dois ativos que formam o par que estamos testando.

Adicionalmente, pode haver alguma auto-correlação entre preços anteriores a $t - 1$. Para este teste foram utilizados 6 lags para calcular o valor estatístico de ρ , de acordo com a literatura da rotina desenvolvida em MATLAB (LeSage, 2008 na documentação “*Spatial Econometrics*”). LeSage, por sua vez, referencia Engle e Granger, 1987, assim como Dickey e Fuller, 1981, no desenvolvimento das rotinas.

Os valores críticos indicados na literatura para os intervalos de 10%, 5% e 1% de significância são:

	<i>Intervalos de significância</i>		
	1%	5%	10%
Valores críticos	-4.025	-3.404	-3.089

Neste estudo, foi utilizado o valor crítico para 99% de significância (-4.025).

3.4. Parâmetros da regressão linear entre as séries de preços - OLS

Identificados os pares cointegrados da amostra, serão então calculados os parâmetros da regressão linear da série dos preços dos ativos P_t , para cada par, pelo método dos mínimos quadrados ordinários (MQO ou OLS).

A equação que irá modelar o comportamento do erro no período de *back-testing* será:

$$P_{a1_t} = HedgeRatio * P_{a2_t} + Z_t \quad (3.8)$$

Onde as séries de preços dos dois ativos, que compõem o par cointegrado, são representas por P_{a1_t} e P_{a2_t} . *HedgeRatio* é o parâmetro que estabelece a relação linear entre os preços e será usada no período subsequente ao de formação dos pares.

Obtidos, pelo MQO, *HedgeRatio* e Z (sendo este último o vetor de erros para cada dado da amostra), pode-se calcular o valor médio do erro \bar{Z} , assim como seu desvio σ_z .

3.4.1. Utilização dos parâmetros encontrados, no período de trading

Determinados *HedgeRatio*, \bar{Z} e σ_z , no período de *back-testing*, pode-se partir para a aplicação destes, no período de *trading*.

Para cada t (15 em 15 minutos), são calculados:

$$Z_t = P_{a1_t} - HedgeRatio * P_{a2_t} \quad (3.9)$$

$$UpperBand = \bar{Z} + n * \sigma_z \quad (3.10)$$

$$MiddleBand = \bar{Z} \quad (3.11)$$

$$LowerBand = \bar{Z} - n * \sigma_z \quad (3.12)$$

As bandas calculadas: *UpperBand*, *MiddleBand* e *LowerBand* são utilizadas como referência para se identificar a variação do *spread* (Z_t), ou seja, se este parâmetro está aumentando, diminuindo ou está estável com relação à média. Dentro da estratégia de *trading*, Z_t é comparado com a *UpperBand* e com a *LowerBand*, como exemplificado a seguir:

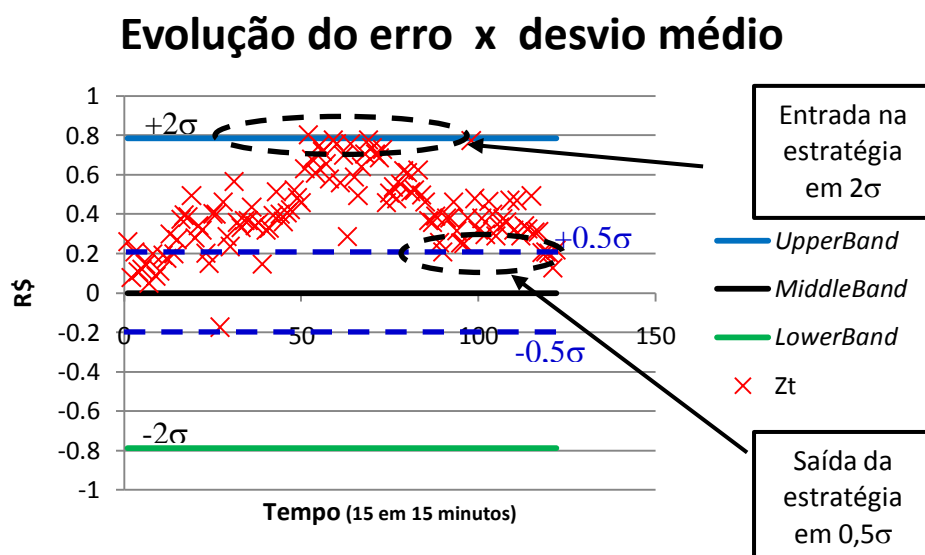


Figura 3-3 - Parâmetros de entrada e saída de estratégia de operação

Como apresentada na figura 3-3, a estratégia de *trading* evolui da seguinte maneira: se o *spread* (Z_t), representado pelos pontos em vermelho no gráfico, ultrapassar a banda superior (*UpperBand*), representada na figura pela linha contínua azul, ao valor de $+2\sigma$, vende-se a ação 1 e compra-se a ação 2.

A situação oposta (não representada na figura), mas também de afastamento do *spread* (Z_t) da média seria aquela em que o Z_t evolui para baixo da banda inferior (*LowerBand*), linha verde, ao valor de -2σ . Neste cenário, compra-se a ação 1 e vende-se a ação 2. As linhas tracejadas da figura exemplificam a saída da estratégia ($\pm 0,5\sigma$), momento de apuração de lucro, visto que significam a reversão parcial à média.

A definição das bandas é feita com base numa fração do desvio do erro. A literatura aponta para utilização de faixas distintas. Para Vidyamurthy, 2004, a distribuição do *spread* obedece à modelagem de uma série *white noise* Gaussiana. A otimização levaria ao valor de $0,75 \sigma_z$.

Neste estudo, fizemos simulações utilizando a fração de $0,75 \sigma_z$, porém este não foi o valor que otimizou o sistema, na estratégia de *trading*. Este resultado pode ser explicado pela menor liquidez e alta volatilidade do mercado Brasileiro quando comparado ao mercado Americano, por exemplo, ambiente para o qual Vidyamurthy, 2004 propôs a modelagem de uma série *white noise* Gaussiana. Seguiu-se, portanto, valores empíricos utilizados pelo mercado, seguindo a experiência de *traders* que operam no mercado Brasileiro. Essa análise levou a considerar $2 \sigma_z$ seguindo estudo de Chui, Lukman, Modarresi e Velayutham, 2011.

As regras de saída abrangem três situações distintas:

- A primeira é aquela na qual o comportamento dos preços das ações que compõem o par em questão segue a expectativa de reversão à média do *spread*. Neste cenário, a ação irá apurar o lucro quando o valor de Z_t atingir $0,5 \sigma_z$, para o cenário em que a estratégia foi aberta em $2 \sigma_z$ ou $-0,5 \sigma_z$, para o cenário em que a estratégia foi aberta em $-2 \sigma_z$.
- A segunda situação é aquela em que, por alguma razão de ruptura no padrão dos preços, ou por um distanciamento temporário grande do *spread* de sua média (o que embutiria um risco muito grande à estratégia caso a posição fosse mantida), o valor de Z_t atinge $\pm 3 \sigma_z$. Neste cenário, as posições são zeradas, mesmo com a apuração de prejuízo.
- Finalmente, cada período de *trading* compreende 4 semanas, com atualização semanal dos parâmetros (média, desvio, bandas e razão entre preços das ações do par). Assim, no intervalo de 4 semanas, um par pode permanecer aberto, mesmo que os parâmetros tenham se alterado (sem rebalanceamento). Contudo, o par é obrigatoriamente fechado ao final do intervalo.

