

**FUNDAÇÃO GETÚLIO VARGAS  
ESCOLA DE ECONOMIA DE SÃO PAULO**

**ALICE YANG**

**COINTEGRAÇÃO ENTRE SÉRIES DE PREÇOS NO MERCADO  
ACIONÁRIO BRASILEIRO**

**SÃO PAULO  
2011**

**ALICE YANG**

**COINTEGRAÇÃO ENTRE SÉRIES DE PREÇOS NO MERCADO  
ACIONÁRIO BRASILEIRO**

Dissertação apresentada à Escola de Economia de  
São Paulo da Fundação Getúlio Vargas, como  
requisito para obtenção do título de Mestre em  
Economia e Finanças.

Campo de Conhecimento:  
Mercado Financeiro e Econometria

Orientador: Prof. Dr. Emerson Marçal

**SÃO PAULO  
2011**

**ALICE YANG**

# **COINTEGRAÇÃO ENTRE SÉRIES DE PREÇOS NO MERCADO ACIONÁRIO BRASILEIRO**

Dissertação apresentada à Escola de Economia de  
São Paulo da Fundação Getúlio Vargas, como  
requisito para obtenção do título de Mestre em  
Economia e Finanças.

Campo de Conhecimento:  
Mercado Financeiro e Econometria

**Data de aprovação:**  
**19/08/2011**

**Banca Examinadora:**

---

**Prof. Dr. Emerson Marçal (Orientador)**  
**FGV – EEESP**

---

**Prof. Dr. João de Mendonça Mergulhão**  
**FGV – EEESP**

---

**Prof. Dr. Claudio Lucinda**  
**FEA – USP-RP**

Yang, Alice.

Cointegração entre séries de preços no mercado acionário brasileiro / Alice Yang - 2011.  
46 f.

Orientador: Emerson Marçal.

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

1. Cointegração. 2. Mercado financeiro – Modelos econométricos. 3. Estratégia – Finanças.  
4. Investimentos. 5. Mercado financeiro – Brasil. I. Marçal Emerson Fernandes. II. Dissertação  
(MPFE) - Escola de Economia de São Paulo. III. Título.

CDU 336.76(81)

*Aos meus pais e ao Rapha pela paciência durante esses dois anos.*

## **RESUMO**

Este trabalho busca avaliar a existência de cointegração das ações do mercado brasileiro e a forma de utilização da estratégia de pair trading pelos gestores. Para isso, utilizou-se preços de ações brasileiras em diferentes frequências e janelas e aplicou-se a metodologia de Johansen e Cavaliere, cuja hipótese nula refere-se a não cointegração dos pares. Os resultados mostram que há poucas relações de cointegração entre as séries analisadas, o que ratifica a necessidade de cautela na forma de implantação da técnica na construção de carteiras.

Palavras-chave: Cointegração, arbitragem estatística, estacionariedade

## **ABSTRACT**

This study seeks to assess the existence of long-run relationships between price series and how Portfolio Managers use pair trading strategy. A quantitative analysis of the Brazilian stocks is done based on Johansen and Cavaliere methodologies, whose null hypothesis assumes the non-existence of cointegration between pairs. The results showed that there are few pairs that cointegrates among the price series analyzed, which reinforce the need to portray the strategy carefully while constructing a portfolio.

Keywords: Cointegration, statistical arbitrage, stationary

## SUMÁRIO

|   |    |
|---|----|
| 1. Introdução .....   | 1  |
| 2. Utilização da estratégia por pair trading pelos gestores entrevistados .....   | 2  |
| 3. Arbitragem estatística e teste de raíz unitária.....                           | 3  |
| 4. Metodologia de Johansen (1988, 1991 e 1994) e Johansen e Juselius (1990) ..... | 4  |
| 5. Testes de cointegração com bootstrap.....                                      | 7  |
| 6. Aplicação do trabalho de Cavaliere (2009) para o mercado brasileiro .....      | 11 |
| 7. Resultados.....  | 13 |
| 8. Avaliação dos pares cointegrados.....  | 15 |
| 9. Alteração da amostra – testes de robustez .....                                | 19 |
| 10. Explicação dos Resultados .....   | 21 |
| 11. Possíveis Extensões .....   | 22 |
| 12. Conclusão .....   | 23 |
| 13. Anexos .....  | 25 |
| 14. Referências .....   | 34 |



## **LISTA DE GRÁFICO**

|  |          |
|--|----------|
| <b>Gráfico 1</b> – Excesso de retorno sobre o CDI (Certificado de Depósito Interbancário) da estratégia de pair trading em frequência mensal, de 3 fundos de investimentos geridos por empresas brasileiras..... | <b>2</b> |
|--|----------|

## LISTA DE TABELAS

|   |           |
|---|-----------|
| <b>Tabela 1</b> – Indicação de não cointegração entre pares, com 1% de significância e dados de 08/06/2006 até 16/03/2011 .....   | <b>14</b> |
| <b>Tabela 2</b> – Indicação de não cointegração entre pares, com 2,5% de significância e dados de 08/06/2006 até 16/03/2011 .....   | <b>14</b> |
| <b>Tabela 3</b> – Indicação de não cointegração entre pares, com 5% de significância e dados de 08/06/2006 até 16/03/2011 .....   | <b>14</b> |
| <b>Tabela 4</b> – Número de empresas que pertencem ao setor discriminado segundo classificação do Economática com indicação do número de pares possíveis de serem combinados dentro do mesmo setor..... | <b>16</b> |
| <b>Tabela 5</b> – Número de pares cointegrados agrupados setorialmente e em termos absolutos ....   | <b>16</b> |
| <b>Tabela 6</b> – Percentual de pares cointegrados dentro do mesmo setor .....  | <b>17</b> |
| <b>Tabela 7</b> – Indicação de não cointegração entre pares, com 1% de significância e dados de 01/03/2001 até 16/03/2011 .....   | <b>19</b> |
| <b>Tabela 8</b> – Indicação de não cointegração entre pares, com 1% de significância e dados de 01/03/2001 até 31/12/2007 .....   | <b>20</b> |
| <b>Tabela 9</b> – Indicação de não cointegração entre pares, com 1% de significância e dados de 01/01/2008 até 16/03/2011 .....   | <b>20</b> |

## 1. Introdução

Uma estratégia amplamente utilizada por gestores de fundos de investimentos domésticos e globais é conhecida como pair trading. Grosso modo, trata-se da compra de uma ação e a venda simultânea de outra ação em um momento que a razão de preços estiver fora de um padrão histórico, sendo que o ganho seria capturado no momento de convergência dessa razão.

Dentro da estratégia de pair trading implementada há diversas variações como: (i) escolha dos ativos (arbitragem<sup>1</sup> de ações que sejam do mesmo setor, de setores pouco ou muito correlacionados entre si); (ii) implementação da estratégia (utilização de dados de alta, baixa frequência); (iii) horizonte de análise (curto, médio longo prazo); (iv) implantação (sistemática ou discricionária).

Independente das diferenças, todas as estratégias buscam a captura de valor através da venda de um ativo sobre-valorizado e a compra do ativo sub-valorizado de forma “relativa”. Em outras palavras, a evolução dos preços em termos absolutos não é importante e sim o desempenho relativo de cada uma das pontas. O gestor ganhará com a estratégia se o ativo comprado subir mais em relação ao ativo vendido ou quando o ativo vendido cair mais do que o ativo comprado.

A idéia de avaliar a evolução entre a razão de preços tem sua origem em 1987, ano em que um time do banco Morgan Stanley, liderado por Nunzio Tartaglia, criou um modelo quantitativo que comprava e vendia simultaneamente diferentes ações em momentos que a razão desviasse do comportamento padrão, sob expectativa de que ela voltaria para a sua média histórica (reversão a média)<sup>2</sup>.

A lógica do modelo era vender o ativo sobre-valorizado e comprar o ativo sub-valorizado de forma relativa, estratégia que foi denominada como arbitragem estatística. De maneira simples, seria posicionar-se quando existisse uma má precificação, ou seja, quando a razão (spread) estivesse longe do valor médio.

<sup>1</sup> Vale ressaltar que o conceito de arbitragem amplamente utilizado nos mercados e da forma que sera tratado daqui em diante difere do conceito de arbitragem no sentido puro da literatura, já que relaciona-se a uma operação com risco e que tem desembolso inicial. O conceito literal de arbitragem trata de uma operação em que não há desembolso inicial e nem tampouco risco na operação.

<sup>2</sup> Ver artigo Chiu, Lukman, Modarresi e Velayutham (2011) - High Frequency Trading – Standford University

O conceito teórico por trás da estratégia é o de estacionariedade, que implica em reversão a média no longo prazo. De maneira mais formalizada, estacionariedade refere-se a uma série que tem média e variância constante no tempo e a covariância entre valores defasados da série dependente apenas da defasagem, isto é, da distancia temporal entre eles. De forma mais técnica, a estacionariedade refere-se a equação estocástica de diferenças em que a condição de estabilidade (ser estacionária) é que as raízes da equação característica sejam menores do que 1 (estejam dentro do círculo unitário).

Contextualizando para a estratégia de arbitragem estatística, dizemos que uma série estacionária pode ser obtida através do conceito de cointegração em que há pelo menos uma combinação linear entre as duas séries de preço integradas de ordem 1 que seja estacionária (reverta a média) no longo prazo.

