



Instituto de  
MATEMÁTICA  
E ESTATÍSTICA

UFRGS



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DEPARTAMENTO DE ESTATÍSTICA

TRABALHO DE CONCLUSÃO DE CURSO

**PAIRS TRADING EM AÇÕES DO RAMO TECNOLÓGICO NORTE-AMERICANO  
BASEADO EM COINTEGRAÇÃO**

**PEDRO CHASSOT CANDIDO ANGELI**

Porto Alegre  
2018

**PEDRO CHASSOT CANDIDO ANGELI**

**PAIRS TRADING EM AÇÕES DO RAMO TECNOLÓGICO NORTE-AMERICANO  
BASEADO EM COINTEGRAÇÃO**

Trabalho de Conclusão apresentado à Comissão de Graduação do Departamento de Estatística da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

Orientador:

Prof. Dr. Eduardo de Oliveira Horta  
Doutor pela Universidade Federal do Rio Grande do Sul, Porto Alegre, RS

Coorientador:

Prof. Dr. João Frois Caldeira  
Doutor pela Universidade Federal do Rio Grande do Sul, Porto Alegre, RS

Instituto de Matemática e Estatística  
Departamento de Estatística

**Pairs Trading em Ações do Ramo Tecnológico Norte-Americano Baseado em  
Cointegração**

Pedro Chassot Candido Angeli

Banca examinadora:

Prof. Dr. Hudson da Silva Torrent  
Doutor em Economia pela Universidade Federal do Rio Grande do Sul

## **AGRADECIMENTOS**

Quero agradecer primeiramente a Deus por ter me dado a benção da vida, aos meus amados pais que sempre deixaram ao meu alcance as melhores condições de estudo, moradia e deslocamento, e por terem me dado irmãs que estiveram presentes quando precisei. Agradeço a minha namorada Luana que esteve sempre ao meu lado durante os cinco anos de curso, me apoiando e sendo compreensiva quanto a meus estudos, e aos amigos que adquiri ao longo da vida e continuam presentes. Agradeço ao professor João pelo conhecimento técnico depositado a este trabalho. Não poderia deixar de agradecer ao professor Eduardo que, não só me guiou durante esse trabalho, mas me apoiou durante toda a graduação sempre me incentivando e se disponibilizando quando preciso.

## SUMÁRIO

<b>RESUMO.....</b>	<b>6</b>
<b>ABSTRACT.....</b>	<b>7</b>
<b>INTRODUÇÃO.....</b>	<b>8</b>
<b>PAIRS TRADING.....</b>	<b>10</b>
SELEÇÃO DOS PARES E COINTEGRAÇÃO .....	12
MODELAGEM.....	13
CUSTOS DE TRANSAÇÃO.....	16
CÁLCULO DOS RETORNOS .....	16
DADOS .....	18
SOFTWARE .....	18
ESTIMAÇÃO E NEGOCIAÇÃO .....	19
<b>RESULTADOS DA ESTRATÉGIA .....</b>	<b>20</b>
<b>CONCLUSÃO .....</b>	<b>33</b>
<b>REFERÊNCIAS.....</b>	<b>35</b>
<b>APÊNDICE.....</b>	<b>36</b>

## RESUMO

Neste trabalho, implementou-se a estratégia de pairs trading selecionando vinte ativos de ações do setor tecnológico de bolsas americanas, onde dados de cotações de fechamento diárias foram coletados de janeiro de 2008 a janeiro de 2013. Na seleção dos pares, utilizou-se a metodologia de Cointegração de Engle-Granger com uso dos testes Augmented Dickey Fuller (ADF), Phillips–Perron (PP) e de Johansen. Dentre os oito períodos de negociação no trabalho, o maior retorno anualizado de um par foi de 81,5% e o maior retorno anualizado médio em uma carteira de pares foi de 26.6%, com um índice de Sharpe de 2,56. Por outro lado, o menor retorno anualizado de um par foi de -52,3% e o menor retorno anualizado médio em uma carteira de pares foi de -16,7%.

**Palavras-chave:** Pairs Trading, Neutralidade de Mercado, Arbitragem Estatística, Reversão à Média, Séries Temporais, Bolsa de Valores.

## ABSTRACT

In this work, pairs trading strategy has been implemented selecting twenty stocks from technological sector quoted at American Stock Exchanges, where data with historical closing prices were collected from January 2008 to January 2013. In pairs selection, Engle-Granger Cointegration methodology was performed using Augmented Dickey Fuller (ADF), Phillips–Perron (PP) and Johansen tests. Among the eight trading periods, the largest annualized return on a pair was 81.5% and the highest average annualized return on a portfolio of pairs was 26.6%, with a Sharpe ratio of 2.56. On the other hand, the lowest annualized return on a pair was -52.3% and the lowest average annualized return on a portfolio of pairs was -16.7%.

**Keywords:** Pairs Trading, Market Neutral, Statistical Arbitrage, Mean Reversion, Time Series, Stock Exchange.

## Introdução

Compra e venda de ações de ativos da bolsa de valores é uma alternativa de investimento bastante propagada e atrativa para investidores audazes. Deter uma ação de uma empresa com capital aberto significa vir a ter uma fração da propriedade desta empresa. Um acionista tem direito a participação nos lucros da empresa e pode vir a ter um aumento do seu patrimônio investido, porém em um cenário negativo, pode-se perder o valor do investimento. Há a possibilidade de se ter uma rentabilidade altíssima em um determinado dia, assim como perder grande parte do valor investido em outro. Devido à volatilidade do mercado, investir em ações é considerado um investimento de alto risco e uma alternativa não indicada a investidores mais conservadores.

Estratégias de alocação de ativos baseadas em reversão à média, mais conhecida na literatura por *mean reversion*, têm sido usadas por bancos de investimento e fundos de cobertura há diversos anos (Caldeira et al., 2013). Dentre essas, que estão sob a alçada da Arbitragem Estatística, encontra-se a estratégia de Pairs Trading, introduzida pelo grupo quantitativo de Nunzio Tartaglia na Morgan Stanley nos anos de 1980. O grupo era formado por matemáticos, físicos e cientistas da computação que desenvolviam sistemas de negociação automatizados para identificar similaridades entre diferentes ativos do mercado norte-americano (Vidyamurthy, 2004). A estratégia fundamenta-se em operações do tipo *long-short*, a qual consiste basicamente em encontrar duas ações cujos preços apresentem um comportamento similar historicamente; quando a discrepância entre os preços das ações se distancia em relação à média, abre-se uma posição no mercado vendendo a ação de maior valor (*short*) e comprando a de menor valor (*long*). A posição é fechada pela reversão à média do *spread* (amplitude) entre os ativos.

Pairs trading é uma arbitragem estatística projetada para explorar desvios de curto prazo de um equilíbrio a longo prazo entre dois ativos. Nessa estratégia, transações *long-short* são construídas para gerar uma posição líquida de zero. Tais estratégias equivalem a um portfólio beta-zero, o que elimina a exposição ao risco



direcional do mercado, de tal forma que o retorno obtido não deve ser correlacionado com o índice de referência de mercado (Caldeira et al., 2013).