3.5. Fluxograma de operação

A modelagem completa, desde a seleção de pares até a simulação de um ambiente de *trading*, utiliza-se de dois sistemas distintos interfaceados: MATLAB, como já mencionado anteriormente, e o Streambase System. Este último trata-se de um sistema cuja tecnologia é conhecida como CEP (*Complex Event Processing*), sendo capaz de integrar o processamento de dados históricos com dados reais. No presente

trabalho, apesar de todas as análises terem sido feitas com dados históricos, o período de *trading* simulado entra no fluxograma de operação com características similares às de uma operação real. Ou seja, os preços na frequência de 15 em 15 minutos alimentam o sistema em seqüência, como se um provedor de dados de mercado estivesse disponibilizando tais informações em tempo real. O passo a passo do processo para cada ano é descrito a seguir:

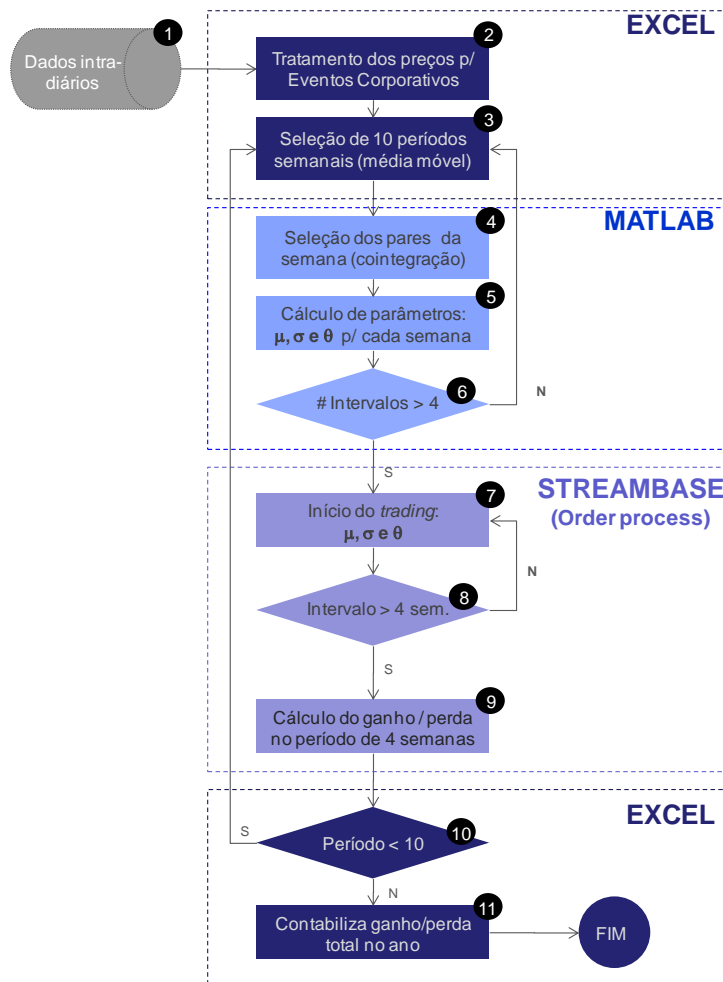


Figura 3-4 - Passo a passo da metodologia

A figura 3-4 apresenta o passo a passo da metodologia, incluindo o tratamento dos dados, seleção dos pares e evolução do *trading*:

Passo 1: Obtenção dos dados para os anos de 2006 a 2010 em fonte de dados de mercado. Neste trabalho foi utilizada fonte Reuters, para os preços de 15 em 15 minutos, dos papéis que compõem o Ibovespa em Março de 2011.

Passo 2: Seleção e armazenamento dos dados em Excel. Tratamento da série de preços para Eventos Corporativos (divulgação de dividendos, juros sobre capital, bonificações, desdobramentos e grupamentos).

Passo 3: Organização dos dados em períodos semanais e seleção de preços em intervalos de 10 semanas. A cada nova rodada, a seleção se desloca em 1 semana para que todos os períodos de cada ano sejam cobertos pela estratégia.

Passo 4: Entrada dos preços das ações, na rotina da ferramenta MATLAB, para o período definido no passo 3, para execução do teste Dickey-Fuller. A saída deste passo é a lista de pares cointegrados.

Passo 5: Também na ferramenta MAT LAB, são calculados os parâmetros: “média do *spread*”, “desvio padrão” e “razão entre os preços” para os pares cointegrados (seleção obtida no passo 4).

Passo 6: Enquanto o número de intervalos de 10 semanas for menor que 4, volta-se ao passo 3, caso contrário, segue-se ao passo 7.

Passo 7: É dado início ao período de *trading* na ferramenta Streambase, que caminha a intervalos de 15 em 15 minutos, por 28 dias corridos. Neste período, a cada novo preço, o algoritmo compara a diferença de preços das ações de cada par com as bandas exemplificadas na Figura 3-3. O algoritmo dispara ordens de compra e venda e armazena os resultados obtidos ao longo dos 28 dias.

Passo 8: O algoritmo volta para o passo 7, até que as 4 semanas sejam percorridas.

Passo 9: Sobre as 4 semanas de operação, são contabilizados, na ferramenta Excel: resultado bruto, resultado líquido, máximo capital investido nas 4 semanas e retorno sobre o investimento.

Passo 10: O algoritmo volta para o passo 3 até que todos os 10 períodos de 4 semanas tenham sido percorridos.

Passo 11: Em Excel, é feita a contabilização dos 10 intervalos de 4 semanas cada um, chegando-se ao retorno total do ano.

3.6. Metodologia para apuração dos resultados

O primeiro passo para definição da metodologia de apuração dos retornos é a definição de como os períodos de *trading* apresentados anteriormente são agrupados, como na figura a seguir.

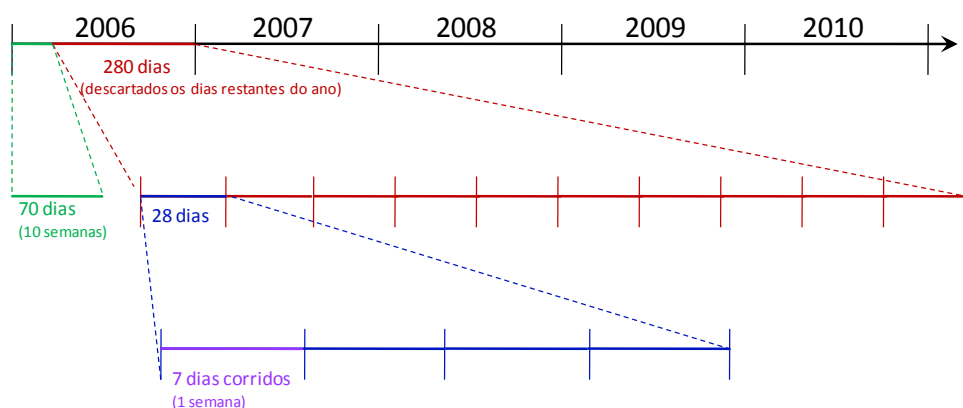


Figura 3-5 - Organização dos períodos para contabilização dos retornos

A figura 3-5 apresenta a subdivisão do período completo de análise, constituído com o propósito de organizar os períodos de: seleção dos pares, *trading* e contabilização dos retornos. Assim, a seleção dos pares dá-se sobre as primeiras 10 semanas (ou 70 dias corridos). Cada uma das semanas subsequentes (7 dias) são períodos de *trading*, tendo sempre as 10 semanas anteriores (média móvel) como referência para: seleção dos pares e cálculo dos parâmetros de *trading*. Os 280 dias da figura representam o período efetivo de *trading* de cada ano. O cálculo dos retornos, sobre estes 280 dias, é subdividido em 10 períodos de 28 dias (segunda régua da figura). É neste intervalo em que se define o máximo de capital aplicado, assim como o resultado absoluto em reais.