Em uma pesquisa feita por um banco global sobre 4 gestores que aplicam a estratégia de arbitragem estatística para o mercado brasileiro, foi apontado a baixa consistência dos resultados e a baixa uniformidade entre os gestores que a utilizam, o que motivou a análise sobre cointegração entre os preços das ações brasileiras.

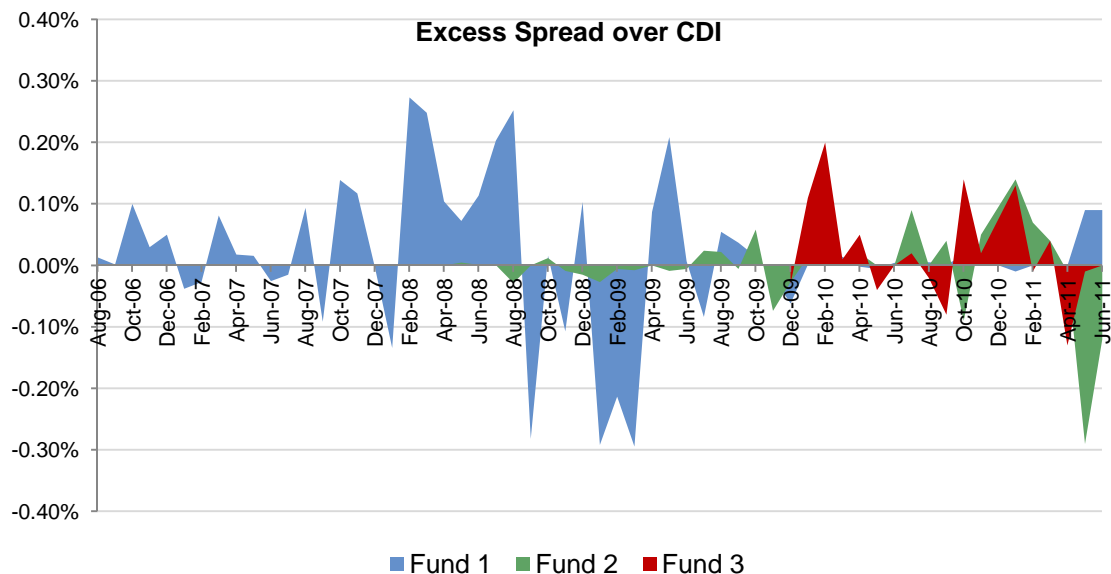
Busca-se no trabalho avaliar as relações de cointegração entre preços de ações brasileiras e verificar se há algum universo de ações mais restrito que apresente mais relações de cointegração do que o universo amplo utilizado por esses gestores.

## **2. Utilização da estratégia de pair trading pelos gestores entrevistados**

A motivação do trabalho surgiu de um estudo feito por um Banco Global em que avaliava-se a estratégia de arbitragem estatística adotada por gestores brasileiros para ativos domésticos. Verificou-se uma baixa consistência do retorno ao longo do tempo e baixa uniformidade entre diferentes gestores que utilizam a mesma estratégia, como ilustrado no Gráfico 1 abaixo.

### **Gráfico 1**

Excesso de retorno sobre o CDI (Certificado de Depósito Interbancário) da estratégia de *pair trading* em frequência mensal, de 3 fundos de investimentos geridos por empresas brasileiras.



Fonte: Elaborada pela autora

Desse 4 gestores entrevistados, 3 utilizam a técnica combinada com outras estratégias no fundo e apenas 1 utiliza exclusivamente a estratégia de *pair trading* como fonte de valor. Todos restringem a aplicação da técnica para ativos brasileiros com filtros relacionado a liquidez da ação, critério que varia entre eles.

Um ponto em comum refere-se a forma de implantação da estratégia já que a maioria deles utiliza o universo de ações que respeite o critério de liquidez, ou seja, a razão dessas ações é monitorada de forma que quando desviarem além da banda pré estabelecida (normalmente 2 desvios padrões da média), é feita a operação de compra ou venda.

### 3. Arbitragem estatística e teste de raiz unitária

A identificação da existência ou não de cointegração em séries financeiras é crucial para o sucesso da implementação da estratégia de arbitragem estatística, já que inferir que as séries cointegram quando na verdade isso não ocorre, é aplicar a estratégia de *pair trading* para ativos que não irão convergir no Longo Prazo.

O teste de cointegração, dado que busca uma relação linear que seja estacionária no Longo Prazo, só faz sentido se tivermos tratando de variáveis que não são estacionárias, que

são o caso de séries financeiras<sup>3</sup> (no caso do trabalho, de séries de preços diários das ações listadas na Bovespa).

A detecção da não estacionariedade das séries pode ser feita através do teste de raiz unitária, sendo que os testes de Dickey-Fuller (ADF –1979), Haza e Fuller (1979) e Dickey e Pantula (1987) estão entre os mais populares para procedimentos univariados. Entretanto, procedimentos multivariados são preferíveis pois não implicam a imposição de premissas restritivas que podem ser potencialmente falsas sobre os dados originais.

Séries não estacionárias são chamadas de Integrada de ordem  $d$  ( $I(d)$ ), sendo  $d \geq 1$ , se precisarem ser diferenciadas  $d$  vezes para apresentarem estacionariedade. Para identificação da ordem de integração de uma série, utiliza-se inicialmente gráficos com evolução da série combinada a análise do correlograma da mesma que unidos aos testes de verificação de raiz unitária, eliminam a subjetividade da análise visual.

As diferenças entre os testes de raiz unitária referem-se aos pressupostos assumidos e à definição da hipótese nula. O teste de Dickey Fuller (Tradicional e o Aumentado), por exemplo, tem como hipótese nula a existência de raiz unitária.

Outros testes amplamente utilizados são: (i) Phillips-Perron (PP), não paramétrico e que permite presença de alguma heterogeneidade e correlação serial nas inovações; (ii) KPSS (Kwiatkowski, Phillips, Schmidt e Shin, 1992) que difere dos anteriores por apresentar hipótese nula de estacionariedade e hipótese alternativa de presença de raiz unitária (KPSS muitas vezes é referido como teste confirmatório já que é usado como complementar aos anteriormente citados).

Apesar de amplamente utilizados, os testes univariados são alvos de críticas em função dos problemas relativos a distorções associadas ao tamanho da amostra e à sua baixa potência. Maddala e Kim (1998) sugerem versões modificadas e mais modernas dos testes ADF (como teste DF-GLS de Elliot, Rothenberg e Stock (1996)) e do PP (Ng e Perron (1996, 2001)).

#### **4. Metodologia de Johansen (1988, 1991 e 1994) e Johansen e Juselius (1990)**

<sup>3</sup> A não estacionariedade das séries financeiras é uma evidência empírica discutida por Kendall (1953), Houthakker (1961) e Osborne (1962). Engle (1982) ainda introduz o modelo ARCH (autoregressive conditional heterocedasticity) para as séries financeiras, provando que o desvio padrão dos preços não é constante no tempo.

A metodologia de Johansen tornou-se padrão para implementação de testes de cointegração. O trabalho do autor foi uma generalização do trabalho de Robert Engle e Clive Granger (1987) em que consiste em estimar a equação de uma relação linear entre duas variáveis originais e verificar se os resíduos gerados são estacionários. O método Engle-Granger (1987) recomenda dois procedimentos: (i) teste ADF e (ii) estatística Durbin-Watson<sup>4</sup>, ambos sobre o erro estimado (chamado de resíduo cointegrado). A hipótese nula  $H_0$  a ser testada é a de que as séries não são cointegradas.

A definição de Engle e Granger (1987) sobre cointegração segue abaixo:

*“Os componentes de  $X_{zt}$  são ditos cointegrados de ordem  $(d,b)$ , denotado por  $X_{zt} \sim CI(d,b)$ , se todos os componentes de  $X_{zt}$  forem integrados de ordem  $d$ , i.e.  $I(d)$ , e existir um vetor  $\beta \neq 0$  tal que  $Y_z t = \beta' X_{zt} \sim I(d-b)$ ,  $b > 0$ ”*

O caso mais comum presente nas séries financeiras são aquelas integradas de ordem 1 ( $I(1)$ ). Supondo uma série de  $N$  variáveis aleatórias  $X_t$ , a variável não estacionária pode ser representada por  $X_t \sim CI(1,1)$  que ao serem combinadas linearmente, podem tornar-se  $I(0)$  ou estacionária.

Se houver cointegração, então há um mecanismo de correção dos erros na dinâmica de curto prazo, chamado de VEC (Vector Error Correction) e cuja representação encontra-se abaixo.

$$\text{Equação 1: } \Delta X_t = \alpha \beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

Em que  $\beta$  é uma matriz ( $p \times r$ ) tal que  $p$  é o número de variáveis,  $r$  é o número de vetores de cointegração e, portanto,  $\beta' X_{t-1} \sim I(0)$ ,  $\Gamma_i$  é uma matriz  $N \times N$  e  $\varepsilon_t$  é um vetor estacionário.

A partir de um VAR ( $k$ ) dado por:

<sup>4</sup> Durbin-Watson é um teste estatístico usado para detectar a presença de autocorrelação dos resíduos de uma regressão.

Equação 2:  $X_t = \pi_1 X_{t-1} + \pi_2 X_{t-2} + \dots + \pi_k X_{t-k} + \varepsilon_t$

Pode-se obter o VEC:

Equação 3:  $\Delta X_t = \pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$

Em que:

Equação 4:  $\pi = \pi_1 + \pi_2 + \dots + \pi_k - I$  e  $\Gamma_i = -\sum_{j=i+1}^k \pi_j$

Na qual  $\varepsilon_t$  tem distribuição normal, os erros são independentes ao longo do tempo e a matriz  $\pi = \alpha\beta'$ .