No presente trabalho, aplicou-se pairs trading em ações do setor de tecnologia Norte Americano. Como regra para a seleção dos pares, utilizou-se a técnica de Cointegração pela metodologia de Engle-Granger. Através da Cointegração, podemos encontrar uma combinação linear estacionária entre duas ou mais séries temporais de preços. Se ambos os ativos compartilham uma relação de equilíbrio a longo prazo, espera-se que se possa lucrar com desvios a curto prazo entre eles.

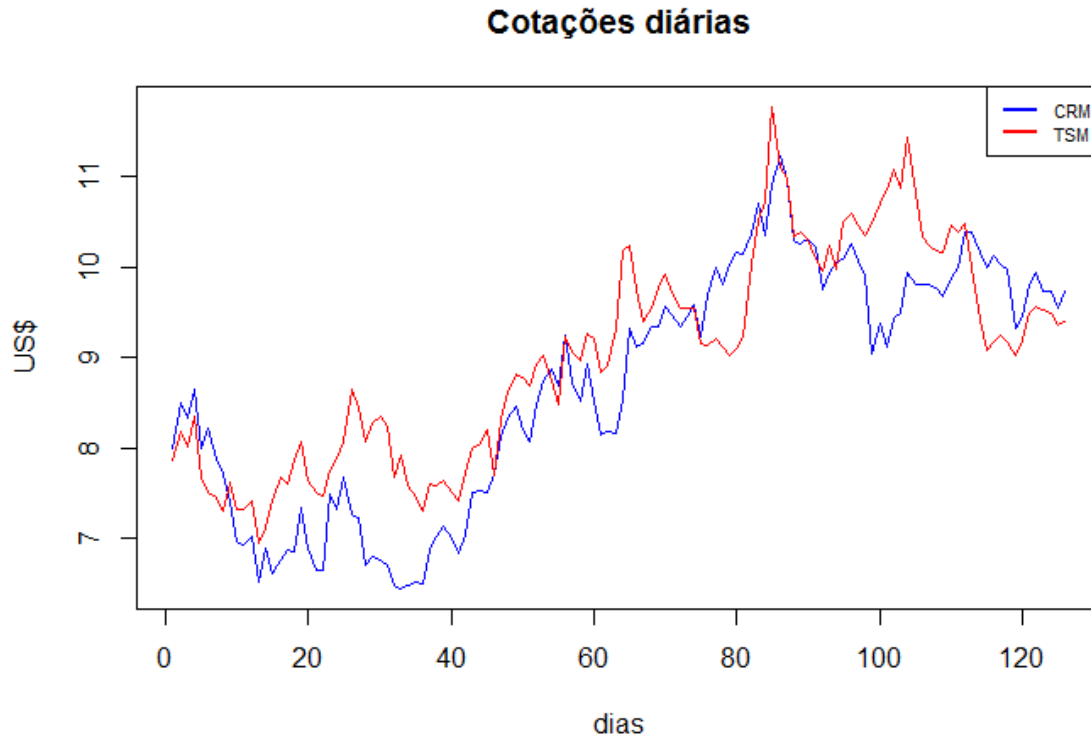
Em trabalhos anteriores, pesquisadores atingiram, através de estratégias de pairs trading, retornos financeiros médios anualizados de 11% (Gatev et al., 2006), 12.5% (Broussard et al., 2012), 16,38% (Caldeira et al., 2013) e 15,05% (Brito, 2014).

## Pairs Trading

Segundo Caldeira et al. (2013), arbitragem estatística é uma abordagem puramente estatística baseada no pressuposto de que padrões observados no passado serão repetidos futuramente. Ela é uma estratégia de negociação altamente quantitativa e computacional, e engloba uma variedade de estratégias de investimento cujo principal objetivo é gerar retornos em excesso a partir de ferramentas estatísticas para explorar ineficiências de mercados financeiros que não estão em equilíbrio. Diferentemente de negociações tradicionais em mercados de ações, onde se lida com tendências estabelecidas para ativos específicos tentando prever o comportamento dos preços dos ativos, em pairs trading lida-se com tendências de equilíbrio a longo prazo em comum a um par de ações.

Broussard & Vaihekoski (2012) argumentam que uma negociação em par tem o intuito de lucrar a partir de forças de convergência que eliminam os desvios de preço temporários em favor de relações históricas de preços de longo prazo. De acordo com Gatev et al. (2006), a ideia de pairs trading consiste em identificar um par de ações com movimentos históricos similares entre si. Quando há um desvio significativo na relação de estabilidade dos preços do par, uma posição de mercado é aberta vendendo a ação sobrevalorizada (*short position*) e comprando a subvalorizada (*long position*). Essa posição *long-short* é estabelecida na expectativa de que essa discrepância entre os pares seja de curto prazo e que, a longo prazo, isso vá reverter à média. Portanto, a negociação é considerada autofinanciada. Na metodologia desta estratégia, admite-se apenas abrir uma posição de mercado ou liquidar totalmente uma posição aberta previamente.

Um exemplo de par de ativos cujos preços apresentam uma relação comum de longo prazo pode ser observado na Figura 1, onde são mostradas as séries de cotações diárias dos ativos da Salesforce.com (CRM) e da Taiwan Semiconductor Manufacturing Company Limited (TSM).



**Figura 1:** Evolução dos preços dos ativos da CRM e da TSM, de Jan/2009 a Jun/2010.

A ideia de pairs trading é investir o mesmo montante nas operações *long* e *short*, de modo que  $P_t^s = \alpha P_t^l$ , onde  $P_t^s$  é o preço na operação *short* no tempo  $t$  e  $P_t^l$  é o preço na operação *long* no tempo  $t$ , tornando esse um investimento com custo inicial igual a zero (Caldeira et al., 2013). Isso é possível tomando emprestado ações de um ativo e vendendo-as imediatamente, usando o valor adquirido para investir essa quantia em  $\alpha$  ações de outro ativo. Alguns autores indicam fazer a operação em função do parâmetro  $\beta$  da combinação linear dos ativos. Inamdar et al. (2016) sugerem abrir uma posição *long* de uma unidade do ativo  $y$  versus  $\beta$  unidades do ativo  $x$ .

Para uma revisão de literatura envolvendo pairs trading, podem ser citados alguns trabalhos. Gatev et al. (2006) documentam que pairs trading gera lucros consistentes no mercado de ações dos Estados Unidos, o que é considerado o mercado mais eficiente e líquido no mundo. Broussard & Vaihekoski (2012) dão evidências de rentabilidade por pairs trading em mercado de ações com liquidez reduzida na Finlândia. Por fim, Caldeira et al. (2013) performam pairs trading em ações

da bolsa de valores do estado de São Paulo (Bovespa), mostrando ser lucrativo mesmo em períodos de crise global.

### **Seleção dos Pares e Cointegração**

Na literatura, encontram-se diversos métodos utilizados na seleção dos pares de ativos, alguns sendo baseados em correlação ou em outras regras de decisão não-paramétricas. Gatev et al (2006) usaram o método da distância, o qual escolhe pares com os menores desvios quadráticos entre os preços normalizados. Rad et al. (2016) avaliam os resultados de pairs trading através de cointegração e do método de cópula contra o método da distância. Na estratégia por cópula, para cada par selecionado, a função cópula e a função de distribuição marginal são estimadas e os pares são selecionados com base nos critérios AIC e BIC.