A estratégia de *pair trading*, no modelo proposto, apresenta bastante oscilação no montante de capital exigido para sua execução. O investimento direto na compra e venda das ações que compõem o par é próximo de zero, já que haverá um casamento entre posições opostas. Esse valor não é zero em função do momento em que a estratégia é aberta: quando o *spread* está aberto. Assim, o custo direto da operação será o quanto o par está deslocado do que se supõe que seja o equilíbrio, multiplicado pelo número de ações da ação 1. No gráfico a seguir, o deslocamento do *spread* é a distância entre Z e a linha de *Middleband*. Como Z_t é negativo, compra-se a ação 1 e vende-se a ação 2 ($Z_t = P_{a1_t} - P_{a2_t} * HedgeRatio$).

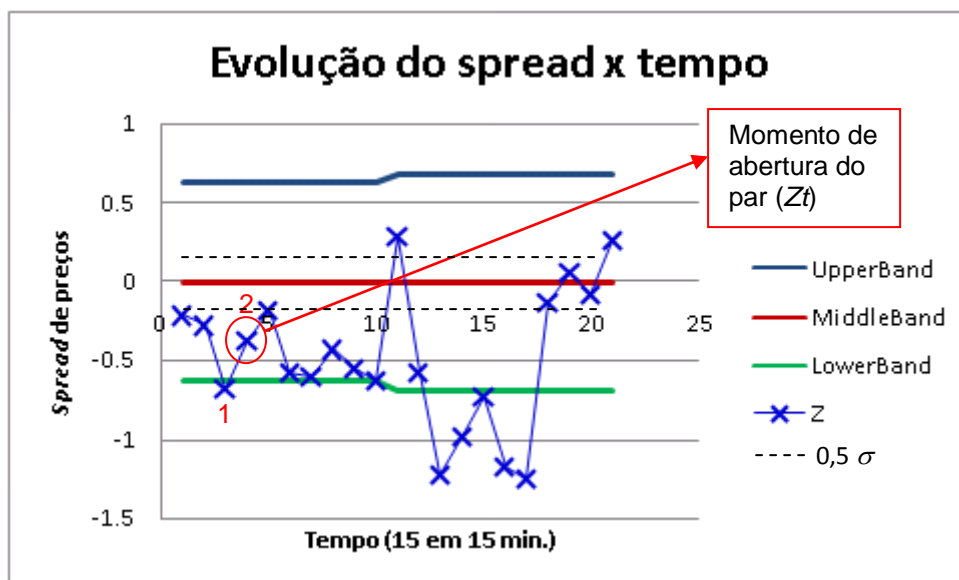


Figura 3-6 - Evolução do *spread*

A figura 3-6 exemplifica a evolução do *spread* de um determinado par, definido no gráfico como Z. O momento de abertura do par sinalizado na figura dá-se quando a diferença de preços rompeu a linha verde, no momento 1 (*LowerBand*) e voltou a aproximar-se da linha vermelha, no momento 2 (média do *spread*, *MiddleBand*). Identifica-se nesse instante que existe uma tendência de reversão à média. Ou seja, o movimento de distanciamento unicamente (momento 1) não é suficiente para a abertura de uma posição (comprada ou vendida). É necessário que o *spread* rompa a banda novamente, aproximando-se da média. A distância entre as bandas, por sua vez (na figura: *Upperband*, *Middleband* e *LowerBand*) é uma razão do desvio padrão (σ). As linhas pontilhadas, por sua vez, representam 0,5 σ , momento de fechamento das posições quando o *spread* retorna à média. Ou seja, esta é a situação de apuração de lucro.

Os valores a seguir exemplificam o modelo:

Par: ALLL3-BRKM5 (ALL America Latina Logística S.A. e Braskem S.A.)

Razão entre preços calculada nos 50 dias anteriores (*HedgeRatio*): 0,497096

Financeiro da ação 1 (por definição): R\$ 100.000,00 (neste exemplo, comprado)

No instante de abertura do par:

- Preço da ALLL3 (P_{a1t}): 6,299050 (preços ajustados com eventos corporativos)

- Preço da BRKM5 (P_{a2t}): 13,220983

Número de ações compradas (Q_1) = R\$ 100.000,00 / P_{a1t} = 15.875 (arredondando)

Número de ações vendidas (Q_2) = $Q_1 * HedgeRatio$ = 7.892 (arredondando)

Financeiro na ponta vendida = $Q_2 * P_{a2t}$ = 104.340,00

Spread (Z_t) = $P_{a1t} - P_{a2t} * HedgeRatio$ = 6,299050 - 13,220983 * 0,497096 = -0,273051

Financeiro para iniciar a estratégia = 15.875 * 6,299050 – 7.892 * 13,220983 = - **4,342.58**

Desvio (σ) = 0,315278 ($> Z_t$)

Desta forma, no momento de abertura da estratégia, existe sempre um aumento momentâneo de capital.

Contudo, a manutenção de posições vendidas exige que sejam alugadas ações no mercado. Além do custo de aluguel, que será abordado mais adiante, a BM&FBOVESPA, como contraparte de todos os contratos registrados de empréstimos, somente autoriza as operações com as garantias, pela ponta tomadora, previamente depositadas. Esta margem não precisa necessariamente ser disponibilizada em moeda corrente nacional, sendo também aceitos títulos públicos, privados e negociados em mercados internacionais, ações pertencentes à carteira do Índice Bovespa, entre outros. Dependendo do risco que apresenta a ação vendida, maior o ágio exigido pela Bolsa. O total exigido de garantias para uma operação de empréstimo é de 100% do valor dos ativos mais um intervalo de margem específico para cada ativo. O intervalo de margem representa a oscilação possível desse ativo em dois dias úteis consecutivos, dentro de um intervalo de confiança.

A hipótese, bastante realista, que será utilizada no presente estudo é a de que toda a margem extra (acima dos 100% da posição alugada, que por sua vez é igual à posição vendida) poderá ser coberta com as posições compradas.

Assim, a posição dos ativos deste modelo de investimentos será composta pelos pares comprados e vendidos, incluindo os títulos públicos no valor total das posições vendidas, de acordo com o cálculo apresentado no início desta seção.

O retorno do investimento será, então, calculado sobre o financeiro da posição vendida. O resultado será um retorno acima do rendimento dos títulos, que podem ser mantidos permanentemente depositados como garantia da operação.

O modelo desenvolvido na ferramenta Streambase armazena um controle instantâneo do investimento durante todo o período de *trading*, para todos os pares formados abertos. Por definição, cada posição aberta (por par) terá R\$ 100,000.00 comprados ou vendidos na ação 1 de cada par:

- Será comprado na ação 1 se o *spread* for negativo, ou seja, a ação 1 está sub-precificada com relação à ação 2.
- Será vendido se o *spread* for positivo, ou seja, a ação 2 está sub-precificada com relação à ação 1.