A matriz  $\pi$  é o ponto central do procedimento de Johansen, sendo as colunas de  $\beta'$  os vetores de cointegração e  $\alpha$  os parâmetros de ajustamento do vetor de cointegração às séries (velocidade do ajuste ao equilíbrio).

Os vetores de cointegração  $\beta'$  representam a relação de Longo Prazo entre as séries presentes no vetor  $X_t$ , sendo que o produto  $\beta'X_{t-1}$  é chamado de termo de correção de erros (TCE). Por hipótese todas as séries presentes em  $X_t$  são integradas de ordem um  $I(1)$  e  $\varepsilon_t$  é estacionária ( $I(0)$ ).

Conclui-se portanto que para existir cointegração, o produto  $\pi X_{t-1}$  deverá ser  $I(0)$ , ou seja, estacionária e como consequência, a matriz  $\pi = \alpha\beta'$  deverá ter posto (rank) reduzido (entre 0 e  $n-1$ ), tal que  $n$  refere-se o número de variáveis.

A matriz  $\pi$ , é estimada pelo método de máxima verossimilhança e normalmente estimada de maneira irrestrita. Posteriormente, são feitos testes de restrições de posto reduzido sobre a matriz  $\pi$ .



A estimação envolve a solução de um problema de autovalor em que o posto da matriz  $\pi$  é determinado pelo número de autovalores que é diferente de 0. O número de autovalores, por sua vez, é determinado pelo número de variáveis endógenas do sistema.

Após estimação da matriz, dois testes estatísticos são implementados para identificação do posto: (i) estatística do traço em que testa-se a hipótese nula de que o número de vetores de cointegração é menor ou igual a  $r$  e (ii) teste do máximo autovalor que testa a hipótese de  $r$  vetores de cointegração contra a alternativa de  $r+1$ . Para implementação desses testes, os autovalores são ordenados em ordem decrescente e testes sequenciais são realizados.

As estatísticas dos teste encontram-se abaixo:

$$\text{Estatística do Traço: } \lambda_{trace}(r) = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i)$$

$$\text{Estatística do máximo autovalor: } \lambda_{max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1})$$

Em que  $\lambda$  são números reais entre 0 e 1, representando os autovalores de  $\pi$ ,  $T$  é o tamanho da amostra e  $n$  o número de variáveis empregadas.

Segundo Enders (1995), os resultados dos testes podem ser conflitantes e, uma vez que o teste de máximo autovalor possui hipóteses alternativas mais rigorosas, é preferido na determinação do número de vetores cointegrantes.

## 5. Testes de cointegração com bootstrap

O artigo de Cavaliere (2009) parte da análise das propriedades do teste convencional de Johansen (1988) e propõe o wild bootstrap para identificação do posto de cointegração, técnica que se diferencia do bootstrap proposto por Swensen basicamente por preservar e ser robusto a eventual heterocedasticidade nos choques do modelo.

A demonstração da performance superior do wild bootstrap em relação ao tradicional para amostras finitas que apresentam heterocedasticidade condicional nos erros, foi feito testes para estrutura a termo de juros para alguns países.

Gonçalves e Kilian (2004) demonstraram que as séries financeiras não são normalmente distribuídas e que diversos dados macroeconômicos são heterocedásticos. Além disso, mostraram que a inferência de estacionariedade em modelos autoregressivos univariados,

precedida pela aplicação do bootstrap, pode levar a resultados distorcidos quando a série original é heterocedástica - diferente do wild bootstrap, que apresenta-se válido sob condições de não estacionariedade como mostrado por Xu (2008) e Cavaliere e Taylor (2008).

A estatística do traço e do máximo autovalor do teste de Johansen são derivados da premissa de Normalidade sendo que Lee e Tse (1996) mostraram que o teste, sob a hipótese de heterocedasticidade ou erros na especificação do GARCH tendem a cometer o erro de dizer que as séries cointegram quando na verdade não cointegram (rejeitam mais do que deveriam a hipótese nula de não cointegração). O artigo mostra que esse problema é irrelevante para amostras maiores – resultado esse corroborado por Rahbek, Hansen e Dennis (2002), comprovando que a suposição de estacionariedade global para séries longas e heterocedásticas continuam válidos.

Os autores ressaltam também que apesar do teste tradicional (Johansen) para verificar o posto da matriz ser assintoticamente válido sob presença de heterocedasticidade dos erros, em nenhum momento do teste é modelado formalmente a heterocedasticidade presente nos choques.

Assim, os autores apresentam a técnica desenvolvida baseada no esquema de wild-bootstrapping, em que diferente das outras metodologias de bootstrap, considera a presença de heterocedasticidade condicional nos choques da série original.

A metodologia foi também usada por Cavaliere, Rahbek e Taylor (2009) em que as inovações apresentavam volatilidade não estacionária, ou seja, presença de heterocedasticidade incondicional, em que a variância do erro varia no tempo de maneira não sistemática e em função das variáveis exógenas (não das variáveis resposta defasadas no tempo).

Os autores ressaltam que não é o objetivo do estudo estabelecer que o wild bootstrap oferece a melhor aproximação para metodologia convencional, sendo a idéia apenas apresentar um conjunto com menos restrições do que da literatura e que possa ser aplicada de forma a considerar as propriedades das séries financeiras e macroeconomicas atuais.

Na primeira parte do artigo de Cavaliere (2009) apresenta-se o o Modelo VAR e logo em seguida discorre-se sobre o *Pseudo LR Tests*. Basicamente, deseja-se avaliar simultaneamente a existência de cointegração e a ordem de integração das variáveis de um sistema.

Utiliza-se o termo “pseudo” pois o teste de máxima verossimilhança puro proposto por Johansen supõe normalidade, ausência de correlação serial e homocedasticidade sendo que o proposto (Pseudo) supõe apenas ausência de correlação serial. O teste exposto segue a mesma lógica de Johansen (1988) que já apresentado anteriormente na Seção 4.

Na segunda parte é apresentado o modelo de Swensen (2006). O autor defende o uso do bootstrap para identificação de cointegração ao invés da inferência assintótica clássica. A metodologia de bootstrap foi introduzida em 1979 por Efron, como uma técnica não paramétrica que procura substituir complicadas análises estatísticas teóricas por métodos de computação intensiva. Sua aplicação é utilizada para estimar características de interesse como variância, quantis e até a distribuição de amostragem do estimador.

O nome da técnica estatística veio da expressão “*to pull one self up by his own bootstraps*”, que induz a percepção de que nada mais é preciso além dos próprios dados. Basicamente, são feitas replicações contínuas em que novas amostras são criadas baseadas na original. Essa nova amostra, por sua vez, tem propriedades estatísticas que são usadas para inferência da amostra original.

Dado que as séries de tempo tem uma lógica na ordem, a utilização da técnica de bootstrap tem seu poder reduzido, já que a estrutura seria perdida com a criação de novas séries através da amostragem com reposição. Para séries temporais estacionárias, foi desenvolvida algumas metodologias como o “*block bootstrap*”, desenvolvido inicialmente por Carlstein (1986) e Kunsch (1989) e cuja idéia principal é dividir a série original em blocos consecutivos e criar novas amostras através da escolha dos blocos ao invés de observações individuais.

Entretanto, esse método não pode ser aplicado para séries não estacionárias dado que não é possível capturar a tendência estocástica com a quebra em blocos. Além disso, as séries não estacionárias são processos não invertíveis, impossibilitando a aplicação da técnica.

Por fim, o problema mais sério de se usar o bootstrap refere-se a premissa adotada de que os erros das séries originais são homocedásticos, o que pode não ser verdadeiro.

Em resumo, a técnica de bootstrap ao replicar a amostra original de uma série temporal (no nosso caso os preços das ações) para inferência sobre cointegração não é indicada pois: (i) não é capaz de identificar tendências estocásticas e (ii) assume que os erros são homocedásticos.

A metodologia proposta por Cavaliere (2009) é apresentada partindo de um teste em que hipótese nula  $H(r)$  testa um VAR(k) restrito contra a hipótese alternativa  $H(p)$  que testa VAR(k) irrestrito, sendo que  $r < p$  (numero de variáveis de  $H(r)$  é menor do que o numero de variáveis de  $H(p)$ ).

A idéia nada mais é que uma extensão da abordagem de Swensen (2006), em que se modifica a forma da amostragem com o objetivo de replicar a presença de heterocedasticidade condicional dos erros.

A estimativa dos parâmetros do VAR segue a mesma metodologia de Johansen tradicional:

**Passo 1:** Gerar T residuos  $\varepsilon_t^b$  sendo que  $t=1, \dots, T$  de acordo com:

$$\varepsilon_t^b = \hat{\varepsilon}_t w_t$$

Em que  $w_t$  é um escalar independente  $N(0, 1)$

O objetivo desse passo é criar artificialmente uma série em que amplifique a heterocedasticidade dos erros através do choque de um escalar no erro, mas sem alterar a estrutura da volatilidade.