A técnica de Cointegração vem sendo aplicada na alocação de ativos desde 1997. Caldeira et al. (2013) justificam que um sistema envolvendo séries de preços de ações não-estacionários em nível pode apresentar uma tendência estocástica comum. Vidyamurthy (2004) alega que qualquer desvio do equilíbrio entre séries temporais cointegradas é causado por uma tendência estocástica comum e que esse desvio será corrigido à medida que as séries evoluem no tempo. Quando comparamos Cointegração a Correlação, a primeira leva vantagem por permitir o uso da informação contida nos níveis de variáveis financeiras. Jakobsson (2015) alega que preços cointegrados se dirigem de volta um para o outro quando o spread é grande. No caso em que as séries de preços se desviam uma da outra e estas só são correlacionadas, nada diz que elas voltarão a se encontrar, já que os preços futuros não dependem do histórico. Assim, séries de preços cointegrados são muito mais adequados na estratégia de pairs trading do que séries de preços com retornos correlacionados.

Uma série temporal não estacionária é dita  $I(1)$  se a série da sua primeira diferença forma um processo estacionário. Duas séries temporais são ditas cointegradas se ambas compartilham ordem de integração  $I(1)$  e podem ser

combinadas linearmente formando uma série temporal  $I(0)$ , ou seja, uma série estacionária. A partir da validação da suposição de estacionariedade dos resíduos obtidos de uma combinação linear estimada por Mínimos Quadrados Ordinários (MQO), espera-se que o *spread* atenda à reversão à média.

Engle e Granger (1987) desenvolveram um teste de duas etapas para detectar Cointegração. Na primeira etapa, estima-se uma relação linear no par de preços através da equação

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (1)$$

onde  $x_t$  é o preço do ativo x no tempo t,  $y_t$  é o preço do ativo y no tempo t,  $\beta_0$  é o intercepto e  $\beta_1$  é o coeficiente angular da regressão. Na segunda etapa, testes de estacionariedade são executados para os resíduos da regressão.

## Modelagem

Baseado em valores históricos do par de ativos, um modelo com respeito aos resíduos entre as duas séries do par é estimado e, com os parâmetros estimados nesse modelo, abrem-se posições de negociação de mercado. Tipicamente (Caldeira et al., 2013) os algoritmos de pairs trading têm duas etapas principais. A primeira etapa consiste na seleção dos pares para realizar a negociação. Para pares detectados cointegrados, a segunda parte do algoritmo consiste em criar sinais de negociação baseados em regras de negociação que determinam quando abre-se e quando fecha-se uma posição.

No presente trabalho, a seleção dos pares foi baseada na abordagem de cointegração de Engle-Granger. Tal abordagem começa testando se todas as séries de ativos pré-selecionadas compartilham mesma ordem de integração  $I(1)$  e, em seguida, avalia-se a estacionariedade do *spread* entre os ativos. Entre os testes existentes para verificar a ordem de integração de uma série temporal, dois comuns são o teste Augmented Dickey-Fuller (ADF) e o teste Phillips-Perron (PP). Ambos os testes podem

ser usados na segunda parte da abordagem (teste de estacionariedade do *spread*), contudo, utilizam-se também outros testes na literatura, como o teste de Johansen (Caldeira et al, 2013).

Os pares utilizados foram selecionados a um nível de 5% de significância pelo teste ADF nas duas etapas do teste de cointegração. Avaliou-se também os resultados, em alguns períodos, das carteiras de pares selecionados utilizando os testes PP e de Johansen.

Para a criação dos sinais de negociação, calcula-se o spread do par no tempo  $t$  como

$$\varepsilon_t = y_t - \beta_1 x_t. \quad (2)$$

Em seguida, computa-se o resíduo padronizado como

$$Z_t = \frac{\varepsilon_t - \mu_\varepsilon}{\sigma_\varepsilon}. \quad (3)$$

Os parâmetros desconhecidos  $\mu_\varepsilon$  e  $\sigma_\varepsilon$  foram estimados como a média e desvio-padrão amostrais dos resíduos. Neste trabalho, usou-se uma janela móvel amostral dos 63 dias correntes (3 meses). No mercado americano, essa quantidade é chamada de um quarter. O escore  $z_t$  mede a distância em relação à média de longo prazo em unidades de desvios padrões.

Como regra na negociação, vamos abrir uma posição quando o escore  $z_t$  atingir os limites de +/- 2 desvios padrão acima ou abaixo. Esta situação implica que as ações são mal precificadas em termos de seus valores relativos. Se  $z_t$  atingir o limiar de desvio-padrão -2, isso significa que o portfólio de ativos está abaixo do seu valor de equilíbrio de longo prazo. Nesse caso, deve-se comprar o portfólio, o que significa comprar a ação  $y$  e vender a ação  $x$ . Se  $z_t$  atingir o limite de 2 desvios padrão acima de 0, o portfólio é supervalorizado e deve ser vendido *short*, o que significa vender a ação  $y$  e comprar a ação  $x$ .

Assumimos, neste trabalho, a arbitragem estatística em seu sentido puro onde não há desembolso de caixa. Assim, vendemos uma ação como short e compramos  $\alpha$  unidades da outra ação. Os sinais utilizados para abrir e fechar posições são descritos no esquema a seguir:

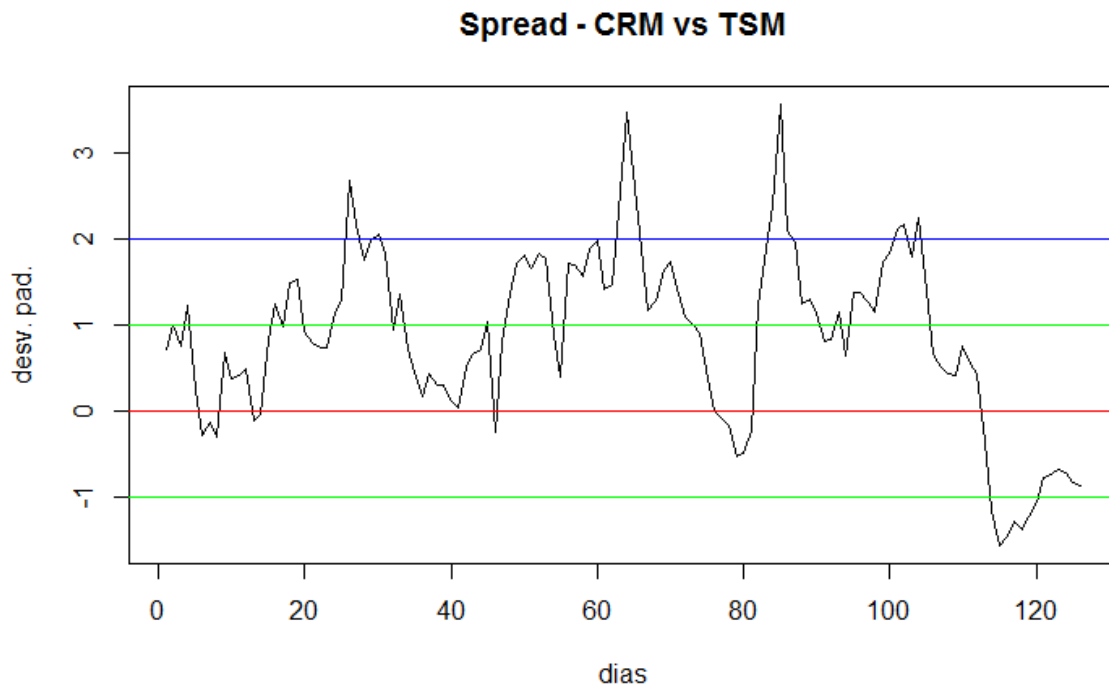
Abre-se a posição *long* se  $z_t \leq -2$ ;

Fecha-se a posição *long* se  $z_t > -1$ ;

Abre-se a posição *short* se  $z_t \geq 2$ ;

Fecha-se a posição *short* se  $z_t < 1$ .