O gráfico a seguir apresenta a distribuição de capital exigido num intervalo de 4 semanas:

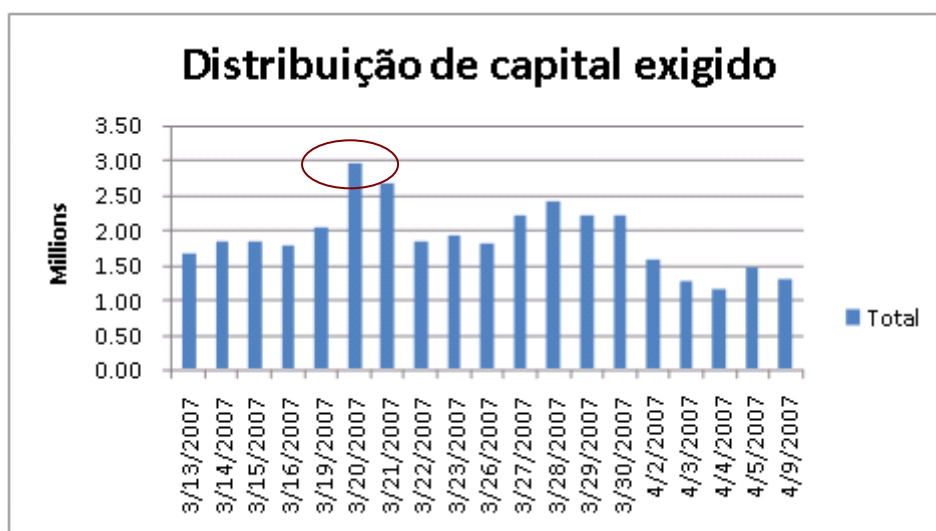


Figura 3-7 - Distribuição de capital num intervalo de 4 semanas

A figura 3.6 apresenta a distribuição de capital exigido ao longo de 1 período de 28 dias (subdivisão do intervalo completo, apresentada na figura 3-6). O pico destacado na figura, no valor de 3 milhões de reais no dia 20 de Março de 2007,

representa o máximo de capital exigido neste período. Esta foi a metodologia aplicada em todos os 10 períodos de 28 dias de cada ano, para o cálculo dos retornos.

Numa operação ativa, o capital excedente nos demais dias poderia ser redirecionado para outras estratégias, caso estivesse num ambiente de um fundo de investimentos, ou na tesouraria de algum banco. Contudo, neste estudo, será considerada a exigência máxima no período como o capital necessário para rodar a estratégia. Como exemplo, no gráfico apresentado, seriam necessários os 3 milhões de reais em caixa ou títulos públicos para este período de 28 dias corridos.

Portanto, para cada intervalo de 4 semanas, o retorno obtido será:

$$Retorno = \frac{(Resultado\ bruto - custos)}{Investimento} \quad (3.13)$$

Os retornos acumulados foram calculados como a evolução de uma cota fictícia, tal qual um Fundo de investimentos, ou seja, a cota é iniciada ao valor de R\$ 100,00, e evolui a cada 4 semanas à medida que os resultados são acumulados.

A seguir um exemplo para o ano de 2006:

		cota inicial = 100	
Ano	Período (4 semanas)	% mes	Cota
2006	1	1.79%	101.79
	2	0.66%	102.46
	3	1.08%	103.56
	4	4.44%	108.16
	5	-3.69%	104.17
	6	-2.42%	101.65
	7	4.75%	106.48
	8	5.67%	112.52
	9	-3.29%	108.82
	10	-2.79%	105.78

Figura 3-8 - Exemplo de cálculo da cota, em intervalos de 4 semanas

Na figura 3-8, a cota inicia-se ao valor de R\$ 100,00 e evolui de acordo com os retornos de cada período de 4 semanas (10 períodos no total, a cada ano). Para o período 1, a cota assume o valor de 101,79 ao final das 4 semanas, visto que o retorno foi de 1,79% neste período. Esta é, portanto, a cota de início do período 2, que evolui neste segundo intervalo de acordo com a rentabilidade de 0,66%, chegando a 102,46:

$$Cota\ ao\ final\ do\ período\ 2 = 101,79 * (1 + 0,66\%) = 102,46 \quad (3.14)$$

Esta foi a metodologia aplicada para cada período de 4 semanas e para todos os anos do intervalo analisado.

3.7. Custos operacionais

Em operações que envolvem grande movimentação de posição, ou seja, com volumes de compras e vendas elevados, a contabilização de custos de operação é crucial para o cálculo preciso dos retornos. Negligenciar os custos pode encobrir prejuízos, visto que as margens de ganhos, principalmente em operações que se beneficiam de volumes elevados, como é o caso do *pair trading*, são reduzidas.

De forma a aproximar a estratégia estudada da realidade do mercado brasileiro, foram considerados 3 tipos de custo:

- Corretagem de operação: incide sobre cada transação de compra e venda;
- Custódia: custo cobrado pelo custodiante para a manutenção dos ativos em sua conta;
- Aluguel: taxa cobrada pelo doador da ação que servirá para viabilizar a posição vendida;

Estes 3 custos foram estimados tomando como base uma operação de um fundo de investimentos de volume médio:

- 40 *basis points* para corretagem e custódia;
- 2% ao ano como valor médio para o aluguel;

Os custos totais serão apresentados mais adiante, junto dos resultados gerais da estratégia. No capítulo a seguir, são apresentados os resultados obtidos para o *trading* simulado no período de *back-testing*, de 2006 e 2010, assim como os retornos calculados para cada ano.

4. APRESENTAÇÃO DOS RESULTADOS

De posse do conjunto de preços históricos das ações do Ibovespa no período de 2006 a 2010, em intervalos intra-diários, como detalhado no capítulo 3, foram estabelecidos os limites de 70 dias corridos para seleção dos pares e cálculo dos parâmetros a serem utilizados nos 7 dias subseqüentes.

Para cada intervalo de seleção, é feita a combinação das 69 ações duas a duas para que sejam identificados todos os pares possíveis, ou seja, no máximo serão formados 2346 pares a cada período. Definidos os pares, é feito cálculo de cointegração na ferramenta MATLAB de forma a selecionar aqueles que apresentam mais de 99% de significância no teste de Dickey-Fuller aumentado, para verificação da raiz unitária dos resíduos da combinação linear entre os preços das ações que compõem cada par. Estes pares alimentam então a ferramenta Streambase que tem a regra de operação, simulando o mercado por 7 dias corridos. Junto do conjunto de pares, os parâmetros de *trading* alimentam também a ferramenta: média do *spread*, desvio padrão e razão de preços.