**Passo 2:** Geração de uma nova série artificial através do bootstrap convencional de forma recursiva sob a lógica apresentada nessa mesma Seção (sorteio aleatório com reposição da amostra). Desse passo, obtem-se:

$$\Delta \tilde{X}_t = \alpha \beta' \tilde{X}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \tilde{X}_{t-i} + \varepsilon_t$$

**Passo 3:** Utilizando a amostra gerada por bootstrap no passo 2, calcula-se a estatística do teste através do máximo autovalor que me indicará a rejeição ou aceitação da hipótese nula de que há  $r$  vetores de cointegração distintos. Lembrando que nesse passo, também poderia ser utilizado a estatística do traço.

**Passo 4:** Ordenação dos p-valores do bootstrap obtidos no passo 3 e criação de um histograma (*Cumulative distribution function*) mais verossímil e que descreverá a probabilidade de rejeição da hipótese nula através da área debaixo do histograma inferior a 5%.

O ponto chave da metodologia proposta é o Passo 1 em que multiplicamos o choque por uma sequência de escalares normais com média 0 e variância 1. Isso permite que amplifiquemos a heterocedasticidade presente nos choques originais sem alterar a estrutura (os “novos erros” também tem média 0 e variância  $\Omega_t \omega_t^2$ ). Verifica-se que a inclusão de variáveis determinísticas não tem propriedades distintas para amostras finitas.

As 3 metodologias (assintótica, bootstrap e wild bootstrap) foram aplicadas para diversos países com o objetivo de verificar a validade da teoria, ou seja, a hipótese de que a taxa de juros de longo prazo é uma média da taxa de juros atual e a expectativa futura de curto prazo durante o período de investimentos. Assim, se a constatação for verdadeira, as taxas de juros seriam bem descritas por um processo I(1) já que as taxas de diferentes maturidades são guiadas por um componente comum (taxa corrente) com spreads das taxas das diferentes maturidades sendo estacionárias. Os testes foram implementados sobre as taxas correntes e as taxas de curto prazo de cada país e evidenciou que o wild bootstrap performa melhor em amostras finitas. Além disso, mostrou-se mais robusta quando na presença de heterocedasticidade condicional dos erros nas séries.

Os testes Johansen (1988) e Cavaliere (2009) que foram aplicados para estrutura da taxa de juros mostraram resultados distintos entre países com alta evidência da existência de cointegração entre diferentes vértices futuros.

## **6. Aplicação do trabalho de Cavaliere (2009) para o mercado brasileiro**

A metodologia de Cavaliere aplicada para a estrutura da taxa de juros reais para diversos países foi replicada para o mercado acionário brasileiro. Assim, através das mesmas ferramentas utilizadas pelo *paper* buscou-se verificar se efetivamente há um componente comum entre as ações que gere preços cointegrados.

Os testes para o mercado brasileiro foram aplicados sobre preços de ações listadas na Bovespa já tratadas (considerando dividendos, proventos e outros). Nenhum filtro relacionado ao porte, característica da gestão, segmento de atuação, controle acionário, alavancagem ou variáveis específicas da empresa foram filtrados<sup>5</sup>.

Ao escolher um critério de liquidez, alguns gestores optam por utilizar o valor de mercado da empresa, o volume de negociação médio diário do papel ou ambos. Ainda há aqueles que criam variáveis compostas de combinações de critérios para o mesmo tema (presença na bolsa, frequência nos pregões no ano...). Optou-se no presente trabalho aplicar um filtro de liquidez baseado no volume de negociação médio diário superior a R\$ 5 milhões por ser um filtro menos restritivo e de ampla utilização por participantes do mercado.

Além disso, excluiu-se ADRs (American Depositary Receipt), que são ativos de empresas brasileiras negociadas no mercado norte-americano dado a complexidade que envolve os diferentes níveis (I, II ou III) e por serem negociados em um mercado com liquidez e outras características muito diferentes do mercado brasileiro. As rotinas foram implementadas em Matlab. O resumo das séries filtradas encontra-se abaixo:

- Apenas ações do mercado brasileiro
- Excluiu-se ADR
- Ações que tenham histórico desde 08/03/2006 e que estejam ativas em 16/03/2011
- Volume de negociação médio diário desde janeiro de 2009 superior a R\$ 5 milhões

O filtro aplicado com as características descritas retornou 65 ações (que se encontram no Anexo 1) que combinadas duas a duas resulta em 2080 pares. Vale ressaltar que as 65 ações não referem-se a 65 empresas diferentes já que a base de dados após filtro de liquidez pode retornar ações de diferentes classes (preferenciais e ordinárias) de uma mesma companhia. Além disso, a base de dados pode apresentar sobreposições como no caso de Itaú Unibanco e Itaúsa (subsidiária e holding) o que em termos práticos significa ampliar a análise de cointegração para diferentes estratégias de pair trading (intrasetorial, intersetorial, holding x subsidiária entre outros).

O bootstrap foi implementado com 1000 replicações e para os dados com frequência diária, semanal e mensal. Aplicou-se a metodologia de Johansen (1988) e Cavaliere (2009).

<sup>5</sup> Optou-se por não aplicar outros filtros sobre as séries de preços de ações listadas na Bovespa dado que os 4 gestores de investimentos no Brasil, consultados para essa utilizam apenas filtros relacionados a liquidez.

A aplicação da metodologia de Cavaliere (2009) busca verificar se há diferenças nos testes que apontam cointegração entre as séries já que Johansen (1988) pressupõe que os erros das séries de preços são homocedásticas.

## **7. Resultados**

Para o caso de Johansen comparou-se as estatísticas do traço extraídos do MatLab com os valores críticos para diferentes níveis de confiança e, para o caso do Cavaliere, comparou-se diretamente os p-valores resultantes da rotina com os diferentes níveis de significância. De outra maneira, o teste foi feito da seguinte maneira:

### **Johansen:**

Há cointegração:

- Se as duas estatísticas do traço forem maiores que o Valor Crítico,  $r=2$  (hipótese nula rejeitada a favor de 2 vetores)
- Se estatística do traço 1  $>$  Valor Crítico e estatística do traço 2  $<$  Valor Crítico,  $r=1$  (hipótese nula rejeitada a favor de 1 vetor)

Não há Cointegração:

- Se as duas estatísticas do traço forem menores que o Valor Crítico,  $r=0$

### **Cavaliere:**

Há cointegração:

- Se rejeito a nula em favor de 1 vetor e, portanto,  $(1-pvalor1) > 95\%$  e  $(1-pvalor2) < 95\%$
- Se rejeito a nula em favor de 2 vetores e, portanto,  $(1-pvalor1) > 95\%$  e  $(1-pvalor2) > 95\%$

Não há cointegração:

- Se aceito a nula e, portanto,  $(1-pvalor1) < 95\%$  e  $(1-pvalor2) < 95\%$

Os resultados para os 2080 pares encontram-se na tabela abaixo, que nada mais é que a somatória dos resultados cujo teste apontaram cointegração rejeitando a nula ou a favor de 1 ou a favor de 2 vetores. Os níveis de significância adotados foram de 1%, 2,5% e 5%.

**Tabela 1**

Percentual dos pares que não cointegram, considerando nível de significância de 1%, frequência diária, semanal e mensal, a partir de 08/06/2006 até 16/03/2011.

| Significância = 1% | # Replicações | Johansen | Cavaliere |
|--------------------|---------------|----------|-----------|
| Diários            | 1000          | 96,30%   | 95,05%    |
| Semanais           | 1000          | 97,90%   | 96,60%    |
| Mensais            | 1000          | 97,90%   | 97,10%    |

Fonte: Elaborada pela autora

**Tabela 2**

Percentual dos pares que não cointegram, considerando nível de significância de 2,5%, frequência diária, semanal e mensal, a partir de 08/06/2006 até 16/03/2011.

| Significância = 2,5% | # Replicações | Johansen | Cavaliere |
|----------------------|---------------|----------|-----------|
| Diários              | 1000          | 91,97%   | 88,13%    |
| Semanais             | 1000          | 94,40%   | 91,30%    |
| Mensais              | 1000          | 94,40%   | 91,80%    |

Fonte: Elaborada pela autora

**Tabela 3**

Percentual dos pares que não cointegram, considerando nível de significância de 5%, frequência diária, semanal e mensal, a partir de 08/06/2006 até 16/03/2011.

| Significância = 5% | # Replicações | Johansen | Cavaliere |
|--------------------|---------------|----------|-----------|
| Diários            | 1000          | 91,68%   | 83,85%    |
| Semanais           | 1000          | 94,20%   | 87,40%    |



|         |      |        |        |
|---------|------|--------|--------|
| Mensais | 1000 | 94,20% | 87,60% |
|---------|------|--------|--------|

Fonte: Elaborada pela autora

Observa-se que utilizando frequência diária, semanal e mensal, há um alto nível de aceitação da hipótese nula para os diferentes níveis de confiança adotado (1%, 2,5% e 5%). Todos apresentaram resultados de não cointegração com proporção superior a 80% e em muitos casos, próximo ao tamanho do teste.

Adicionalmente, verifica-se que a utilização de dados semanais, aumenta a aceitação da hipótese de não cointegração quando comparados a utilização de dados diários, que pode estar ligado à diminuição dos ruídos nas séries de maior frequência.

## 8. Avaliação dos pares cointegrados

Pelo resultado do teste feito pela metodologia de Cavaliere da Tabela 3, para os dados diários, com 5% de significância e 1000 replicações, verifica-se que dos 2080 pares avaliados, 1744 (83,85%) pares não apresentaram evidência de cointegração contra 336 (16,15%) que apresentaram. Desses que apontaram para cointegração, 107 rejeitaram a hipótese nula a favor de 1 vetor e 229 rejeitaram a favor de 2 vetores.

Para verificar se esses pares que indicam cointegração fazem parte do mesmo setor, utilizou-se o critério de classificação setorial do Economática descrito com mais detalhes no Anexo 2, o qual agrupa o total de ações em 20 setores diferentes. A suspeita é de que desse total de pares cointegrados, a maior parte seja de empresas do mesmo setor, dado que espera-se que os choques exógenos tenham impactos semelhantes para duas empresas que possuem fatores de risco em comum.

O número de ações para cada setor, utilizando a classificação setorial do Economática para as 65 ações avaliadas, encontra-se abaixo. A segunda coluna da Tabela 4 indica o número de pares possíveis dentro do setor respectivo.

**Tabela 4**

Número de empresas que pertencem ao setor discriminado segundo classificação do Economática com indicação do número de pares possíveis de serem combinados dentro do mesmo setor.

| Setor                | # Empresas | Combinações |
|----------------------|------------|-------------|
| Alimentos e Beb      | 3          | 3           |
| Comércio             | 5          | 10          |
| Construção           | 3          | 3           |
| Energia Elétrica     | 10         | 45          |
| Finanças e Seguros   | 5          | 10          |
| Máquinas Industriais | 1          | -           |
| Mineração            | 2          | 1           |
| Outros               | 7          | 21          |
| Papel e Celulose     | 3          | 3           |
| Petróleo e Gas       | 2          | 1           |
| Química              | 3          | 3           |
| Siderur & Metalur    | 6          | 15          |
| Software e Dados     | 1          | -           |
| Telecomunicações     | 9          | 36          |
| Transporte Serviç    | 4          | 6           |
| Veiculos e peças     | 1          | -           |
| <b>TOTAL</b>         | <b>65</b>  | <b>2080</b> |

Fonte: Elaborada pela autora

Observa-se pela Tabela 4 que os setores de Energia Elétrica e Telecomunicações são os que apresentaram maior número empresas e consequentemente maior número de pares possíveis. Com a classificação, criou-se uma tabela para verificar se as empresas do mesmo setor apresentam de fato maior número de pares cointegrados quando comparados com pares de setores distintos.

**Tabela 5**

## Número de pares cointegrados agrupados setorialmente e em termos absolutos

| Setor                | Alimentos e Beb | Comércio  | Construção | Energia Elétrica | Finanças e Seguros | Máquinas Industriais | Mineração | Outros    | Papel e Celulose | Petróleo e Gas | Química   | Siderur & Metalur | Software e Dados | Telecomu nicações | Transporte Serviç | Veiculos e peças | Total      |
|----------------------|-----------------|-----------|------------|------------------|--------------------|----------------------|-----------|-----------|------------------|----------------|-----------|-------------------|------------------|-------------------|-------------------|------------------|------------|
| Alimentos e Beb      | 1               | 0         | 0          | 8                | 2                  | 0                    | 0         | 2         | 0                | 0              | 0         | 0                 | 1                | 4                 | 3                 | 0                | 21         |
| Comércio             | 0               | 1         | 1          | 0                | 1                  | 1                    | 0         | 3         | 1                | 0              | 0         | 0                 | 1                | 1                 | 0                 | 0                | 9          |
| Construção           | 0               | 1         | 2          | 0                | 0                  | 2                    | 0         | 3         | 2                | 0              | 0         | 3                 | 0                | 4                 | 0                 | 0                | 17         |
| Energia Elétrica     | 0               | 3         | 0          | 25               | 6                  | 0                    | 2         | 10        | 0                | 0              | 6         | 1                 | 5                | 19                | 0                 | 0                | 77         |
| Finanças e Seguros   | 0               | 3         | 1          | 3                | 5                  | 4                    | 4         | 9         | 1                | 0              | 1         | 1                 | 2                | 7                 | 1                 | 0                | 42         |
| Máquinas Industriais | 0               | 0         | 0          | 0                | 0                  | 0                    | 0         | 0         | 0                | 0              | 0         | 0                 | 0                | 0                 | 0                 | 0                | 0          |
| Mineração            | 0               | 0         | 0          | 0                | 0                  | 2                    | 1         | 0         | 0                | 0              | 1         | 0                 | 0                | 1                 | 0                 | 0                | 5          |
| Outros               | 0               | 2         | 1          | 5                | 4                  | 4                    | 5         | 4         | 2                | 2              | 2         | 2                 | 1                | 11                | 0                 | 0                | 45         |
| Papel e Celulose     | 0               | 1         | 3          | 0                | 0                  | 1                    | 0         | 3         | 0                | 0              | 0         | 2                 | 0                | 5                 | 0                 | 0                | 15         |
| Petróleo e Gas       | 0               | 0         | 0          | 0                | 0                  | 0                    | 0         | 0         | 0                | 0              | 0         | 0                 | 0                | 9                 | 0                 | 0                | 9          |
| Química              | 0               | 0         | 0          | 0                | 0                  | 1                    | 0         | 2         | 1                | 0              | 0         | 0                 | 0                | 2                 | 0                 | 0                | 6          |
| Siderur & Metalur    | 0               | 2         | 0          | 0                | 0                  | 0                    | 3         | 4         | 1                | 0              | 0         | 9                 | 0                | 10                | 0                 | 0                | 29         |
| Software e Dados     | 0               | 0         | 0          | 0                | 0                  | 0                    | 0         | 0         | 0                | 0              | 0         | 0                 | 0                | 1                 | 0                 | 0                | 1          |
| Telecomunicações     | 0               | 0         | 0          | 2                | 0                  | 5                    | 5         | 0         | 0                | 2              | 7         | 5                 | 0                | 13                | 0                 | 0                | 39         |
| Transporte Serviço   | 0               | 0         | 0          | 5                | 1                  | 0                    | 0         | 2         | 0                | 0              | 1         | 0                 | 1                | 8                 | 1                 | 0                | 19         |
| Veículos e peças     | 0               | 0         | 0          | 0                | 0                  | 0                    | 0         | 0         | 0                | 0              | 0         | 0                 | 0                | 2                 | 0                 | 0                | 2          |
| <b>Total</b>         | <b>1</b>        | <b>12</b> | <b>8</b>   | <b>48</b>        | <b>19</b>          | <b>20</b>            | <b>20</b> | <b>42</b> | <b>8</b>         | <b>4</b>       | <b>18</b> | <b>23</b>         | <b>11</b>        | <b>97</b>         | <b>5</b>          | <b>0</b>         | <b>336</b> |

Fonte: Elaborada pela autora

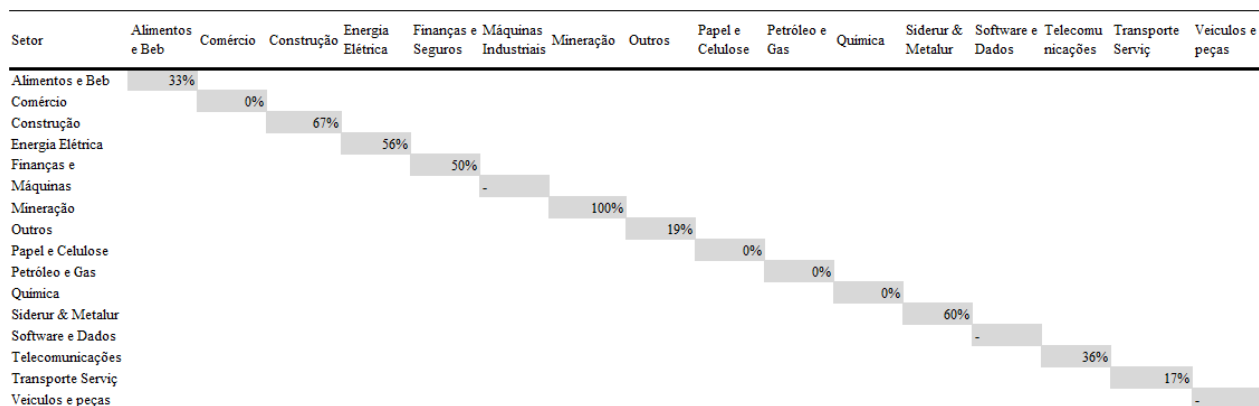
A tabela 5 mostra que dos 336 pares cointegrados, apenas 61 (soma da diagonal principal da Tabela 5) são pares cujas empresas pertencem ao mesmo setor (18,15% dos pares cointegrados), sendo que a soma dos 3 maiores setores representam mais de 75% desse total (Energia Elétrica, Telecomunicações e Siderurgia e Metalurgia).

Os valores apresentados acima em termos absolutos não permite comparar os setores diretamente dado que o número de empresas difere para cada setor e, consequentemente, o número de combinações possíveis é distinto.

Assim, avaliando a Tabela 5 de forma relativa, ou seja, comparando o número de pares cointegrados em relação ao total de combinações possíveis intrasetorialmente, observa-se que o setor de Mineração e Construção são os que apresentam maior número de pares cointegrados.

**Tabela 6**

Percentual de pares cointegrados dentro do mesmo setor



Fonte: Elaborada pela autora

O setor de mineração resultou em 100% de cointegração intrasetorialmente por se tratar de apenas duas ações referentes à mesma empresa e que se diferem pela classe (ordinária e preferencial). Era de se esperar que os choques que afetam a empresa Vale do Rio Doce tenham efeitos semelhantes para a VALE3 (ordinária) e a VALE5 (preferencial).

Já o setor de Construção, que apresentou 67% de cointegração intrasetorialmente, refere-se a 3 empresas distintas – Gafisa, Cyrela e Rossi Residencial. Apesar de atuarem majoritariamente no setor residencial, cada uma das empresas possuem características diferentes em relação a governança corporativa e ao público alvo. A Gafisa, por exemplo, tem 100% de *free float* (percentual do capital social que se encontra disperso em bolsa) enquanto que a Rossi e a Cyrela têm por volta de 65%. Apesar das diferenças, possivelmente os fatores setoriais são predominantes para determinar a direção dos preços das empresas que são altamente dependentes do crédito para financiamento imobiliário e da atividade econômica.

Para o setor de Siderurgia e Metalurgia, trata-se de 6 ações distintas e referentes a 3 conglomerados diferentes (Gerdau, Usiminas e CSN). A empresa Gerdau possui 3 ações listadas na Bovespa, sendo GGBR3 e GGBR4 referentes à Gerdau S/A e a GOAU4 à divisão da Metarlúrgica. Mesmo entre as ações pertencentes a diferentes conglomerados, observa-se cointegração, como é o caso da GGBR3/GGBR4 com USIM5, as quais possuem o preço do aço como um dos principais componentes dos custos da empresa e consequentemente para o lucro, o que se reflete diretamente nos preços das respectivas ações.

Assim como no caso da Vale do Rio Doce que apresentaram cointegração no par das ações de diferentes classes, dos 336 pares avaliados, 9 referem-se a pares em que as duas ações

pertencem a um mesmo conglomerado. Desses 9 pares, 6 rejeitaram a hipótese nula de não cointegração (Ambev, Eletrobrás, Gerdau, Telemar, Tim Participações e Vale do Rio Doce) e 3 aceitaram a hipótese nula (Bradesco, Petrobrás e Usiminas), excluindo a hipótese de cointegração.

Uma hipótese para explicar os resultados de não cointegração entre ações de um mesmo conglomerado refere-se a alta concentração dos detentores das ações ordinárias (no caso da Petrobrás, a concentração das ações ordinárias pelo Governo).

## 9. Alteração da amostra – testes de robustez

A base de dados inicial inclui o período da crise financeira, que representou uma forte queda nas ações nos últimos 50 anos, que ocorreu no segundo semestre de 2008, momento em que diversos ativos tiveram seu valor de mercado divididos pela metade.

Dessa forma, para eliminar a suspeita de que os testes estão indicando baixa cointegração das séries com alto grau de confiança, em função do período da crise na base de dados, utilizar-se-a uma amostra mais longa (de março 2001 até março 2011), com os mesmos filtros da amostra anterior.

A amostra final considerando os criterios de liquidez adotado anteriormente, possui 40 ações (Anexo 3), o que resulta em 780 pares diferentes. Os resultados encontram-se abaixo:

**Tabela 7**

Percentual dos pares que não cointegram, considerando nível de significância de 1%, frequencia diária, semanal e mensal, a partir de 01/03/2001 até 16/03/2011.

| Significância = 1% | # Replicações | Johansen | Cavaliere |
|--------------------|---------------|----------|-----------|
| Diários            | 1000          | 87,10%   | 78,70%    |
| Semanais           | 1000          | 92,10%   | 84,50%    |

Fonte: Elaborada pela autora

Observa-se que o teste considerando dados semanais e sobre a base mais longa (10 anos), assim como na amostra anterior, teve uma aceitação maior da hipótese de não cointegração do que a amostra com dados diários.

Verifica-se também que mesmo considerando um período longo de 10 anos, em que o efeito da crise não estaria tão potencializado quanto na base anterior, persiste a alta aceitação da nula de não cointegração. Vale ressaltar que a rejeição de não cointegração diminuiu comparado com os testes feitos com o período mais curto (de março 2006 até março 2011), o que é o mesmo que dizer que o teste apontou para um percentual maior de pares que cointegram entre si.

Isso levanta a hipótese de que a alta aceitação da nula nos testes feito anteriormente esteja relacionado a um período de turbulência em que os retornos dos ativos comportaram-se de forma atípica apontando uma sobre-aceitação que não pode ser levada em consideração para inferência de não cointegração.

Dessa forma, segregou-se a amostra de 10 anos em duas partes, de forma que a primeira compreenderá o período de março 2001 até dezembro 2007 e a segunda de janeiro 2008 até março 2011. A divisão busca isolar o efeito da Crise de 2008. Os resultados encontram-se abaixo:

**Tabela 8**

Percentual dos pares que não cointegram, considerando nível de significância de 1%, frequência diária e semanal, a partir de 01/03/2001 até 31/12/2007.

| Significância = 1% | # Replicações | Johansen | Cavaliere |
|--------------------|---------------|----------|-----------|
| Diários            | 1000          | 91,80%   | 87,90%    |
| Semanais           | 1000          | 94,00%   | 89,40%    |

Fonte: Elaborada pela autora

**Tabela 9**

Percentual dos pares que não cointegram, considerando nível de significância de 1%, frequência diária e semanal, a partir de 01/01/2008 até 16/03/2011.

| Significância = 1% | # Replicações | Johansen | Cavaliere |
|--------------------|---------------|----------|-----------|
| Diários            | 1000          | 91,40%   | 87,90%    |
| Semanais           | 1000          | 94,20%   | 91,80%    |

Fonte: Elaborada pela autora

Os resultados mostram que a hipótese nula de não cointegração continua a ser aceita com alta probabilidade, mesmo isolando o efeito da Crise.

## 10. Explicação dos resultados

Após avaliação dos preços de ações brasileiras em diferentes períodos e diferentes frequências para diferentes amostras, conclui-se que há poucas evidências sobre o alto número de pares cointegrados.

Uma possível explicação para isso foi citada no paper de Richards (1996) em que cita o caso de empresas da mesma indústria e atuantes no mesmo país, cujos retornos possui componente comum que reflete variáveis macroeconômicas e setoriais. Entretanto, diz ser razoável pensar que cada uma dessas ações contem um componente próprio que reflete a forma de gestão das empresas e assim, a cointegração só seria verificada se os choques sobre esse componente tivessem o mesmo impacto em termos de direção para todas elas – premissa improvável de ocorrer.

O paper escrito Lin, McCrae e Gulati (2006) usa os princípios da teoria de cointegração para desenvolver um procedimento que contem uma condição de lucro mínimo através da estratégia de pair trading. Inicialmente, verificam as condições necessárias para implementar o procedimento e então definem um processo de 5 passos para identificação dos ativos elegíveis para as operações. A validade estatística é então verificada através de uma simulação histórica.

A abordagem se inicia com a identificação de pares potenciais sob a óptica da teoria de cointegração, apontando que as ineficiências da estratégia ocorrem pois: (i) há eventos inesperados no mercado, (ii) há ineficiência persistente na precificação dos ativos e (iii) alterações estruturais nos preços.

Os diversos autores que tentam explicar a não cointegração entre ações para diferentes mercados se dividem basicamente em dois grupos. O primeiro, defende que o retorno das ações de fato cointegram, em função dos diversos fatores comuns ligados que influenciam a evolução dos preços, mas que eventos exógenos – legislação, movimentos especulativos, eventos corporativos entre outros – distorcem esse equilíbrio de Longo Prazo.

O segundo grupo, atribui a alta aceitação da nula pelas especificidades de cada uma das empresas no que se diz respeito principalmente a gestão corporativa, variável que difere muito entre elas e que se sobrepõe aos seus fatores comuns.

Por fim, verificou-se que dentre os pares que se mostraram cointegrados, menos de 20% são ações do mesmo setor. Adicionalmente, observou-se 3 casos de uma mesma empresa em que ações de classes distintas não apresentaram evidências de cointegração, resultados que aumentam ainda mais a complexidade da seleção do universo para aplicação da técnica de arbitragem estatística.

## **11. Possíveis Extensões**

Engle e Smith (1999), tenta preencher a lacuna entre processos com choques permanentes e choques transitórios, formulando o processo STOPBREAK. O processo é basicamente caracterizado por ter alterações estruturais aleatórias e em intervalos também aleatórios. Em outras palavras, o processo está relacionado a variação do impacto dos choques permanentes ao longo do tempo, que está intimamente ligado ao conceito de quebras estruturais.

No paper, a aderência do processo STOPBREAK é avaliado para o comportamento relativo entre pares de ações, sob a premissa de que os preços entre os pares podem se mover juntos temporariamente, ou seja, pode haver uma cointegração temporária entre o preço relativo de ações. Os resultados para os dados diários de ações americanas de 1988 até 1995 mostraram que há fortes evidências de que os spreads são bem modelados pelo processo STOPBREAK.

Uma possível extensão para estudo seria avaliar a aderência desse processo para as ações no mercado brasileiro mesmo que temporariamente, possibilitando assim que a estratégia de pair trading seja aplicada com maior eficiência.

Granger (1986), por sua vez, inicia o artigo mencionando a crença de que pares de variáveis econômicas devem ser convergentes pelo menos no longo prazo, mesmo com possíveis



desvios no curto prazo. Adiciona ainda que se esses desvios de curto prazo perdurarem por determinado tempo em função de fatores sazonais, forças econômicas como mecanismos do mercado ou intervenção do governo farão com que haja convergência novamente (cointegração).

Entretanto, ao inferir a teoria para pares de preços em um mercado eficiente, descarta a possibilidade de cointegração entre eles já que a existência de cointegração implica presença de Causalidade de Granger em pelo menos uma direção. Ou seja, se preços forem cointegrados, haveria uma previsibilidade dos preços futuros e, portanto, seria contraditório a premissa de mercados eficientes.

Os resultados do trabalho mostraram alta aceitação da hipótese de não cointegração, o que pode estar relacionado a aplicação da técnica de pair trading em mercados eficientes em que os ativos não cointegram entre si.

Uma possível extensão para o trabalho seria testar a Hipótese de Eficiência de Mercado para o mercado acionário brasileiro, amplamente discutido na literatura por Fama (1969, 1970), Haugen (2001) e Minardi (2004) e simular operações para verificar se de fatos são fontes de valor em mercados eficientes e não eficientes.

## **12. Conclusão**

Nessa dissertação procurou-se avaliar a existência de cointegração dos preços das ações do mercado brasileiro. Para isso, aplicou-se os testes propostos por Johansen (1986) e Cavaliere (2009), sendo esse último uma forma de tornar o teste robusto a eventual heterocedasticidade nos choques do modelo.

A amostra inicial (de 2006 a 2011) mostrou que há baixa evidência de alto nível de cointegração entre o preços das 65 ações levantadas que no período apresentaram média de negociação diária superior a R\$5 milhões (apenas 16,15%). Dentre esses pares, cujo resultado apontou para cointegração, foi avaliado se a maioria tratava-se de pares de ações do mesmo setor. O resultado mostrou que apenas 61 pares (18,15% do total dos pares cointegrados) referem-se a pares intrasetoriais. Nessa amostra, avaliou-se também pares de um mesmo conglomerado e observou-se que 3 dos 9 pares apontaram para não cointegração.

O teste inicial levantou a hipótese de que o alto nível de aceitação da hipótese de não cointegração poderia estar relacionado ao período de crise em que os preços foram afetados

negativamente e atipicamente. Sendo assim, aplicou-se o teste para uma amostra mais longa e em outra que não contém o período da crise para tornar o teste robusto a amostra. Adicionalmente, o teste foi feito em base diária, semanal e mensal para que fosse robusto também à frequência.

Os resultados obtidos apontaram novamente para a baixa evidência de cointegração entre as séries de preços de ações brasileiras em todos os casos e com alto nível de confiança, o que ratifica a necessidade de cautela na forma de implantação da técnica na construção de carteiras.

### 13. Anexos

#### Anexo 1: Código das 65 Ações filtradas para o 1º teste

| Empresa |              |
|---------|--------------|
| 1       | GETI4        |
| 2       | ALLL11+ALLL3 |
| 3       | AMBV3        |
| 4       | AMBV4        |
| 5       | SUBA3+BTOW3  |
| 6       | BBDC3        |
| 7       | BBDC4        |
| 8       | BRAP4        |
| 9       | BBAS3        |
| 10      | BRTO4        |
| 11      | BRKM5        |
| 12      | CCRO3        |
| 13      | CMIG4        |
| 14      | CPLE6        |
| 15      | CSAN3        |
| 16      | CPFE3        |
| 17      | CYRE3        |
| 18      | DASA3        |
| 19      | ELET3        |
| 20      | ELET6        |
| 21      | EMBR3        |
| 22      | ENBR3        |
| 23      | VCPA4+FIBR3  |
| 24      | GFSA3        |
| 25      | GGBR3        |
| 26      | GGBR4        |
| 27      | GOAU4        |
| 28      | GOLL4        |

|    |       |
|----|-------|
| 29 | ITSA4 |
| 30 | ITUB4 |
| 31 | KLBN4 |
| 32 | LIGT3 |
| 33 | RENT3 |
| 34 | LAME4 |
| 35 | LREN3 |
| 36 | NATU3 |
| 37 | NETC4 |
| 38 | PCAR5 |
| 39 | PETR3 |
| 40 | PETR4 |
| 41 | PSSA3 |
| 42 | RSID3 |
| 43 | SBSP3 |
| 44 | CSNA3 |
| 45 | CRUZ3 |
| 46 | SUZB5 |
| 47 | TAMM4 |
| 48 | TELB4 |
| 49 | TNLP3 |
| 50 | TNLP4 |
| 51 | TMAR5 |
| 52 | TLPP4 |
| 53 | TCSL3 |
| 54 | TCSL4 |
| 55 | TOTS3 |
| 56 | TBLE3 |
| 57 | TRPL4 |
| 58 | UGPA4 |
| 59 | USIM3 |
| 60 | USIM5 |
| 61 | VALE3 |

|    |             |
|----|-------------|
| 62 | VALE5       |
| 63 | FFTL4       |
| 64 | VIVO4       |
| 65 | WEGE4+WEGE3 |

## Anexo 2: Classificação setorial do Economática

O Economática classifica as ações listadas em 20 setores distintos e discriminados abaixo. O número na segunda coluna refere-se a quantidade de ações classificadas na categoria das 640 listadas e ativas.

| Setor Economática  | Total      |
|--------------------|------------|
| Agro e Pesca       | 6          |
| Alimentos e Beb    | 33         |
| Comércio           | 28         |
| Construção         | 36         |
| Eletroeletrônicos  | 11         |
| Energia Elétrica   | 85         |
| Finanças e Seguros | 61         |
| Fundos             | 1          |
| Máquinas Indust    | 6          |
| Mineração          | 9          |
| Minerais não Met   | 7          |
| Papel e Celulose   | 11         |
| Petróleo e Gas     | 10         |
| Química            | 23         |
| Siderur & Metalur  | 47         |
| Software e Dados   | 5          |
| Telecomunicações   | 20         |
| Textil             | 54         |
| Transporte Serviç  | 22         |
| Veiculos e peças   | 33         |
| Outros             | 132        |
| <b>Grand Total</b> | <b>640</b> |

## Empresas classificadas como Outros pelo Economática

| Ação          | Classe | Classificação Economática |
|---------------|--------|---------------------------|
| Abril Educa   | UNT N2 | Outros                    |
| Alfa Consorc  | ON     | Outros                    |
| Alfa Consorc  | PNA    | Outros                    |
| Alfa Consorc  | PNB    | Outros                    |
| Alfa Consorc  | PNC    | Outros                    |
| Alfa Consorc  | PND    | Outros                    |
| Alfa Consorc  | PNE    | Outros                    |
| Alfa Consorc  | PNF    | Outros                    |
| Alfa Holding  | ON     | Outros                    |
| Alfa Holding  | PNA    | Outros                    |
| Alfa Holding  | PNB    | Outros                    |
| Aliansce      | ON     | Outros                    |
| Amil          | ON     | Outros                    |
| Anhanguera    | ON     | Outros                    |
| Anhanguera    | ON     | Outros                    |
| Bahema        | ON     | Outros                    |
| Baumer        | ON     | Outros                    |
| Baumer        | PN     | Outros                    |
| BHG           | ON     | Outros                    |
| Biommm        | ON     | Outros                    |
| Biommm        | PN     | Outros                    |
| BR Brokers    | ON     | Outros                    |
| BR Malls Par  | ON     | Outros                    |
| BR Properties | ON     | Outros                    |
| Bradespar     | ON     | Outros                    |
| Bradespar     | PN     | Outros                    |
| Brazilian Fr  | ON     | Outros                    |
| Casan         | ON     | Outros                    |
| Casan         | PN     | Outros                    |

|                |     |        |
|----------------|-----|--------|
| Cemepe         | ON  | Outros |
| Cemepe         | PN  | Outros |
| Cims           | ON  | Outros |
| Cims           | PN  | Outros |
| Coari Part     | ON  | Outros |
| Coari Part     | PN  | Outros |
| Contax         | ON  | Outros |
| Contax         | PN  | Outros |
| Copasa         | ON  | Outros |
| Cor Ribeiro    | ON  | Outros |
| Cor Ribeiro    | PN  | Outros |
| Csu Cardsystem | ON  | Outros |
| Cyre Com-Ccp   | ON  | Outros |
| Dasa           | ON  | Outros |
| Docas          | ON  | Outros |
| Docas          | PN  | Outros |
| Dtcom Direct   | ON  | Outros |
| Dtcom Direct   | PN  | Outros |
| Duratex        | ON  | Outros |
| Ecodiesel      | ON  | Outros |
| Estacio Part   | ON  | Outros |
| Estrela        | ON  | Outros |
| Estrela        | PN  | Outros |
| Eucatex        | ON  | Outros |
| Eucatex        | PN  | Outros |
| Fleury         | ON  | Outros |
| Generalshopp   | ON  | Outros |
| Gp Invest      | A   | Outros |
| GPC Part       | ON  | Outros |
| Grucai         | ON  | Outros |
| Grucai         | PNB | Outros |
| Habitasul      | ON  | Outros |
| Habitasul      | PNA | Outros |

|              |        |        |
|--------------|--------|--------|
| Habitasul    | PNB    | Outros |
| Hoteis Othon | ON     | Outros |
| Hoteis Othon | PN     | Outros |
| Hypermarcas  | ON     | Outros |
| Ideiasnet    | ON     | Outros |
| Iguatemi     | ON     | Outros |
| Inepar       | ON     | Outros |
| Inepar       | PN     | Outros |
| Invest Bemge | ON     | Outros |
| Invest Bemge | PN     | Outros |
| Itaitinga    | ON     | Outros |
| Itaitinga    | PRB    | Outros |
| Itausa       | ON     | Outros |
| Itausa       | PN     | Outros |
| Jereissati   | ON     | Outros |
| Jereissati   | PN     | Outros |
| Kroton       | ON     | Outros |
| Kroton       | PN     | Outros |
| Kroton       | UNT N2 | Outros |
| Lark Maqs    | ON     | Outros |
| Lark Maqs    | PN     | Outros |
| Llx Log      | ON     | Outros |
| Localiza     | ON     | Outros |
| Lopes Brasil | ON     | Outros |
| Mills        | ON     | Outros |
| Mont Aranha  | ON     | Outros |
| Multiplan    | ON     | Outros |
| Multiplus    | ON     | Outros |
| Net          | ON     | Outros |
| Net          | PN     | Outros |
| Odontoprev   | ON     | Outros |
| Par Al Bahia | ON     | Outros |
| Par Al Bahia | PN     | Outros |



|              |     |        |
|--------------|-----|--------|
| Petropar     | ON  | Outros |
| Petropar     | PN  | Outros |
| Polpar       | ON  | Outros |
| Pq Hopi Hari | ON  | Outros |
| Pq Hopi Hari | PN  | Outros |
| Qualicorp    | ON  | Outros |
| Sabesp       | ON  | Outros |
| Sanepar      | ON  | Outros |
| Sanepar      | PN  | Outros |
| Sansuy       | ON  | Outros |
| Sansuy       | PNA | Outros |
| Sansuy       | PNB | Outros |
| Sao Carlos   | ON  | Outros |
| Saraiva Livr | ON  | Outros |
| Saraiva Livr | PN  | Outros |
| Sauipe       | ON  | Outros |
| Sauipe       | PN  | Outros |
| Sierrabrasil | ON  | Outros |
| Sondotecnica | ON  | Outros |
| Sondotecnica | PNA | Outros |
| Sondotecnica | PNB | Outros |
| Souza Cruz   | ON  | Outros |
| SPturis      | ON  | Outros |
| SPturis      | PNA | Outros |
| SPturis      | PNB | Outros |
| Suzano Hold  | ON  | Outros |
| Suzano Hold  | PNA | Outros |
| Technos      | ON  | Outros |
| Tecnosolo    | ON  | Outros |
| Tecnosolo    | PN  | Outros |
| Tectoy       | ON  | Outros |
| Tectoy       | PN  | Outros |
| Tempo Part   | ON  | Outros |

|              |    |        |
|--------------|----|--------|
| Time For Fun | ON | Outros |
| Trevisa      | ON | Outros |
| Trevisa      | PN | Outros |
| Valid        | ON | Outros |

### **Anexo 3: Código das 40 Ações filtradas para o teste de Robustez**

|    | Empresa     |
|----|-------------|
| 1  | AMBV4       |
| 2  | BBDC3       |
| 3  | BBDC4       |
| 4  | BRAP4       |
| 5  | BBAS3       |
| 6  | BRTO4       |
| 7  | BRKM5       |
| 8  | CMIG4       |
| 9  | CPLE6       |
| 10 | ELET3       |
| 11 | ELET6       |
| 12 | EMBR3       |
| 13 | EBTP4       |
| 14 | VCPA4+FIBR3 |
| 15 | GGBR4       |
| 16 | GOAU4       |
| 17 | ITSA4       |
| 18 | ITUB4       |
| 19 | KLBN4       |
| 20 | LIGT3       |
| 21 | LAME4       |
| 22 | NETC4       |
| 23 | PCAR4       |
| 24 | PETR3       |
| 25 | PETR4       |

|    |       |
|----|-------|
| 26 | SBSP3 |
| 27 | CSNA3 |
| 28 | CRUZ3 |
| 29 | SUZB5 |
| 30 | TELB4 |
| 31 | TNLP3 |
| 32 | TNLP4 |
| 33 | TCSL4 |
| 34 | TBLE3 |
| 35 | TRPL4 |
| 36 | UGPA4 |
| 37 | USIM5 |
| 38 | VALE3 |
| 39 | VALE5 |
| 40 | VIVO4 |

#### **14. Referências:**

**Booth, G. G. & Lee, T. H., & Tse, Y. (1996);** “International Linkages in the Nikkei Stock Index Futures Markets”; In: Pacific Basin Finance Journal, 4:59-76

**Carlstein, E. (1986);** “The use of subseries methods for estimating the variance of a general statistic from a stationary time series”; In: The Annals of Statistics

**Cavaliere, G., Rahbek, A. & Taylor, A.M.R. (2010);** “Co-integration rank testing under conditional heteroskedasticity”; In: Econometric Theory, 26, p. 1719-1780

**Cavaliere, G., Rahbek, A. & Taylor, A.M.R. (2008);** “Testing for Co-Integration in Vector Autoregressions with Non-Stationary Volatility”; In: CREATES Research Paper No. 2008-50

**Chiu, J. & Lukman, D. & Modarresi, K. & Velayutham, A. (2011);** “High-frequency Trading”; In: Stanford University Working Paper

**Dickey, D. & W. Fuller (1979);** “Distribution of the Estimators for Autoregressive Time Series with a Unit Root”; In: Journal of the American Statistical Association

**Dickey, D. A. & Pantula, S. (1987);** “Determining the Order of Differencing in Autoregressive Processes”; In: Journal of Business and Economic Statistics

**Durbin, J. & Watson, G.S. (1950);** “Testing for Serial Correlation in Least Squares Regression I”; In: Biometrika 37:409-428

**Efron, B. (1979);** “Bootstrap Methods: Another Look at the Jackknife”; In: The Annals of Statistics

**Elliott, G. & Rothenberg, T. J. & J. H. Stock (1996);** “Efficient Test for an Autoregressive Unit Root”; In: Econometrica, 64:813-836

**Enders, W. (1995);** “Applied Econometric Time Series”; Series. John Wiley and Sons

**Engle, R. F. & Smith, A. D. (1999);** “Stochastic Permanent Breaks”; In: The Review of Economics and Statistics

**Engle, R. F. (1982);** “Autoregressive Conditional Heteroskedasticity”; In: Journal of Econometrics 31:307-27

**Engle, R. F. & Granger, C. W. J. (1987);** “Co-integration and error-correction: Representation, estimation and testing”; In: Econometrica 56:119-139

**Goncalves, S. & Kilian, L. (2003);** “Asymptotic and Bootstrap Inference for AR (Infinity) Processes with Conditional Heteroskedasticity”; In: Scientific Series

**Granger, C.W.J. (1986);** “Development in the Study of Cointegrated Economic Variables”; In: Oxford Bulletin of Economics and Statistics

**Haza, D. P. & Fuller, W.A. (1979);** “Estimation for Autoregressive Processes with Unit Roots”; In: The Annals of Statistics

**Houthakker, H. (1961);** “An International Comparison of Personal Saving”; In: Bulletin of the International Statistical Institute

**Johansen, S. (1988);** “Statistical Analysis of cointegration vectors”; In: Journal of Economic Dynamics and Control

**Johansen, S. & Juselius, K. (1990);** “Some Structural Hypotheses in a Multivariate Cointegration Analysis of the Purchasing Power Parity and the Uncovered Interest Parity for UK”; In: Discussion Papers 90-05, University of Copenhagen - Department of Economics

**Johansen, S. (1991);** “Determination of Cointegration Rank in the Presence of a Linear Trend”; In: Papers 78, Helsinki – Department of Economics

**Johansen, S. (1994);** “The role of the constant and linear terms in cointegration analysis of nonstationary variables”; In: Econometric Reviews, 13:205-229

**Kendal, M. G. & Bradford, H. A. (1953);** “The Analysis of Economic Time-Series-Part I: Prices”; In: Journal of the Royal Statistical Society

**Cavaliere, G., Rahbek, A. & Taylor, A.M.R. (2010);** “Co-integration rank testing under conditional heteroskedasticity”; In: Econometric Theory, 26, p. 1719-1780

**Kwiatkowski, D. & Phillips, P. C. B. & Schmidt, P. & Shin, Y. (1992);** “Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root”; In: Journal of Econometrics 54:159-178

**Kunsch, H. R. (1989);** “The Jackknife and the Bootstrap for General Stationary Observations”; In: The Annals of Statistics

**Lin Y. X., McCrae M. & Gulati C. (2006),** “Loss Protection in Pairs Trading through minimum profit bounds: a cointegration approach”; In: Journal of Applied Mathematics and Decision Sciences

**Maddala, G. & Kim, I. (1998),** “Unit Roots, Cointegration and Structural Change”; In: Cambridge University

**Ng, S. & Perron, P. (201);** “Lag length selection and the construction of unit root tests with good size and power”; In: Econometrica 69:1519-1554

**Osborne, M. F. M. (1962);** “Periodic Structure in the Brownian Motion os Stock Prices”;  
In: Operations Research, Vol. 10

**Rahbek, A., J.G. Dennis and E. Hansen (2002);** “ARCH Innovations and their Impact  
on Cointegration Rank”; In: Working Paper, Department of Statistics and Operations Research -  
University of Copenhagen

**Richards, Anthony J. (1995);** “Comovements in National Stock Market Returns:  
Evidence of Predictability but not Cointegration”; In: NBER Working Papers 3609

**Xu, K. L. (2008),** “Bootstrapping Autoregression under Non-stationary Volatility”; In:  
The Econometrics Journal, 11:1–26.