Pode-se ter uma ideia melhor das regras de operação observando a figura a seguir. Durante o período, abriu-se a posição short no par em quatro momentos distintos (quando  $z_t$  cruza 2 desvios acima). Fechou-se a posição short quatro vezes, sempre que  $z_t$  se desloca de 2 desvios acima para abaixo de 1 desvio padrão.



**Figura 2:** spread padronizado entre os ativos da CRM e da TSM, de Jan/2009 a Jun/2010.

Neste trabalho, assim como Caldeira et al. (2013), os limites para abertura e fechamento de posição foram selecionados de maneira ad hoc.

## Custos de Transação

Os custos transacionais desempenham um papel vital na lucratividade de uma estratégia de pairs trading (Rad et al., 2016). Cada execução completa de pair trading envolve uma negociação de ida e outra de volta. Além disso, existem custos de corretagem e de aluguel de ativos (no caso de *short selling*), o que pode degradar os lucros da estratégia.

Rad et al. (2016) envolvem ações oriundas das mesmas bolsas de valores das ações utilizadas neste trabalho. Não obstante, essa é uma publicação relativamente recente. Portanto, utilizou-se neste trabalho os mesmos custos transacionais relatados por eles, que foram de 20 pontos base (0,2%) a serem descontados do retorno de cada operação *long-short*. Pontos base geralmente são usados quando se lida com quantidades baixas em percentagem e seu uso também remove a ambiguidade que pode surgir quando se está em dúvida se falamos de uma alteração relativa ou absoluta. Por exemplo, uma taxa de 10% acrescida de 5% pode ser interpretada como 10,5% ou como 15%. Se a taxa sobe em 50 bps, ela vale 10,5%, mas se sobe em 500 bps, ela vai para 15%.

## Cálculo dos Retornos

Conforme documentado em Broussard & Vaihekoski (2012), para se calcular o retorno de um par de ações em um determinado período de negociação, é preciso acumular retornos diários das posições long e short. Os retornos percentuais diários para os ativos x e y no tempo t são computados do seguinte modo:

$$r_t = \frac{p_t}{p_{t-1}} - 1 \quad (4)$$

Obtém-se o log-retorno diário do i-ésimo par de ativos em uma carteira pela equação

$$r_{i,t} = \ln \left[ \frac{P_t^l}{P_{t-1}^l} \right] - \ln \left[ \frac{P_t^s}{P_{t-1}^s} \right]. \quad (5)$$

Caldeira et al. (2013) sugere que o retorno diário de uma carteira composta por N pares negociados no mesmo dia pode ser avaliado dessa forma:



$$R_t^{net} = \sum_{i=1}^N w_{i,t} r_{i,t} + 2n \ln \left( \frac{1-C}{1+C} \right) \quad (6)$$

onde  $C$  diz respeito ao custo de transação e  $w_{i,t}$  ao peso do  $i$ -ésimo par na carteira no dia  $t$ . Neste trabalho, ponderamos todos os pares de forma igual assumindo  $w_{i,t} = \frac{1}{N}$ .

Para avaliar o resultado na estratégia de pairs trading, utilizou-se a rentabilidade anualizada ( $R^A$ ), desvio padrão anualizado dos retornos ( $\sigma^A$ ) e o índice de Sharpe anualizado ( $SR^A$ ). Calculamos essas estatísticas como:

$$R^A = \left[ \prod_{i=1}^T (1 + r_t) \right]^{\frac{252}{T}} - 1 \quad (7)$$

$$\sigma^A = \sqrt{\frac{1}{T} \sum_{t=1}^T (r_t - \mu_r)^2} \quad (8)$$

$$SR^A = \frac{R^A}{\sigma^A} \quad (9)$$

O retorno anualizado se refere ao retorno geométrico anual, mostrando os ganhos reais que o investimento poderia gerar em um ano. O desvio padrão anualizado avalia o risco associado ao portfólio de ativos (a volatilidade do investimento) mensurando o grau de dispersão em torno do retorno médio (Alencastro e Martins, 2013). O índice de Sharpe, proposto por William Sharpe (1994), é uma medida ajustada do risco para retornos de investimento. Esse indicador expressa com qual nível de risco uma aplicação oferece determinada rentabilidade. Quanto maior for o índice de Sharpe, melhor é performance do investimento em relação ao risco que ele oferece.

## Dados

Geralmente em estratégias de pairs trading, requer-se que os pares dos ativos sejam do mesmo segmento de mercado. Assim, os dados utilizados neste trabalho consistem de cotações de fechamento diários de 20 ações do ramo tecnológico cotadas em bolsas norte-americanas, onde selecionou-se ativos com volume diário atual (novembro de 2018) de ao menos um milhão de dólares (US\$1 milhão) e com os maiores valores de mercado. Essa filtragem na busca das ações foi feita no intuito de se trabalhar com ações que apresentem liquidez elevada, dado que utilizar ações com baixa liquidez pode implicar em maiores custos operacionais (Caldeira et al., 2013). Dentre as ações selecionadas, sete são cotadas na bolsa de Nova York – NYSE – enquanto treze são cotadas na bolsa de tecnologia norte-americana – Nasdaq. Coletou-se, através da plataforma Yahoo Finance, séries de cotações diárias ao longo de cinco anos, entre os dias 1 de janeiro de 2008 e 2 de janeiro de 2013, acumulando 1260 observações diárias. O mercado de ações norte-americano é aberto, em média, durante 252 dias ao longo de um ano.