Os valores obtidos ano a ano são apresentados a seguir:

								cota inicial = 100
Ano	Periodo (4 semanas)	Resultado gross	Custo	Net	Net acumulado	Margem/Capital	% mes	Cota
2006	1	41,694.84	4,672.73	37,022.11	37,022.11	2,070,709.05	1.79%	101.79
	2	15,338.15	3,380.28	11,957.87	48,979.97	1,819,435.48	0.66%	102.46
	3	24,160.65	5,325.71	18,834.93	67,814.91	1,748,333.02	1.08%	103.56
	4	133,384.03	6,989.84	126,394.19	194,209.10	2,847,965.77	4.44%	108.16
	5	(72,009.75)	7,637.63	(79,647.38)	114,561.72	2,160,731.27	-3.69%	104.17
	6	(45,690.91)	7,692.20	(53,383.11)	61,178.61	2,203,050.38	-2.42%	101.65
	7	142,746.11	6,227.91	136,518.20	197,696.81	2,871,605.61	4.75%	106.48
	8	140,753.97	6,286.75	134,467.21	332,164.02	2,370,433.59	5.67%	112.52
	9	(83,449.38)	8,515.21	(91,964.59)	240,199.43	2,795,230.35	-3.29%	108.82
	10	(88,558.05)	8,472.76	(97,030.81)	143,168.62	3,476,000.25	-2.79%	105.78

								cota inicial = 100
Ano	Periodo (4 semanas)	Resultado gross	Custo	Net	Net acumulado	Margem/Capital	% mes	Cota
2007	1	31,392.13	11,195.21	20,196.92	20,196.92	2,957,753.03	0.68%	100.68
	2	(3,977.11)	7,591.56	(11,568.67)	8,628.25	2,354,174.71	-0.49%	100.19
	3	(119,858.20)	7,375.76	(127,233.96)	(118,605.70)	2,140,945.89	-5.94%	94.23
	4	(97,603.46)	6,249.65	(103,853.10)	(222,458.81)	2,484,253.00	-4.18%	90.29
	5	(76,932.55)	5,599.42	(82,531.97)	(304,990.78)	1,594,264.77	-5.18%	85.62
	6	138,357.29	12,472.62	125,884.68	(179,106.11)	2,856,060.40	4.41%	89.39
	7	105,931.34	9,546.58	96,384.76	(82,721.35)	2,799,927.57	3.44%	92.47
	8	50,807.03	8,847.25	41,959.77	(40,761.58)	3,032,494.65	1.38%	93.75
	9	22,124.86	7,843.16	14,281.70	(26,479.87)	2,857,761.14	0.50%	94.22
	10	222,887.43	10,963.12	211,924.31	185,444.44	3,119,824.43	6.79%	100.62

								cota inicial = 100
Ano	Período (4 semanas)	Resultado gross	Custo	Net	Net acumulado	Margem/Capital	% mes	Cota
2008	1	111,788.88	6,767.02	105,021.86	105,021.86	1,941,050.30	5.41%	105.41
	2	(8,957.97)	4,041.07	(12,999.03)	92,022.83	1,400,000.00	-0.93%	104.43
	3	(40,444.99)	6,367.71	(46,812.70)	45,210.12	1,629,459.93	-2.87%	101.43
	4	76,261.85	2,953.93	73,307.92	118,518.05	836,339.07	8.77%	110.32
	5	(159,435.29)	6,123.37	(165,558.67)	(47,040.62)	2,489,394.26	-6.65%	102.99
	6	84,816.18	6,084.02	78,732.16	31,691.54	2,176,340.51	3.62%	106.71
	7	(197,786.91)	13,040.24	(210,827.15)	(179,135.61)	2,731,355.89	-7.72%	98.47
	8	(628,900.90)	33,732.16	(662,633.05)	(841,768.66)	4,570,624.74	-14.50%	84.20
	9	140,719.14	18,085.59	122,633.56	(719,135.10)	3,969,533.79	3.09%	86.80
	10	85,554.45	8,709.77	76,844.69	(642,290.42)	1,978,130.98	3.88%	90.17

								cota inicial = 100
Ano	Período (4 semanas)	Resultado gross	Custo	Net	Net acumulado	Margem/Capital	% mes	Cota
2009	1	(134,973.00)	25,606.28	(160,579.29)	(160,579.29)	3,449,093.39	-4.66%	95.34
	2	(133,960.74)	24,328.60	(158,289.33)	(318,868.62)	1,869,314.10	-8.47%	87.27
	3	125,237.29	7,377.19	117,860.10	(201,008.52)	2,066,445.16	5.70%	92.25
	4	(77,326.59)	4,166.13	(81,492.72)	(282,501.24)	1,238,138.98	-6.58%	86.18
	5	(39,278.10)	8,471.38	(47,749.48)	(330,250.72)	1,670,600.62	-2.86%	83.71
	6	(24,237.30)	6,225.76	(30,463.06)	(360,713.77)	2,341,576.78	-1.30%	82.62
	7	(2,245.60)	6,190.78	(8,436.38)	(369,150.15)	1,069,585.05	-0.79%	81.97
	8	(68,991.98)	22,302.99	(91,294.97)	(460,445.12)	2,552,418.19	-3.58%	79.04
	9	80,530.88	8,000.27	72,530.61	(387,914.51)	1,539,838.42	4.71%	82.76
	10	(94,026.86)	6,320.37	(100,347.23)	(488,261.74)	1,662,621.44	-6.04%	77.77

								cota inicial = 100
Ano	Período (4 semanas)	Resultado gross	Custo	Net	Net acumulado	Margem/Capital	% mes	Cota
2010	1	7,233.62	6,025.51	1,208.11	1,208.11	2,015,641.40	0.06%	100.06
	2	(14,207.64)	7,204.88	(21,412.51)	(20,204.40)	2,872,536.25	-0.75%	99.31
	3	43,282.35	12,134.27	31,148.08	10,943.68	2,430,717.64	1.28%	100.59
	4	96,868.92	7,099.94	89,768.98	100,712.66	2,242,618.45	4.00%	104.61
	5	12,477.60	8,784.24	3,693.36	104,406.02	1,363,983.33	0.27%	104.90
	6	33,305.17	10,878.33	22,426.84	126,832.86	1,416,728.49	1.58%	106.56
	7	15,032.36	19,694.41	(4,662.05)	122,170.81	1,655,938.30	-0.28%	106.26
	8	(5,708.93)	14,006.57	(19,715.50)	102,455.31	1,331,542.85	-1.48%	104.68
	9	90,773.06	8,710.93	82,062.14	184,517.45	2,024,499.83	4.05%	108.93
	10	49,156.17	7,978.18	41,177.99	225,695.44	2,343,601.45	1.76%	110.84

Figura 4-1 - Resultados obtidos de 2006 a 2010

Os resultados finais obtidos, de acordo com a figura 4-1, consideram os custos de operação e não incluem a rentabilidade dos títulos públicos. Ou seja, como todo o capital está aplicado em títulos públicos, com a finalidade de viabilizar a operação completa, os resultados podem ser resumidos como:

- **2006:** Selic + 5,78%;
- **2007:** Selic + 0,62%;
- **2008:** Selic - 9,83%;
- **2009:** Selic - 22,23%;
- **2010:** Selic + 10,84%;

Analisando-se as grandes perdas ocorridas em 2008 e 2009, pode-se associa-las à alta volatilidade dos retornos este período. Uma nova simulação foi feita para o oitavo período de 2008, onde se observou a maior perda absoluta de todo o

intervalo (R\$ 629 mil) e -14,5%. Foram substituídos os parâmetros de entrada no *trading* de: 2σ para cima e para baixo da média do *spread* para formação das bandas (valores iniciais) para 4σ (novos parâmetros, de acordo com recomendações de especialistas do mercado). O resultado obtido inverte o retorno de -14,5% para 3,9% positivo. Este comportamento sugere que, à medida que a volatilidade dos retornos aumente, deve-se incorporar esta oscilação à estratégia de operação. Ou seja, quanto maior a volatilidade, maiores deveriam ser as bandas que determinam os momentos de entrada e saída do *trading*. Contudo, este comportamento somente pode ser confirmado com simulações de um conjunto maior de variação dos parâmetros e para todo o período. A simulação em apenas 1 mês tem o objetivo apenas de sugerir o caminho para ampliação do presente estudo.

Analisando ainda o detalhe de cada par, podemos observar que existem 3 tipos de operação: as que foram finalizadas após a reversão do *spread* à média, ou seja, geraram lucro, as que foram finalizadas pelo final do período de 4 semanas, podendo ou não gerar prejuízo, ou no *stop-loss* de 3 desvios padrão, estas sim com perdas.

Não foi possível estabelecer uma relação entre o número de operações realizadas e uma tendência a ganhos e perdas.

Ganho/Perda	Número de operações	Total de pares	Percentual por resultado
Zero	Zero	387	100.00%
Perda	1 operação	234	53.42%
	Mas de 1 operação	204	46.58%
Ganho	1 operação	265	56.99%
	Mas de 1 operação	200	43.01%
1290			

Tabela 4-2 - Distribuição dos pares selecionados em 2006 pelo número de *trades* executados

A tabela 4-2 mostra a distribuição dos pares selecionados em 2006 versus o número de *trades* executados, de acordo com o resultado (ganho, perda ou resultado zerado). Observa-se que, dentro do universo de pares que tiveram alguma transação, a maior parte pôde ser operada apenas uma vez, tanto para resultados positivos quanto negativos. Ou seja, não foi possível notar uma tendência de resultado de acordo com a frequência de execução de operações.

A tabela a seguir apresenta a distribuição geral das operações, pelo número de *trades* executados (abertura e fechamento a operação), independente do resultado:

Número de operações	Total de pares	Percentual por # <i>trades</i>
Zero	387	30%
1 operação	499	39%
Mas de 1 operação	404	31%

Tabela 4-3 - Distribuição geral, independente do retorno

Apesar de o modelo permitir alta frequência de operações, por conta do intervalo selecionado (15 em 15 minutos), a tabela 4-3 mostra que a maior parte dos pares operou com apenas 1 compra e venda (39% contra 31% para pares com frequência maior). Este comportamento pode ser explicado exatamente pela teoria de cointegração. Os estudos apresentados na literatura (Vidyamurthy, 2004 e Ernest P. Chan, 2009) indicam que cointegração é uma característica observada ao longo de períodos maiores, ou seja, a reversão dos resíduos à média é uma característica que não deveria ser observada em apenas 1 dia de operação, sendo necessários movimentos com prazos maiores.

Outro dado que fortalece esta interpretação é o tempo médio em que as posições permaneceram abertas. A tabela 4-4 a seguir apresenta: o tempo médio de todas as operações e também segregadas, dependendo do resultado obtido (ganho ou perda):

Tempo médio de posições abertas =	3.56	dias
Desvio geral =	3.51	dias

Tempo médio de posições com ganho =	3.39	dias
Tempo médio de posições com perda =	3.73	dias

Figura 4-4 - Duração média das posições abertas

Numa operação de alta frequência, era esperado que a maior parte das operações fossem intra-diárias. O que se observa, contudo, é que menos de 15% delas são encerradas no mesmo dia. O gráfico a seguir apresenta esta distribuição:

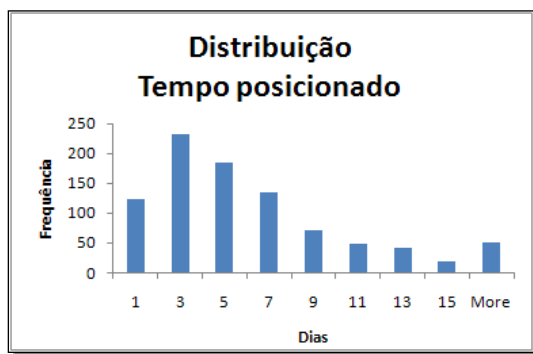


Gráfico 4-5 – Distribuição do tempo em que os pares se mantiveram abertos (máximo 20 dias)

A primeira coluna do gráfico 4-5, cujo tempo de duração da operação é menor ou igual a 1 dia, tem o percentual de cerca de 15%. Este comportamento reforça a teoria que a cointegração é observada para períodos mais longos que as variações intra-diárias, para a maioria das operações testadas.

Os resultados individuais dos pares com maiores e menores retornos podem ser visto no **Anexo 2**.

Complementando a análise, o impacto dos custos operacionais é claramente observado, quando comparado ao resultado bruto obtido a cada ano:

- **2006:** 31,3% de custos operacionais totais sobre resultado bruto;
- **2007:** 32,1% de custos;
- **2008:** 19,7% de custos;
- **2009:** 32,2% de custos;
- **2010:** 31,2% de custos;

Nota-se, portanto, que o resultado líquido simulado é fortemente impactado pelos custos considerados. As estimativas utilizadas neste estudo consideram a operação real de um fundo de investimentos de tamanho médio, com volumes entre 0,5 e 1,0 bilhão de reais ao mês. Assim, alterações nos parâmetros de custos podem inviabilizar a estratégia.

Excluindo-se o período de crise (alta volatilidade), este estudo obteve retornos positivos para os períodos de 2006, 2007 e 2010 (Selic + 5,78%, Selic + 0,62% e Selic + 10,84%, respectivamente). Comparando-os aos resultados obtidos por Kabashima, 2005, observa-se que ele apresenta retornos de Selic + 3,65%.

Já quando são incluídos os anos de 2008 e 2009, os resultados são bastante díspares. Deve-se considerar que o período analisado por Kabashima, considera

cenários econômicos distintos dos presentes neste trabalho, principalmente com relação à crise de 2008. O trabalho de Kabashima trata do período de 1993 a 2005.

Perlin, 2007, também apresentou retornos positivos em seu trabalho ao simular estratégias de pares no mercado brasileiro. Contudo a forma de seleção dos pares foi diferente, sendo feita através do cálculo da distância mínima quadrada para os preços normalizados. Os retornos dos pares selecionados por este método foram apresentados de forma relativa a pares selecionados aleatoriamente, diferentemente da referência de CDI adotada no presente estudo. Por conta desta diferença, os valores não puderam ser comparados diretamente, contudo observa-se que ambos os trabalhos apresentam retornos positivos.

Alexander e Dimitriu, 2002 estudaram o comportamento dos preços de carteiras de ações no mercado americano entre 1995 e 2001. Eles testaram a estratégia *long-short* baseada nas ações que compõem o DJIA, com seleção das ações também utilizando cointegração. Os resultados encontrados por estes pesquisadores foram bastante similares aos apresentados no presente estudo. Alexander e Dimitriu encontraram retornos que variaram entre os anos estudados, sendo que o período de melhor resultado anual gerou 10% de retorno, livre de custos de transação. Houve também anos com retornos negativos (1996 e 2000), como os aqui apresentados.

Caldeira, 2010 reproduziu estratégia similar a Kabashima, 2005 e encontrou resultados positivos no intervalo de 2006 a 2009 no mercado brasileiro (17% de retorno anualizado). Como Caldeira não detalha os parâmetros de entrada e saída das operações compradas e vendidas, a comparação com o estudo atual pode ser comprometida, mas os retornos positivos mesmo nos anos de crise (principalmente 2008) sugerem que a calibragem destes parâmetros (bandas de operação) pode ajustar os resultados negativos aqui obtidos.