## Software

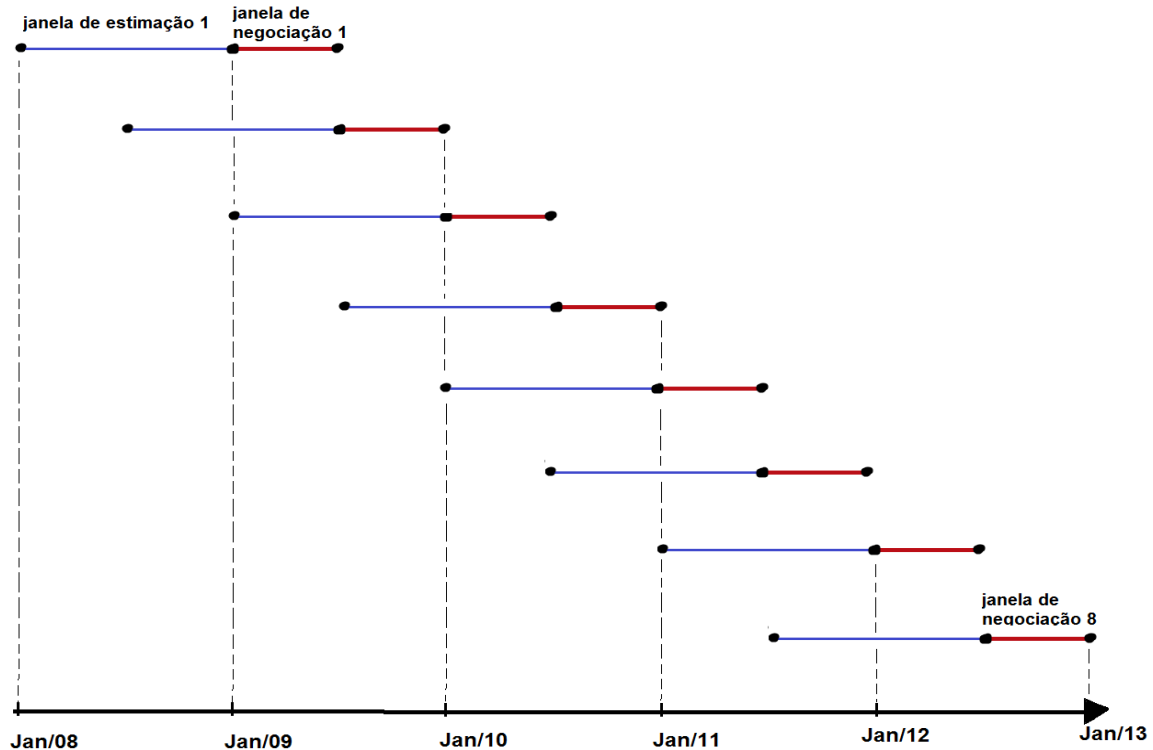
Todas os dados e análises neste trabalho foram computados no software R, com a interface R Studio (versão 1.1.383). Utilizou-se o recurso dos seguintes pacotes no R:

Pacote	Título
egcm	Engle-Granger Cointegration Models
quantmod	Quantitative Financial Modelling Framework
PerformanceAnalytics	Econometric Tools for Performance and Risk Analysis
urca	Unit Root and Cointegration Tests for Time Series Data
tseries	Time Series Analysis and Computational Finance

## Estimação e Negociação

Os dados amostrais foram divididos em um período de treinamento (estimação) e um período de teste (negociação). No período de treinamento, estimam-se os parâmetros do experimento. Logo após o término do período de estimação, seguimos com o período de teste, onde aplicamos a estratégia de negociação usando os parâmetros estimados no período anterior. Neste trabalho, utilizou-se o mesmo critério estabelecido por Gatev et al. (2006): 12 meses no período de treinamento (252 dias de cotações) e 6 meses de negociação (126 dias).

Testes de cointegração são realizados em todos os possíveis pares cujas séries dos ativos foram detectadas da ordem  $I(1)$  e os parâmetros da regressão são estimados com base nos primeiros 252 dias da amostra. Selecionados os pares e computados os parâmetros, seis meses de negociação são executados nos próximos dias da amostra. Ao final de cada período de negociação, as posições atuais que permaneceram abertas são fechadas e um novo período de estimação é escolhido de forma que o período de negociação seguinte se inicie imediatamente ao período de negociação atual. Nessa altura, uma nova seleção de pares é realizada e os parâmetros são re-estimados para um período subsequente de negociação. Essa janela móvel continua até chegar no final da amostra, tendo um total de oito períodos de trading. A figura abaixo ilustra como os períodos de estimação e negociação se interseccionam.



**Figura 3:** Esquema de negociação ao longo do tempo.

### Resultados da estratégia

Sendo a nossa base composta por vinte séries de preços diários de ações, existem  $\binom{20}{2} = 190$  pares possíveis de negociação em cada período de trading. Na abordagem testando cointegração utilizando o teste Augmented Dickey-Fuller, obteve-se um número mínimo de 3 pares cointegrados e um número máximo de 38 pares cointegrados. Na abordagem pelo teste de Phillips-Perron, os números mínimo e máximo foram, respectivamente, 4 e 40. Já pelo teste de Johansen, o número mínimo foi 3 e o máximo 32.

Quando rodamos a metodologia de EG utilizando o teste Phillips-Perron, nos primeiros dois períodos de negociação, os retornos anualizados das carteiras foram negativos, assumindo, respectivamente, -5,37% e -9,33% (sem custos de corretagem). Além disso, diversos pares que passaram no teste de cointegração por PP foram

reprovados pelo teste ADF. Dentre esses pares reprovados, a maioria trouxe retornos negativos.

Com a nova seleção de pares por ADF, os retornos anualizados nos dois primeiros períodos foram de 5,3% e -4,1% (sem corretagens). Ao adotar o teste de Johansen na metodologia de EG, obtivemos retornos anualizados de 4,6% (período 1) e 3,2% (período 2).

Testamos também um critério mais rigoroso na seleção dos pares para realizar a estratégia de pairs trading, onde pares passaram na metodologia de Engle-Grenger tanto pelo teste ADF quanto pelo teste de Johansen. Dos 27 pares que cointegraram previamente por ADF no primeiro período de negociação, apenas 15 permaneceram levando em conta os dois testes. No segundo período, esse número foi de 26 a 9 pares. Os retornos, desvios padrões e índices de Sharpe anualizados nos primeiros dois períodos foram:

**Tabela 1: Retornos estratégias pelos testes ADF e Johansen**

		<b>log retorno anualizado</b>	<b>Desvio Padrão Anualizado</b>	<b>Índice de Sharpe Anualizado</b>
<b>1º período</b>	<b>Sem corretagens</b>	12.0%	0.0772	1.5578
	<b>Com corretagens</b>	11.9%	0.0772	1.5442
<b>2º período</b>	<b>Sem corretagens</b>	2.9%	0.0622	0.4589
	<b>Com corretagens</b>	2.7%	0.0621	0.4388

Pode-se perceber que, nesses 12 meses de operação, o teste ADF conjuntamente com o teste de Johansen trouxe retornos anualizados e índices de Sharpe mais atrativos que os apresentados previamente.

Abaixo, são mostrados os resultados da estratégia utilizando apenas o teste ADF na abordagem de Engle-Grenger.

### **Resultados do período 1 (30/12/2008 a 01/07/2009)**

**Tabela 2:** Resultados dos pares no primeiro período de negociação.

Par	log retorno (ao ano)	log retorno com corretagem (ao ano)
CRM_TSM	81.5%	81.3%
TSM_TMUS	81.1%	80.9%
ADBE_QCOM	66.4%	66.2%
IBM_GOOG	51.0%	50.7%
IBM_GOOGL	51.0%	50.7%
TSM_GOOG	33.4%	33.2%
TSM_GOOGL	33.4%	33.2%
CRM_GOOG	21.5%	21.4%
CRM_GOOGL	21.5%	21.4%
BIDU_IBM	19.5%	19.4%
ADBE_IBM	16.5%	16.3%
ACN_ORCL	9.3%	9.2%
IBM_ORCL	6.4%	6.3%
IBM_VZ	3.5%	3.5%
CRM_IBM	3.0%	2.9%
T_VZ	-3.4%	-3.5%
IBM_MSFT	-8.8%	-8.9%
CRM_INTC	-13.7%	-13.7%
IBM_INTC	-17.8%	-17.9%
TSM_VZ	-18.9%	-19.0%
ADBE_ACN	-21.9%	-22.0%
CRM_ORCL	-29.0%	-29.1%
ACN_QCOM	-29.6%	-29.7%
BIDU_INTC	-32.7%	-32.8%
TMUS_VZ	-32.8%	-32.9%
BIDU_ORCL	-34.7%	-34.8%
CRM_VZ	-34.8%	-34.8%

**Tabela 3:** Resultado da carteira no primeiro período de negociação.

	Log-Retorno Anualizado	Desvio Padrão Anualizado	Índice de Sharpe Anualizado
Sem corretagens	5.32%	0.0635	0.8376
Com corretagens	5.22%	0.0635	0.8216

**Resultados do período 2 (02/07/2009 a 30/12/2009)**

**Tabela 4:** Resultados dos pares no segundo período de negociação.

Par	log retorno anualizado	log retorno anualizado (c/ corretagem)
IBM_QCOM	24.7%	24.5%
CRM_NVDA	19.8%	19.6%
QCOM_TXN	19.2%	19.0%
ORCL_QCOM	18.5%	18.3%
BIDU_TXN	16.0%	15.8%
IBM_GOOG	8.8%	8.7%
IBM_GOGL	8.8%	8.7%
ADBE_CSCO	8.0%	7.9%
BIDU_GOOG	1.2%	1.1%
BIDU_GOGL	1.2%	1.1%
INTU_TXN	0.6%	0.5%
INTC_TXN	-0.2%	-0.2%
CRM_INTC	-0.9%	-1.0%
GOOG_NVDA	-5.3%	-5.4%
GOOGL_NVDA	-5.3%	-5.4%
BIDU_INTU	-5.6%	-5.7%
TSM_NVDA	-10.7%	-10.7%
IBM_NVDA	-13.0%	-13.1%
INTU_GOOG	-14.9%	-15.0%
INTU_GOGL	-14.9%	-15.0%
BIDU_ORCL	-16.8%	-16.9%
ORCL_GOOG	-26.2%	-26.3%
ORCL_GOGL	-26.2%	-26.3%
GOOG_QCOM	-26.6%	-26.7%
GOOGL_QCOM	-26.6%	-26.7%
QCOM_NVDA	-36.9%	-36.9%

**Tabela 5:** Resultado da carteira no segundo período de negociação.

	Log-Retorno Anualizado	Desvio Padrão Anualizado	Índice de Sharpe Anualizado
Sem corretagens	-4.1%	0.0533	-0.7774
Com corretagens	-4.2%	0.0533	-0.7963

**Resultados do período 3 (31/12/2009 a 01/07/2010)**

**Tabela 6:** Resultados dos pares no terceiro período de negociação.

Par	log retorno anualizado	log retorno anualizado (c/ corretagem)
NVDA_VZ	39.1%	39.0%
ADBE_IBM	17.5%	17.1%
ADP_TMUS	16.6%	16.5%
GOOG_MSFT	13.3%	13.2%
GOOGL_MSFT	13.3%	13.2%
ACN_IBM	12.1%	11.9%
ADP_T	10.2%	10.1%
IBM_T	5.6%	5.5%
TXN_INTU	3.4%	3.2%
TXN_ORCL	-1.3%	-1.4%
GOOG_T	-1.6%	-1.7%
GOOGL_T	-1.6%	-1.7%
NVDA_T	-6.2%	-6.3%
IBM_ORCL	-7.4%	-7.5%
ACN_TMUS	-9.5%	-9.6%
IBM_INTC	-12.4%	-12.5%
CRM_T	-15.3%	-15.3%
CSCO_IBM	-20.3%	-20.4%
IBM_MSFT	-27.9%	-28.0%
ADBE_BIDU	-36.2%	-36.3%

**Tabela 7:** Resultados da carteira no terceiro período de negociação.

	Log-Retorno Anualizado	Desvio Padrão Anualizado	Índice de Sharpe Anualizado
Sem corretagens	-0.9%	0.0439	-0.1928
Com corretagens	-1.0%	0.0439	-0.2189

### Resultados do período 4 (02/07/2010 a 30/12/2010)

**Tabela 8:** Resultado dos pares no quarto período de negociação.

Par	log retorno anualizado	log retorno anualizado (c/ corretagem)
MSFT_ADBE	49.1%	49.0%
ADBE_T	37.3%	37.2%
TSM_IBM	27.1%	26.9%
IBM_NVDA	21.9%	21.8%
GOOG_ADBE	19.1%	18.9%



GOOGL_ADBE	19.1%	18.9%
GOOG_IBM	18.4%	18.4%
GOOGL_IBM	18.4%	18.4%
QCOM_T	17.0%	16.9%
TXN_ORCL	13.9%	13.9%
BIDU_IBM	7.5%	7.4%
TMUS_IBM	5.6%	5.5%
MSFT_ACN	5.3%	5.2%
TSM_TXN	4.5%	4.4%
TMUS_TXN	4.5%	4.4%
GOOG_TXN	1.3%	1.1%
GOOGL_TXN	1.3%	1.1%
T_VZ	0.8%	0.7%
INTU_IBM	-0.7%	-0.8%
IBM_T	-0.9%	-1.0%
ACN_ADP	-1.4%	-1.5%
ADP_IBM	-1.6%	-1.6%
ACN_IBM	-2.4%	-2.5%
IBM_QCOM	-3.7%	-3.7%
MSFT_IBM	-4.8%	-4.8%
CRM_IBM	-5.0%	-5.2%
NVDA_TXN	-7.5%	-7.5%
VZ_TXN	-10.1%	-10.2%
INTU_TXN	-10.6%	-10.6%
IBM_ORCL	-13.2%	-13.3%
ACN_TXN	-13.3%	-13.3%
MSFT_TXN	-13.6%	-13.8%
IBM_VZ	-15.6%	-15.6%
QCOM_TXN	-17.5%	-17.5%
ADP_TXN	-17.8%	-17.8%
TSM_VZ	-24.8%	-24.9%
T_TSM	-28.0%	-28.1%
BIDU_TXN	-37.4%	-37.5%

**Tabela 9:** Resultado da carteira no quarto período de negociação.

	<b>Log-Retorno Anualizado</b>	<b>Desvio Padrão Anualizado</b>	<b>Índice de Sharpe Anualizado</b>
Sem corretagens	1.0%	0.0532	0.1824
Com corretagens	0.9%	0.0531	0.1652

### Resultados do período 5 (31/12/2010 a 30/06/2011)

**Tabela 10:** Resultado dos pares no quinto período de negociação.

Par	log retorno anualizado	log retorno anualizado (c/ corretagem)
TMUS_BIDU	38.7%	38.5%
ACN_QCOM	24.3%	24.1%
ORCL_QCOM	13.9%	13.8%

**Tabela 11:** Resultado da carteira no quinto período de negociação.

	Log-Retorno Anualizado	Desvio Padrão Anualizado	Índice de Sharpe Anualizado
Sem corretagens	26.6%	0.1036	2.5678
Com corretagens	26.4%	0.1035	2.5536

### Resultados do período 6 (01/07/2011 a 29/12/2011)

**Tabela 12:** Resultado dos pares no sexto período de negociação.

Par	log retorno anualizado	log retorno anualizado (c/ corretagem)
BIDU_INTU	14.9%	14.8%
ADBE_QCOM	-5.3%	-5.4%
TSM_ADBE	-11.1%	-11.1%
VZ_ADBE	-22.8%	-22.9%
ADBE_IBM	-24.8%	-24.8%
ORCL_VZ	-29.7%	-29.8%
ADBE_NVDA	-38.3%	-38.3%

**Tabela 13:** Resultado da carteira no sexto período de negociação.

	Log-Retorno Anualizado	Desvio Padrão Anualizado	Índice de Sharpe Anualizado
Sem corretagens	-16.7%	0.1289	-1.2947
Com corretagens	-16.8%	0.1289	-1.3016

### Resultados do período 7 (30/12/2011 a 29/06/2012)

**Tabela 14:** Resultado dos pares no sétimo período de negociação.

Par	log retorno anualizado	log retorno anualizado (c/ corretagem)
ADBE_TXN	17.3%	17.2%
TSM_QCOM	10.2%	10.1%
ADBE_QCOM	8.0%	7.9%
CSCO_BIDU	1.8%	1.7%

**Tabela 15:** Resultado da carteira no sétimo período de negociação.

	Log-Retorno Anualizado	Desvio Padrão Anualizado	Índice de Sharpe Anualizado
Sem corretagens	10.6%	0.0931	1.1389
Com corretagens	10.5%	0.0931	1.1263

### Resultados do período 8 (02/07/2012 a 02/01/2013)

**Tabela 16:** Resultado dos pares no oitavo período de negociação.

Par	log retorno anualizado	log retorno anualizado (c/ corretagem)
TSM_ADBE	48.7%	48.4%
TMUS_VZ	33.6%	33.5%
ADBE_IBM	26.9%	26.7%
CSCO_GOOG	26.7%	26.6%
CSCO_GOOGL	26.7%	26.6%
GOOG_ADBE	16.9%	16.7%
GOOGL_ADBE	16.9%	16.7%
ADP_GOOG	11.2%	11.2%
ADP_GOOGL	11.2%	11.2%
IBM_QCOM	10.6%	10.5%
ADP_IBM	5.2%	5.1%
INTU_IBM	3.3%	3.2%
GOOG_QCOM	2.7%	2.7%
GOOGL_QCOM	2.7%	2.7%
TXN_CSCO	2.1%	2.0%
MSFT_ADBE	-5.9%	-6.1%
GOOG_TXN	-7.8%	-7.9%
GOOGL_TXN	-7.8%	-7.9%
ACN_IBM	-7.9%	-8.0%
TMUS_BIDU	-12.9%	-13.0%









**Tabela 25:** Retornos anualizados dos pares no período 8.

	MSFT	GOOGL	GOOG	VZ	T	INTC	CSCO	TSM	ORCL	NVDA	ADBE	IBM	CRM	ACN	QCOM	TXN	BIDU	ADP	TMUS	INTU	
MSFT	-																				
GOOGL		-																			
GOOG			-																		
VZ				-																	
T					-																
INTC						-															
CSCO		26.7%	26.7%				-														
TSM								-													
ORCL									-												
NVDA										-											
ADBE	-5.9%	16.9%	16.9%			-52.3%		48.7%													
IBM						-21.3%					26.9%										
CRM																					
ACN																					
QCOM		2.7%	2.7%																		
TXN		-7.8%	-7.8%				2.1%														
BIDU																					
ADP		11.2%	11.2%			-30.1%					5.2%										
TMUS				33.6%	-47.5%																
INTU						-19.4%					3.3%										

No período de operação 1, atingimos retornos anualizados acima de 80% em dois pares e retornos positivos em outros 13 pares da carteira, contra 12 pares com retornos negativos. Os pares que mais detratarem a rentabilidade da estratégia atingiram retornos anualizados em torno de -30%. Ainda assim, obtivemos um retorno anualizado médio da carteira maior que 5%. No período 2, os pares com resultados positivos não alcançaram retornos anualizados maiores que 25% e a proporção de pares com retornos negativos foi maior que a de pares com retornos positivos. O retorno mais negativo de um par ficou perto de -37% e o menor retorno médio em carteira ficou em torno de -4% ao ano. Nos períodos 3, 4 e 8, os retornos médios do portfólio ficaram em torno de -2% e 1%. O portfólio do período 5 foi o que trouxe o maior retorno anualizado (> 26%) e o maior índice de Sharpe (> 2,55) da estratégia, dado que todos os pares apresentaram retornos em excesso. Nos períodos 6 e 7, obtivemos retornos anualizados de -16.8% e 10.5% respectivamente, levando em conta os custos de corretagem estimados.



## Conclusão

Neste trabalho, implementou-se a estratégia de pairs trading com pares de ações das bolsas americanas Nasdaq e NYSE, onde dados históricos foram coletados de janeiro de 2008 a janeiro de 2013, acumulando 1260 cotações de fechamento diárias. Foram selecionados os vinte ativos com maiores valores de mercado e com volume diário de pelo menos 1 milhão de dólares (no mês de novembro de 18) no intuito de não se operar com ações com baixa liquidez.

Na seleção dos pares, utilizou-se a metodologia de Cointegração de Engle-Granger, a qual consiste de duas etapas, a primeira testando se as séries de preços dos ativos compartilham a mesma ordem de integração  $I(1)$  e em seguida avaliando se a combinação linear formada entre os ativos é uma série estacionária. Para testar ordem de integração, foram utilizados os testes Augmented Dickey Fuller (ADF), Phillips–Perron (PP) e Johansen. Nos oito períodos de negociação, havia 190 pares possíveis de operação, onde o número mínimo de pares cointegrados foi 3 e o número máximo de pares foi 40.

No primeiro ano de negociação, compararam-se os retornos das carteiras de pares selecionados pelos três testes citados previamente na metodologia de cointegração de EG. O teste PP demonstrou baixa performance nos dois períodos de seis meses de negociação, atingindo retornos anualizados (sem custos de corretagem) de -5,4% no período 1 e de -9,33% no período 2. Obtiveram-se também retornos nesse mesmo período usando o teste ADF, o teste de Johansen, e um teste mais rígido combinado entre esses dois últimos testes. No teste combinado, atingiu-se um retorno anualizado em torno de 12% no primeiro período e em torno de 3% no segundo período.

No período completo de negociação da estratégia de pairs trading (4 anos), o maior retorno anualizado de um par foi de 81,5% e o maior retorno anualizado médio em uma carteira de pares foi de 26.6%, com um índice de Sharpe de 2,56. Por outro

lado, o menor retorno anualizado de um par foi de -52,3% e o menor retorno anualizado médio em uma carteira de pares de -16,7%. Nos oito períodos de operação, obteve-se retornos médios positivos em quatro dos períodos, contra retornos negativos nos outros quatro períodos.

## Referências

FROIS CALDEIRA, João; VALLE MOURA, Guilherme. Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy. **Revista Brasileira de Finanças**, v. 11, n. 1, 2013.

GATEV, Evan; GOETZMANN, William N.; ROUWENHORST, K. Geert. Pairs trading: Performance of a relative-value arbitrage rule. **The Review of Financial Studies**, v. 19, n. 3, p. 797-827, 2006.

BROUSSARD, John Paul; VAIHEKOSKI, Mika. Profitability of pairs trading strategy in an illiquid market with multiple share classes. **Journal of International Financial Markets, Institutions and Money**, v. 22, n. 5, p. 1188-1201, 2012.

VIDYAMURTHY, Ganapathy. **Pairs Trading: quantitative methods and analysis**. John Wiley & Sons, 2004.

ENGLE, Robert F.; GRANGER, Clive WJ. Co-integration and error correction: representation, estimation, and testing. **Econometrica: journal of the Econometric Society**, p. 251-276, 1987.

RAD, Hossein; LOW, Rand Kwong Yew; FAFF, Robert. The profitability of pairs trading strategies: distance, cointegration and copula methods. **Quantitative Finance**, v. 16, n. 10, p. 1541-1558, 2016.

JAKOBSSON, Erik. A new approach to Pairs Trading: Using fundamental data to find optimal portfolios. Master Thesis, UMEA UNIVERSITY, 2015.

BRITO, Igor Arantes. **Uma nova estratégia de negociação em pares**. Tese de Doutorado, FGV-SP, 2014.

INAMDAR, Aniket et al. Statistical Arbitrage: Asset clustering, market-exposure minimization, and high-frequency explorations. **MS&E 448, Spring**, Stanford, 2016.

ALENCASTRO, Denilson; MARTINS, Marco Antônio dos Santos. Interpretando o Índice Sharpe. Disponível em: < <http://gilsonrossato.livejournal.com/6060.html>>. Acesso em: 24 jun. 2013.

## Apêndice

```

#funcao para visualizar pares que cointegram:
make_matrix= function(matriz,tab){
  colnames(tab)<- colnames(matriz)
  rownames(tab)<- colnames(matriz)
  for (i in 1:ncol(matriz)) {
    for (j in 1:ncol(matriz)) {
      temp=egcm(matriz[,i],matriz[,j], iltest = "adf", urtest = "adf")
      tab[i,j]= ifelse(as.numeric(temp$r.p)>0.05 |
rownames(tab)[i]==colnames(tab)[j],NA,paste(rownames(tab)[i],",",colnames(tab)[j]))
    }
  }
  tab}
#DEMAIS FUNCOES:
xy_plot<- function(x,y){
  plot.zoo(cbind(x,y),plot.type = "single",lty=c(1,2),lwd =2.2,
  col =c("blue","red") , xlab = "dias",
  ylab = "Cotações de Fechamento Diário");
  legend("topleft",legend=c(colnames(x),colnames(y)),col=c("blue","red"),lty=1,cex=0.75)
}
testing_func<- function(database,start,end){
  #beta da regressao
  x<- database[,1]
  y<- database[,2]
  t0<- start-252
  t1<- start-1
  temp= egcm(x[t0:t1],y[t0:t1],iltest = "adf",urtest = "adf")
  estatistica_teste= temp$r.p
  estatistica_teste}
#funcao para criacao de colunas:
Add_col= function(database){
  database$residuals=0
  database$std_res=0
  database$Trading_signals=NA
  database$alpha_betazero=0
  database$x_trade_price=0
  database$y_trade_price=0
  database$pair_log_ret=0
  database$returns_c=0
  database}
function(database,start,end){
  #beta da regressao
  x<- database[,1]
  y<- database[,2]
  t0<- start-252
  t1<- start-1
  temp= egcm(x[t0:t1],y[t0:t1], iltest = "adf", urtest = "adf")
  b1<- temp$beta
  #residuos
  database$residuals[t0:end]= y[t0:end] - b1*x[t0:end]
  #residuos padronizados (janela movel de 3 meses)
  for(i in (start-1):end){database$std_res[i]= (database$residuals[i]-mean(database$residuals[(i-
63+1):(i)]))/
sd(database$residuals[(i-63+1):(i)])}
}
#sinais
database$Trading_signals[start-1]<- "close"
for(i in start:end){
  database$Trading_signals[i]<- ifelse(( database$std_res[i]< -2 & database$std_res[i-1]> -2 )|(
database$std_res[i] < -2 &
database$Trading_signals[i-1]=="close" ),
"open long y",
ifelse(( database$std_res[i] > 2 & database$std_res[i-1] < 2 )|( database$std_res[i] > 2 &
database$Trading_signals[i-1]=="close" ),
"open short y",
ifelse(( database$std_res[i] > -1 & database$std_res[i-1]< -1 & database$Trading_signals[i-1]
=="open long y" ),
"close",

```

```

ifelse(( database$std_res[i] < 1 & database$std_res[i-1] > 1 &
database$Trading_signals[i-1] == "open short y" ),
"close", database$Trading_signals[i-1]))
}
database$Trading_signals[end]<- "close"
#alpha para estrategia betazero:
database$alpha_betazero[start-1]<- 0
for(i in start:end){database$alpha_betazero[i] <- ifelse(database$Trading_signals[i] == "open
short y" & database$Trading_signals[i-1] ==
"close",
database[i,2]/database[i,1],
ifelse(database$Trading_signals[i] == "open long y" & database$Trading_signals[i-1]
== "close",
database[i,1]/database[i,2],
ifelse(database$Trading_signals[i] == "close" & database$Trading_signals[i-
1] == "close",
0,database$alpha_betazero[i-1]))))
}
#precos de negociacao com market neutral:
for(i in start:end){
database$y_trade_price[i]<- ifelse(( database$Trading_signals[i] == "open short y" | (
database$Trading_signals[i] == "close" &
database$Trading_signals[i-1] == "open short y" )),
database[i,2],
ifelse(( database$Trading_signals[i] == "open long y" | ( database$Trading_signals[i] == "close" &
database$Trading_signals[i-1] == "open long y" )),
database[i,2]*database$alpha_betazero[i],0))
database$x_trade_price[i]<- ifelse(( database$Trading_signals[i] == "open short y" | (
database$Trading_signals[i] == "close" &
database$Trading_signals[i-1] == "open short y" )),
database[i,1]*database$alpha_betazero[i],
ifelse(( database$Trading_signals[i] == "open long y" | ( database$Trading_signals[i] == "close" &
database$Trading_signals[i-1] == "open long y" )),
database[i,1],0))
}
#log retorno:
for(i in start+1:end){
database$pair_log_ret[i]<- ifelse((database$Trading_signals[i]=="open long y" &
database$Trading_signals[i-1]=="open long y") | (
database$Trading_signals[i]=="close" & database$Trading_signals[i-1]=="open long y" ),
log(database$y_trade_price[i]/database$y_trade_price[i-1]) -
log(database$x_trade_price[i]/database$x_trade_price[i-1]),
ifelse((database$Trading_signals[i]=="open short y" & database$Trading_signals[i-1]=="open short
y") | (
database$Trading_signals[i]=="close" & database$Trading_signals[i-1]=="open short y" ),
log(database$x_trade_price[i]/database$x_trade_price[i-1]) -
log(database$y_trade_price[i]/database$y_trade_price[i-1]),0))
}
#retorno com custos transacionais:
for(i in start+1:end){
database$returns_c[i]<- ifelse(( database$Trading_signals[i-1]=="open long y" &
database$Trading_signals[i]=="close" ) |
( database$Trading_signals[i-1]=="open short y" & database$Trading_signals[i]=="close" ),
database$pair_log_ret[i] - 0.0002,
database$pair_log_ret[i])
}
database}

```