Consolidando os resultados obtidos nos diversos estudos que relacionam a operação de pares no mercado brasileiro utilizando cointegração para seleção das ações, os resultados são encorajadores. O presente estudo acrescenta a análise para dados intra-diários e encontra resultados similares àqueles obtidos com preços de fechamento. Por conta do tempo médio em que as operações permanecem abertas, nota-se que apesar dos dados terem frequência bem mais alta que os demais estudos, a maioria das operações não é fechada no mesmo dia, sugerindo que a característica

de cointegração apresenta evidências para períodos superiores à frequência intra-diária.

5. CONCLUSÃO E SUGESTÃO DE PESQUISAS FUTURAS

Este estudo aborda a operação com pares cointegrados no mercado de ações, selecionados a partir do teste de Dickey-Fuller aumentado, para verificação da raiz unitária dos resíduos da combinação linear entre os preços dos ativos que compõem cada par.

A operação foi simulada com dados de mercado de *back-testing* através da modelagem de um algoritmo que dispara ordens de compra e venda quando é detectada distância maior que $\pm 2\sigma$ do *spread* de sua média, definido como a diferença entre os preços das ações que compõem o par, normalizado pela razão entre os preços. O universo de dados compreende preços em intervalos de 15 em 15 minutos, no período de 2006 a 2010, tomando-se as ações listadas no Ibovespa em Março de 2011. A estratégia terá retornos positivos quando a diferença dos preços das ações que compõem o par reverte à média, calculada no período anterior.

Os resultados obtidos mostram que a estratégia é fortemente impactada pela volatilidade dos retornos dos preços das ações. Os parâmetros calibrados para o primeiro mês de 2006 puderam ser aplicados com sucesso para o restante do intervalo (retorno de Selic + 5,8% no ano de 2006), para 2007, onde o resultado foi bastante próximo da Selic e para 2010, com resultado de Selic + 10,8%. Nos anos de maior volatilidade, os testes com os mesmos parâmetros de 2006 apresentaram prejuízos. Em 2008, ano de crise mundial desencadeada pelos problemas no mercado hipotecário nos Estados Unidos, sugere-se que a forte oscilação dos preços dos ativos no mercado financeiro deve ser incorporada aos parâmetros da estratégia. Foram repetidos então os testes em 2 dos meses com resultados negativos, porém com bandas de oscilação mais estendidas para acompanhamento do *spread* entre os preços das ações de cada par. Os novos parâmetros foram de $\pm 4\sigma$ para entrada na estratégia e $\pm 6\sigma$ para saída (stop loss). Os resultados desta nova simulação sugerem que testes futuros podem incorporar a volatilidade dos retornos, do período de seleção, na definição da banda ótima de operação.

Apesar da proposta de risco reduzido por conta das posições compradas e vendidas, *pair trading* não pode ser considerada uma estratégia livre de risco, como os resultados confirmam. Sendo assim, este estudo pode ser complementado com uma

análise de risco para a evolução dos retornos mês a mês. Podem ser utilizados índices como o de Sharpe, ou mesmo a volatilidade dos retornos.

Por questões operacionais, os períodos selecionados descartam a operação nas 10 primeiras semanas de cada ano (período exclusivo de formação dos pares). Esta simplificação reduz o período total analisado. Assim, testes futuros poderiam simular a operação como um período único de 2006 a 2010, ou maiores, sem a segmentação anual.

Estas bandas, calculadas em função do desvio σ , desenham a estratégia de *trading*, ou seja, estabelecem os parâmetros para disparar compras e vendas em função do deslocamento do *spread*. Pesquisas futuras poderiam testar variações dos valores de 2σ para entrada na posição e $0,5\sigma$ e 3σ para saída. Como apresentado no capítulo de resultados, para períodos de alta volatilidade, recomenda-se que as bandas sejam aumentadas.

Outra limitação deste estudo está no custo de *bid* e *ask*, que não foram aqui incorporados, ou seja, foram utilizados unicamente os preços de fechamento de cada período de 15 em 15 minutos. Estender a análise incluindo este custo pode ser uma proposta para trabalhos futuros, visto que principalmente em períodos de forte volatilidade, estes custos podem ser significativos.

Como os preços utilizados são em intervalos razoavelmente grandes, mais uma possibilidade de aprofundamento de pesquisas seria a redução do intervalo para 5 ou 1 minuto, por exemplo, de forma a aproximar a análise de uma série contínua.

A literatura apresenta ainda outros testes para seleção dos pares, tais como teste de raiz unitária de DF-GLS e KPSS (*paper* de Longo, 2008). Estas novas seleções de pares podem ser comparadas aos resultados obtidos no presente estudo, que utiliza o teste de Dickey-Fuller Aumentado.

Finalmente, o presente estudo não considera o re-investimento dos ganhos na mesma estratégia, sendo mais uma oportunidade de aprofundamento das pesquisas.

6. BIBLIOGRAFIA

ALEXANDER, Carol. *Market Models. A Guide To Financial Data Analysis*. John Wiley & Sons, Inc., 2001.

ALEXANDER, Carol e DIMITRIU, Anca. *The cointegration alpha: Enhanced Index Tracking and Long-Short Equity Market Neutral Strategies*. Working Paper. ISMA Centre, Henley University of Reading, 2002.

CALDEIRA, João. *Arbitragem estatística e estratégia long-short, pair trading, abordagem da cointegração aplicada a dados do mercado brasileiro*. Paper. ANPEC, 2010.

CHAN, Ernest P. *Quantitative Trading. Wiley Trading*. John Wiley & Sons, Inc., 2009.

CHUI, J., LUKMAN, D., MODARRESI, K. e VELAYUTHAM, A. *High-frequency Trading*. Paper. Stanford University, 2011.

DEWAN, Sanjeev e MENDELSON, Haim. *Information Technology and Time-Based Competition in Financial Markets*. Paper. Management Science, Vol. 44, No. 5, pp. 595-609, 1998.

DICKEY, David e FULLER, Wayne A. *Distribution of the Estimators for Autoregressive Time Series With a Unit Root*. Journal of the American Statistical Association, Vol. 74, No. 366, pp. 427-431, 1979.

DICKEY, David e FULLER, Wayne A. *Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root*. Paper. Econometrica, Vol. 49, No. 4, pp. 1057-1072, 1981.

EHRMAN, Douglas. *Pairs Trading: New Look at an Old Strategy*. Paper. Futures Magazine, pp. 32 – 34, 2004.

ENDERS, Walter e GRANGER, C. W. J. *Unit-Root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates*. Paper. Journal of Business & Economic Statistics, Vol. 16, No. 3, pp. 304-311, 1998.

ENDERS, Walter. *Applied Econometric Time Series*. New York: John Wiley & Sons, Inc., 2004.

ENGLE, R. and GRANGER, C. *Co-integration and error correction: Representation, estimation and testing*. Papes. Econometrica, Vol. 55, No. 2, pp. 251-276, 1987.

HERLEMONT, Daniel (2000). *Pairs trading, convergence trading, cointegration*. Working paper. YATS Finances & Technologies, 2000.

KABASHIMA, Ernesto M. *Pairs Trading: Aplicação no mercado acionário brasileiro*. Dissertação (MPFE). Fundação Getúlio Vargas, Escola de Economia de São Paulo, 2005.

LESAGE, James P. *Spatial Econometrics*. Paper. Toledo: University of Toledo, Department of Economics, 2008.

LONGO, Eduardo M. L. *Pairs Trading: Uma aplicação ao mercado acionário brasileiro*. Dissertação (MPFE). Fundação Getúlio Vargas, Escola de Economia de São Paulo, 2008.

PERLIN, Marcelo S. *Evaluation of Pairs Trading Strategy at the Brazilian Financial Market*. Journal of Derivatives & Hedge Funds, Vol. 15, pp. 122-136, 2006.

SAID, Said E. e DICKEY, David A. *Testing for unit roots in autoregressive-moving average models of unknown order*. Paper. Biometrika, Vol. 71, No. 3, pp. 599-607, 1984.

SHARPE, William, e ALEXANDER, Gordon, *Investments*, 4th edition, Prentice Hall, Engle-wood Cliffs, N.J., 1990.

SHLEIFER, Andrei e VISHNY, Robert. *The Limits of Arbitrage*. Paper. The Journal of Finance, Vol. 52, No. 1, pp. 35-55, 1997.

STOCK, James H. e WATSON, Mark W. *Testing for Common Trends*. Journal of the American Statistical Association, Vol. 83, No. 404, pp. 1097-1107, 1988.

VIDYAMURTHY, Ganapathy. *Pairs Trading, Quantitative Methods and Analysis*. John Wiley & Sons, Inc., 2004.

WOOLDRIDGE, Jeffrey M. *Introdução à Econometria. Uma abordagem Moderna*. Thomson Learning, Inc., 2006.

ANEXO 1

Lista das empresas que compuseram o Ibovespa em março de 2011, referência usada para selecionar o universo de ações mais líquidas do mercado brasileiro:

	Empresa	Símbolo
1	PETROBRAS - PETROLEO BRAS-PR	PETR4
2	CIA VALE DO RIO DOCE-PREF A	VALE5
3	OGX PETROLEO E GAS PARTICIPA	OGXP3
4	ITAU UNIBANCO BANCO MULTIPLO SA	ITUB4
5	BM&F BOVESPA SA	BVMF3
6	PETROBRAS - PETROLEO BRAS	PETR3
7	USINAS SIDER MINAS GER-PF A	USIM5
8	GERDAU SA-PREF	GGBR4
9	BANCO BRADESCO -PREF	BBDC4
10	CIA VALE DO RIO DOCE	VALE3
11	BANCO DO BRASIL SA	BBAS3
12	PDG REALTY SA	PDGR3
13	CIA SIDERURGICA NACIONAL SA	CSNA3
14	ITAUSA-INVESTIMENTOS ITAU-PR	ITSA4
15	CIELO SA	CIEL3
16	REDECARD SA	RDCD3
17	CYRELA BRAZIL REALTY	CYRE3
18	GAFISA SA	GFSA3
19	BRF - Brasil Foods SA	BRFS3
20	CIA ENERGETICA MINAS GER-PRF	CMIG4
21	MRV ENGENHARIA	MRVE3
22	FIBRIA CELULOSE SA	FIBR3
23	MMX MINERACAO E METALICOS SA	MMXM3
24	JBS SA	JBSS3
25	ALL AMERICA LATINA LOGISTICA	ALLL3
26	LOJAS RENNER S.A.	LREN3
27	LOJAS AMERICANAS SA-PREF	LAME4
28	ROSSI RESIDENCIAL SA	RSID3
29	TIM PARTICIPACOES SA	TCSL4
30	TELE NORTE LESTE PART-PREF	TNLP4
31	CIA BRASILEIRA DE DISTR-PREF	PCAR5
32	BRADESPAR SA -PREF	BRAP4
33	BANCO SANTANDER (BRASIL) SA	SANB11
34	VIVO PARTICIPACOES SA-PREF	VIVO4
35	CCR SA	CCRO3
36	CENTRAIS ELETRICAS BRASILIER	ELET3
37	HYPERMARCAS S.A	HYPE3
38	NATURA COSMETICOS SA	NATU3
39	CIA DE BEBIDAS DAS AME-PREF	AMBV4
40	TAM SA-PREF	TAMM4
41	BRASIL ECODIESEL INDUSTRIA E	ECOD3
42	GERDAU METALURGICA SA-PREF	GOAU4
43	COSAN SA INDUSTRIA COMERCIO	CSAN3
44	CENTRAIS ELETRICAS BRAS-PR B	ELET6
45	GOL LINHAS AEREAS - PREF	GOLL4
46	MARFRIG FRIGORIFICOS E COMER	MRFG3
47	BROOKFIELD INCORPORACOES SA	BISA3
48	ELETROPAULO METROPOLI-PREF	ELPL4

	Empresa	Símbolo
49	EMBRAER SA	EMBR3
50	CIA PARANAENSE DE ENERGI-PFB	CPLE6
51	USINAS SIDERURGICAS DE MINAS	USIM3
52	CIA ENERGETICA DE SP-PREF B	CESP6
53	BRASKEM SA-PREF A	BRKM5
54	LIGHT SA	LIGT3
55	DURATEX SA	DTEX3
56	B2W Companhia Global Do Varejo	BTOW3
57	KLabin SA-PREF	KLBN4
58	LLX LOGISTICA SA-ORD	LLXL3
59	CPFL ENERGIA SA	CPFE3
60	ULTRAPAR PARTICIPACOES-PREF	UGPA4
61	PORTX ON	PRTX3
62	SOUZA CRUZ SA	CRUZ3
63	BRASIL TELECOM SA - PREF	BRTQ4
64	CIA SANEAMENTO BASICO DE SP	SBSP3
65	TELE NORTE LESTE PART	TNLP3
66	CIA DE TRANSMISSAO DE ENE-PF	TRPL4
67	TIM PARTICIPACOES SA	TCSL3
68	TELEMAR NORTE LESTE SA-PR A	TMAR5
69	TELECOMUNICACOES DE SAO-PREF	TLPP4

ANEXO 2

Pares de maiores e menores retornos operados em 2006, e seu resultado final, após custos de corretagem, aluguel e custódia:

Par	Resultado no período (R\$)	Número de Operações
USIM3 - LIGT3	49,794.76	15
USIM3 - BRTO4	49,134.24	10
USIM3 - CRUZ3	46,403.52	9
USIM3 - TRPL4	42,788.83	8
KLBN4 - TCSSL3	39,415.26	17
USIM3 - BRKM5	33,778.37	11
LAME4 - SBSP3	31,626.33	13
USIM3 - CPFE3	31,034.44	10
USIM3 - TNLP3	26,689.64	5
ELPL4 - LIGT3	25,360.17	4
USIM5 - USIM3	(20,747.67)	13
CMIG4 - USIM3	(20,848.52)	10
CRUZ3 - TMAR5	(20,850.14)	5
CRUZ3 - TCSSL3	(23,107.27)	3
TCSSL4 - ELPL4	(23,316.56)	5
ELPL4 - TLPP4	(23,852.30)	8
TNLP4 - TMAR5	(27,928.49)	36
TCSSL3 - TLPP4	(30,471.06)	7
AMBV4 - SBSP3	(39,766.22)	17
ELPL4 - TCSSL3	(39,800.78)	6

A tabela do ANEXO 2 apresenta os 10 pares mais rentáveis no ano de 2006, assim como os 10 de maiores prejuízos, de um total de 557 pares no ano. A segunda coluna apresenta os resultados absolutos em reais, já incluídos os custos. A última coluna apresenta o número de operações efetuadas para este par ao longo de todo o período.

ANEXO 3

Histograma da distribuição de resultados líquidos para cada par exposto no ANEXO 2:

