



Revista Brasileira de Finanças

ISSN: 1679-0731

rbfin@fgv.br

Sociedade Brasileira de Finanças

Brasil

Soto, Paula Andrea; Ruilova Teran, Juan Carlos  
Arbitragem Estatística: Uma Abordagem por VECM  
Revista Brasileira de Finanças, vol. 15, núm. 4, octubre-diciembre, 2017, pp. 537-568  
Sociedade Brasileira de Finanças  
Rio de Janeiro, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=305855645002>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal  
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

# **Arbitragem Estatística: Uma Abordagem por VECM**

**(A VECM Approach of Statistical Arbitrage)**

**Paula Andrea Soto\***

**Juan Carlos Ruilova Teran\*\***

## **Resumo**

Neste trabalho foi construído um modelo de arbitragem estatística, aplicado e testado no mercado brasileiro de ações. As séries de preços foram modeladas usando a técnica de VECM (*Vector Error Correction Models*) para aplicar uma estratégia de compra e venda de ações long-short, auto-financiável e neutra ao mercado. Nesta estratégia, os desvios no equilíbrio a longo prazo dos preços são identificados para criar sinais de compra e venda. Foi utilizada a técnica de componentes principais para a seleção de portfólios com tendências comuns. A viabilidade desta estratégia foi abordada empiricamente usando simulações nessas carteiras. Seu desempenho também foi comparado com outras estratégias de negociação do tipo *long-short* e todas foram analisados em termos de retornos, volatilidade e oportunidades de arbitragem estatística. Em cada estratégia os resultados mostram que a técnica proposta é adequada para a modelagem dos portfólios, e todas as estratégias testadas foram capazes de agregar valor para o investidor. A estratégia proposta destaca-se por apresentar arbitragem estatística. **Palavras-chave:** Arbitragem estatística, Cointegração, PCA, VECM.

**Códigos JEL:** G11, G14, G17.

---

Submetido em 18 de janeiro de 2017. Reformulado em 28 de agosto de 2017. Aceito em 18 de abril de 2018. Publicado on-line em 20 de junho de 2018. O artigo foi avaliado segundo o processo de duplo anonimato além de ser avaliado pelo editor. Editor responsável: Márcio Laurini.

\*Escola de Economia de São Paulo, Fundação Getúlio Vargas, São Paulo, SP, Brasil. E-mail: paulaandreasoto@hotmail.com

\*\*Escola de Economia de São Paulo, Fundação Getúlio Vargas, São Paulo, SP, Brasil. E-mail: juan.teran@fgv.br

*Rev. Bras. Finanças (Online), Rio de Janeiro, 15, No. 4, Oct. 2017, pp. 537-568*

*ISSN 1679-0731, ISSN online 1984-5146*

©2017 Sociedade Brasileira de Finanças, under a Creative Commons Attribution 3.0 license -  
<http://creativecommons.org/licenses/by/3.0>

## Abstract

This work develops a statistical arbitrage model which was tested on the Brazilian stock market. Prices were modeled using VECM (Vector Error Correction Models) to create a self-financing, market-neutral, long/short trading strategy. In this strategy, deviations in the long-term equilibrium of prices are identified in order to create buy and sell signals. Portfolios with common trends were selected by means of Principal Component Analysis. The viability of this strategy was empirically addressed using simulations on these portfolios. Its performance was also compared to other long/short trading strategies and were all analyzed in terms of returns, volatility and statistical arbitrage opportunities. The methodology used in this paper shows good results for modeling prices, and though all trading strategies offer considerable gains for the investor, the proposed strategy stands out by presenting statistical arbitrage. **Keywords:** Statistical arbitrage, cointegration, PCA, VECM.

## 1. Introdução

No escopo de gerenciamento de portfólios, são comuns práticas de otimização de alocação de ativos e de controle de riscos baseadas no retorno médio e correlação de retornos. Ao usar retornos na modelagem tem-se a vantagem, em geral, de trabalhar com séries estacionárias e com isso tem-se a possibilidade de usar muitos dos modelos de séries temporais conhecidos. No entanto, a diferenciação remove as tendências comuns das séries financeiras, as quais trazem consigo muita informação que pode ser usada na criação de estratégias de *trading*. Grupos de ativos que apresentam uma reversão a uma ou mais tendências comuns se dizem ativos cointegrados, e o estudo dessas tendências é particularmente interessante por poderem explicar equilíbrios de longo prazo, muitas vezes resultado de fatores econômicos relevantes.

Neste trabalho, são utilizados modelos de correção de erro (VECM) em portfólios de ativos cointegrados. Os portfólios são montados de forma a refletir as principais tendências do mercado brasileiro de ações, definidas previamente pela análise de componentes principais (PCA). Eventuais desvios dos preços do equilíbrio de longo prazo, representado pelos vetores de cointegração estimados, são interpretados como oportunidades de arbitragem. O objetivo do trabalho é propor uma estratégia de negociação cujos sinais sejam definidos pelos desvios desse equilíbrio. Verifica-se, através de simulações, que

a estratégia proposta cria oportunidades de arbitragem estatística e que ela não é apenas viável, mas apresenta baixa volatilidade e retornos superiores à taxa livre de risco e ainda de outras estratégias usuais do tipo *long-short*.

A invenção das técnicas de *long-short* é atribuída ao *trader* do banco Morgan Stanley, Nunzio Tartaglia, na década de 1980. Como o próprio nome sugere, estratégias *long-short* possuem tanto posições vendidas (aqui chamadas de posições *short* pela sua nomenclatura usual em inglês) quanto posições compradas (*long*). A operação em *short* em ações no Brasil é feita através do aluguel do ativo e sua venda imediata. Desta forma, o investidor obtém recursos que podem ser utilizados na operação comprada. Este é um tipo de alavancagem, e a estratégia pode inclusive se tornar autofinanciável caso os dois lados da operação sejam casados. Com origem nos anos 1980, o *pairs trading* é um subgrupo das estratégias *long-short* e é considerado o pai da arbitragem estatística. A ideia consiste em identificar um par de ativos cujos preços sigam uma tendência comum, de modo que a distância entre eles, chamada de *spread*, apresente reversão à média. Por se tratar de um processo estacionário, este *spread* pode ser modelado para gerar sinais de *trade* em estratégias *long-short*. Dentre os trabalhos relevantes de *pairs trading*, destacam-se Elliott et al. (2005) que sugerem a modelagem do *spread* através de um processo Ornstein-Uhlenbeck. O trabalho foi depois estendido por Triantafyllopoulos and Montana (2011), introduzindo coeficientes dependentes do tempo. No contexto brasileiro, sabe-se que é possível encontrar diversos pares de ações cointegrados através de testes de cointegração (Caldeira and Moura (2013)) ou selecionando ativos de uma mesma empresa ou índices e ETFs (*Exchange Traded Funds*), (Pucciarelli (2014)).

Com o passar dos anos, a técnica de *pairs trading* foi generalizada de modo a aumentar o número de ativos no *trade*, e o primeiro passo nesta direção é estabelecer métodos para montar os portfólios cointegrados. Para isto, é fundamental perceber que a procura por um grupo de ativos que sejam cointegrados é análoga à procura por grupos com tendências comuns. Esta noção foi descrita por Stock and Watson (1988), onde é mostrado que séries cointegradas compartilham ao menos uma tendência comum. Por conseguinte, a construção do portfólio pode ser feita com ênfase em uma dessas duas frentes.

Sobre o estudo de tendências, surgiu uma vasta literatura para explorar a aplicação de métodos existentes de modelagem multivariada em séries não estacionárias. O estudo de tendências comuns em um número grande de séries através de modelos de fatores dinâmicos foi abordado em Bai and Ng (2002), Bai and Ng (2013), e Amengual and Watson (2006). Em Peña and Poncela (2006) fazem uma abordagem formal de modelos de fatores dinâmicos para séries não estacionárias e propõem um método para determinar o número de tendências comuns pelo número de fatores não estacionários. Ainda sobre o número de tendências estocásticas em um conjunto de ativos, importantes avanços foram feitos por Chang et al. (2005), e Chang et al. (2008). Esses autores propõem uma reformulação do VECM com *lags* infinitos dentro de um modelo de espaço-estado, cuja estimação é feita através do filtro de Kalman. Focardi et al. (2016), fazem uma ponte interessante entre o uso de modelos de fatores dinâmicos e estratégias de arbitragem estatística: o trabalho mostra que modelos de fatores dinâmicos proporcionam boas previsões para todo o conjunto das séries de logpreços dos ativos pertencentes ao índice S&P. Eles defendem que existe uma tendência comum predominante entre os ativos do índice, que é bem capturada pelo modelo, o que permite aplicar estratégias simples do tipo *long-short* baseadas em retornos esperados.

Uma estratégia muito comum é a técnica de *Index Tracking*. Muitas estratégias de *long-short* são derivadas desta técnica baseada na ideia de que um índice e seus ativos componentes são cointegrados por construção. Os primeiros modelos de *Index Tracking* tiveram um perfil passivo e procuravam apenas replicar os retornos de um índice escolhido. Eles nasceram da necessidade de medir a performance de um portfólio em relação a um *benchmark*. O desempenho do *Index Tracking* (Rudd (1980)) é definido pela análise do *tracking error*, diferença entre os retornos do portfólio e os retornos do índice ao longo do tempo. O racional é que o portfólio deve conseguir os mesmos retornos do índice, com volatilidade mínima no *tracking error*. Os pesos de alocação para todos os ativos do índice têm uma solução analítica (Roll (1992)) mas, geralmente, o portfólio é constituído por um número reduzido de ativos de modo a diminuir o custo de transações. A grande limitação desses modelos é que sua otimização ainda é feita em termos da matriz de covariância dos re-

tornos dos ativos, que reflete equilíbrios de curto prazo e não garante uma boa performance a longo prazo. Assim, se a escolha do subgrupo de ativos e os respectivos pesos não for ótima, a performance do portfólio pode se afastar frequentemente do índice em questão, gerando maiores custos de ajuste.

Gregoriou et al. (2011), fazem uma revisão da literatura da técnica e separam as estratégias em três subgrupos: a forma tradicional de modelos baseados na média-variância, modelos baseados em cointegração, e modelos baseados em fatores comuns. Para Alexander (1999), a existência de cointegração entre um portfólio candidato e o índice é pode ser verificada pelo teste de Engle-Granger e é usada como principal critério de seleção de ativos. Com o uso de modelos de correção de erros, as séries são modeladas na forma de logpreços, e não mais logretornos. Os pesos são definidos regredindo o índice contra as séries, o que minimiza o *tracking error* em portfólios devidamente cointegrados. O método tenta assegurar que os pesos reflitam uma relação de longo prazo, precisando de menos rebalanceamentos e diminuindo os custos de operação. A técnica apresenta um ganho de performance em relação ao *benchmark* e à forma tradicional de média-variância do *Index Tracking* (Alexander and Dimitriu (2004a), Alexander and Dimitriu (2005b)). Estratégias de *long-short* são uma importante extensão deste tipo de *Index Tracking* baseado em cointegração (Alexander and Dimitriu (2002); Alexander and Dimitriu (2005a)). Caldeira and Portugal (2010), fazem uma aplicação dessa técnica no mercado brasileiro de ações e comprovam que é possível montar carteiras rentáveis e com baixo custo de rebalanceamento. A diversidade de aplicações tornou o *Index Tracking* e suas vertentes amplamente utilizados na prática.

Uma terceira variação do método é abordada em Alexander and Dimitriu (2004b), onde generaliza-se a técnica ao tentar replicar uma tendência ao em vez de um índice específico. A seleção de ativos cointegrados é baseada na análise de componentes principais (PCA), considerando os ativos pela sua maior participação no primeiro componente e criando assim um portfólio de boa representatividade do universo considerado. O próprio PCA pode ser utilizado para obter estimativas dos vetores de cointegração (Harris (1997)), e graças à sua simplicidade, ganhou força como método de identificação de tendências.

Neste trabalho, os portfólios candidatos são selecionados pela contribuição dos ativos nos fatores não estacionários estimados por PCA. Seguidamente, os preços das séries são modelados por VECM, e propõe-se uma estratégia de *trading*, na qual os pesos de alocação são determinados pelos vetores de cointegração estimados e os desvios do equilíbrio histórico são aproveitados para criar ganhos através de estratégias *long-short*. Salienta-se que essa estratégia não tem como objetivo replicar um índice ou uma tendência comum, e sim apenas constatar os erros de especificação que geram variações no equilíbrio de longo prazo dos ativos. Para fins de comparação, três estratégias de *trading*, também do tipo *long-short*, baseadas em retornos esperados na mesma linha de Focardi et al. (2016), são simuladas no mesmo cenário, com a ressalva de que os retornos esperados são gerados através do VECM e não do modelo de fatores dinâmicos. O trabalho é relevante dentro da literatura brasileira por dois aspectos principais: (i) é apresentada uma metodologia para detectar oportunidades de arbitragem ao observar desvios do equilíbrio de longo prazo estimados pelos vetores de cointegração, e (ii) é estudada a viabilidade da estratégia proposta e de outros tipos de estratégia *long-short*, em portfólios que representam as principais tendências do mercado brasileiro. Os resultados das diferentes estratégias são comparados em termos de retornos, volatilidade e existência de arbitragem estatística através do teste proposto em Jarrow et al. (2004); e Jarrow et al. (2012).

## 2. Metodologia

### 2.1 Modelo vetorial de correção de erros (VECM)

Quando se tem um portfólio com tendências comuns, a reversão dos preços a estas tendências pode ser modelada através dos vetores de cointegração. Uma forma de obter estimativas destes parâmetros é através de uma adaptação dos familiares modelos autorregressivos, conhecida como modelos de correção de erros (VECM). O modelo aqui abordado é da forma

$$\Delta X_t = AB'X_{t-1} + \sum_{j=1}^b B_j \Delta X_{t-j} + \epsilon_t, \quad (1)$$

onde  $X_t$  é uma matriz  $N \times 1$  que contém os valores das  $N$  séries

de preços no tempo  $t$ ;  $\Delta X_{t-j}$  é uma matriz  $N \times 1$  contendo a primeira diferença das séries no tempo  $t - j$ ;  $b$  é a quantidade de *lags* considerados, e  $\epsilon_t$  é uma matriz  $N \times 1$  de erros não correlacionados serialmente.

O chamado termo de correção  $B'X_{t-1}$  mede a distância das séries em relação a um equilíbrio estacionário definido pelas tendências comuns. A matriz  $B$  de dimensão  $N \times q$  contém em suas colunas os vetores de cointegração, sendo  $q$  o número de vetores de cointegração incluídos no modelo. Os vetores em  $B$  podem ser utilizados para formar a combinação linear  $B'X = S$  que é estacionária. A matriz  $A$  de dimensão  $N \times q$  contêm os parâmetros que indicam a velocidade com que as séries revertem às tendências comuns.

Para definir a estrutura do modelo, o procedimento usual envolve realizar testes de cointegração para o conjunto de séries consideradas e identificar à priori o número de vetores de cointegração  $q$ . Este número está diretamente ligado ao de tendências comuns: se em um conjunto de  $N$  séries existem  $r < N$  tendências comuns, então existem  $q = N - r$  vetores de cointegração. Por isso, uma abordagem equivalente a procurar o número de vetores de cointegração é procurar pelo número de tendências comuns. A estimativa dos parâmetros é feita através de máxima verossimilhança.

## 2.2 As estratégias de *trading*

Uma estratégia de *trading* define quando e de que maneira são feitas e desfeitas as posições em um portfólio. O ato de tomar uma posição é chamado de *enter* e o ato de desfazer uma posição tomada é chamado de *exit*. Por sua vez, as posições podem ser compradas (chamadas de *long*, especificadas por pesos positivos) ou vendidas a descoberto (chamadas de *short* e especificadas por pesos negativos). Quando uma estratégia assume posições em *short*, os ativos são alugados e vendidos em seguida. Não é abordado o uso de derivativos. O aluguel de um ativo no mercado brasileiro é feito através do sistema BTC da BMF&BOVESPA. No período de 2011 a 2015, o mercado de aluguel de ações já estava bem desenvolvido e movimentava em média 65 bilhões de reais mensais. As condições de cada contrato são específicas para cada operação e é comum que ativos mais líquidos, como é o caso dos que compõem os portfólios deste trabalho, possuam também maior oferta de contratos de aluguel e com taxas mais

acessíveis.

A primeira estratégia definida abaixo remete ao objetivo principal deste trabalho: aproveitar o modelo descrito em [1] para gerar sinais de negociação quando existirem evidências de falhas de precificação, e nessas ocasiões definir os pesos de alocação de modo a formar uma estratégia *long-short*. São também definidas outras estratégias que seguem os moldes de Focardi et al. (2016) e servem para comparação de resultados.

### **2.2.1 Estratégia definida em função dos vetores de cointegração**

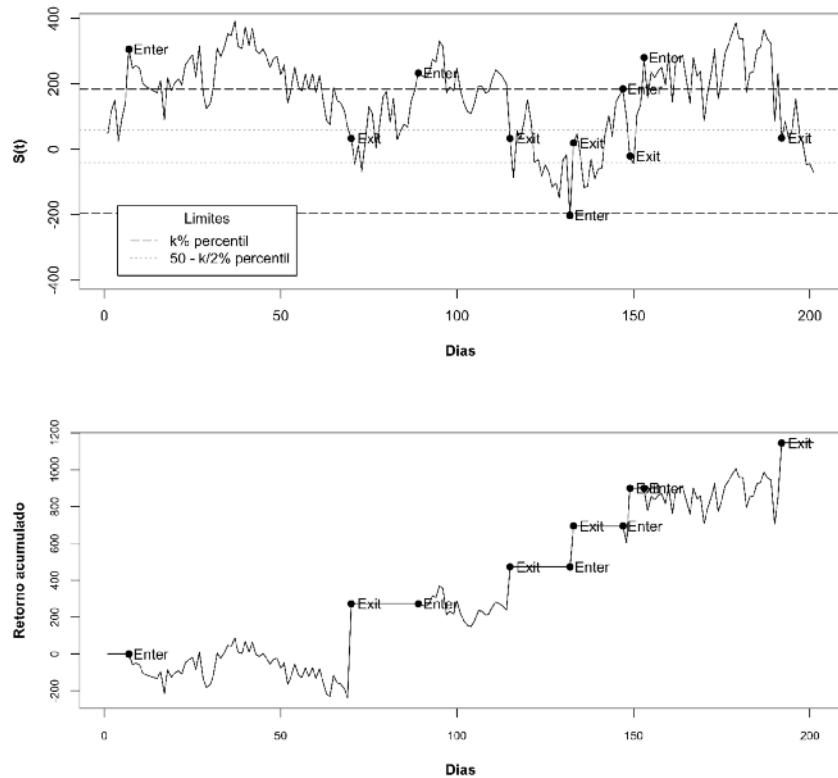
A estratégia explora o fato de que, dado um conjunto de vetores de cointegração  $B$ , as séries  $S_t$  dadas por  $S_t = B'X_t$  são estacionárias. Para facilitar a compreensão, detalha-se a seguir o caso de modelos com apenas um vetor de cointegração (isto é, a matriz  $B$  contém apenas um vetor) e depois é abordado o tratamento para incluir mais vetores.

Sabendo-se que  $S_t$  é uma série estacionária, quanto mais distante  $S_t$  estiver da sua média histórica, maior a reversão que deverá ocorrer nas próximas observações. Já que os vetores  $B$  são assumidos constantes, este processo de reversão será atribuído somente a alterações nos preços  $X_t$ . Os vetores  $B$  representam relações de longo prazo dos ativos, e um valor extremo de  $S_t$  em um determinado momento é um indício de que há desvios na precificação de um ou mais ativos e que existe uma possibilidade de arbitragem. Para recriar um portfólio que imite o movimento de reversão observado em  $S_t$ , as quantidades dos ativos que devem ser compradas ou vendidas a descoberto a cada sinal de *enter* são definidas diretamente pelos vetores de cointegração. Se em um dado momento o valor de  $S_t$  estiver abaixo da média histórica, então as quantidades dos ativos no portfólio devem ser proporcionais aos valores de  $B$ . Da mesma forma, se o valor de  $S_t$  estiver acima da sua média histórica, as quantidades devem ser proporcionais a  $-B$ . Os ganhos devem ser recolhidos quando acontecer a mencionada reversão. É este processo de reversão de  $S_t$  que permite prever o movimento de um portfólio de ativos como um todo, sem de fato precisar fazer previsões das séries.

Um exemplo de como os sinais de *enter* e *exit* são definidos através do posicionamento de  $S_t$  é apresentado na Figura 1. Neste trabalho

**Figura 1**

Exemplo de como são gerados os sinais de *trading* na estratégia de valores extremos de  $S_t$  e o retorno acumulado referente à estratégia.



sugere-se que valores extremos de  $S_t$  sejam identificados a partir dos percentis. Os extremos indicam uma possibilidade de arbitragem e é o momento de tomar uma posição. As posições devem ser desfeitas quando ocorrer uma reversão para a mediana histórica. Os percentis exatos que determinam aquilo que será considerado um valor extremo, e aquilo que será considerado próximo à mediana devem ser definidos a priori pelo pesquisador.

Se houver mais de um vetor de cointegração considerado no modelo, eles são somados para criar um novo vetor  $B_{cons} = \sum_{i=1}^q B_i$  e a lógica da estratégia continuará a mesma. Como qualquer combinação linear dos vetores de cointegração é vetor de cointegração também, sabe-se que a nova série  $S_t = B'_{cons} X_t$  será estacionária. Seja  $S_k\%$  o  $k$ -ésimo percentil histórico da série  $S_t$ , os sinais de negociação e os pesos de alocação são determinados da seguinte forma:

#### Estratégia de valores extremos de $S_t$ (1)

- Se  $S_t < S_k\%:$  Enter

Toma-se posição com quantidades proporcionais a  $B_{cons}$ . Se já havia posição tomada, mantém-se a posição anterior.

- Se  $S_t > S_{(50-\frac{k}{2})\%}:$  Exit

Se existe posição tomada nas condições do item anterior, desfaz-se a posição recolhendo os ganhos ou perdas.

- Se  $S_t > S_{(100-k)\%}:$  Enter

Toma-se posição com quantidades proporcionais ao contrário de  $B_{cons}$ . Se já havia posição tomada, mantém-se a posição anterior.

- Se  $S_t < S_{(50+\frac{k}{2})\%}:$  Exit

Se existe posição tomada nas condições do item anterior, desfaz-se a posição recolhendo os ganhos ou perdas.

- Se nenhum dos anteriores, mantém-se a posição atual.

Na estratégia proposta, os usuais benefícios de poucas transações e baixa volatilidade dos ativos do portfólio devem ser questionados, especialmente quando comparada a estratégias com rebalanceamentos periódicos em horizontes fixos de tempo, como é o caso das estratégias de *Index Tracking* aplicadas em Alexander (1999) e suas

extensões, ou as estratégias *long-short* baseadas em retornos esperados aplicadas em Focardi et al. (2016). A metodologia sugerida não possibilita antever o total nem a periodicidade de transações, visto que as posições do portfólio são feitas e desfeitas de acordo com a velocidade de reversão de  $S_t$  e não há outros rebalanceamentos. No caso de uma reversão lenta, serão realizadas menos transações e as posições deverão ser mantidas por maiores períodos de tempo, não havendo garantia de uma periodicidade fixa. Um número alto de transações não é por si só um fator preocupante, visto que elas podem ser resultado de múltiplas oportunidades de arbitragem. Em contrapartida, esta estratégia também possibilita não tomar nenhuma posição enquanto não forem observados valores extremos de  $S_t$ , o que permite não adquirir risco nem gerar custos enquanto erros de precificação não forem detectados.

A metodologia traz mais duas particularidades importantes. Em primeiro lugar, o modelo não pode ser estimado utilizando logpreços e devem ser utilizadas as próprias séries de preços. Não há dúvidas de que logpreços são preferidos na literatura por formarem séries mais suaves do que as séries de preços propriamente ditas e se um grupo de séries é cointegrado seu logaritmo também será. No entanto, a transformação logarítmica elimina a natureza linear do equilíbrio estimado pelos vetores de cointegração  $B$  e as quantidades sugeridas para a alocação não podem ser transformadas apropriadamente. A outra particularidade é que o total das posições em *long* não necessariamente irá coincidir com o total das posições em *short*. Isto é relevante para a implementação, pois a estratégia não tem garantia de ser autofinanciável. Para contornar este problema nas simulações deste trabalho, foi imposto um controle sobre o volume negociado: os pesos são sempre ajustados para que tanto o total comprado quanto o total vendido estejam limitados a um capital predeterminado C (desconsiderando os custos de transação). A cada negociação, os recursos obtidos pela venda a descoberto são usados em sua totalidade para cobrir os gastos da compra. Quando não for suficiente, novos recursos são captados através da taxa livre de risco. Também são captados recursos para cobrir os custos de transação. Na prática, esta medida faz com que seja possível garantir a operação ao manter um capital próprio C disponível.

## 2.2.2 Estratégias definidas em função de retornos esperados

Para poder ter uma referência do desempenho da estratégia proposta acima, os resultados foram comparados àqueles de três estratégias usuais baseadas em retornos esperados, descritas em Focardi et al. (2016). Nesta segunda abordagem, atualizações nas posições são feitas a cada horizonte de previsão  $hf$ . Para cada atualização, novas previsões são geradas para o próximo horizonte  $[t, t + hf]$  pelas estimativas dadas pelo VECM, e as posições são sempre definidas com base nos retornos esperados  $R_{t+hf,i} = (X_{t+hf,i} - X_{t,i})/X_{t,i}$ . Como será descrito a seguir, neste tipo de abordagem as quantidades dos ativos no portfólio não dependem diretamente dos vetores de cointegração, portanto a estimação do modelo pode ser feita sob logpreços e consequentemente as previsões obtidas devem ser devidamente transformadas. Além disso, os pesos sempre podem ser controlados para que a estratégia seja autofinanciável.

Os pesos para cada estratégia são determinados da seguinte forma:

### **Dummy - Estratégia de pesos iguais (2)**

Dada uma porcentagem limite  $l$ , selecionar os  $l\%$  ativos com maiores retornos esperados e alocar posições compradas com pesos iguais. Selecionar os  $l\%$  ativos com menores retornos esperados e alocar posições vendidas com pesos iguais.

### **Max Sharpe - Estratégia de otimização do Sharpe Ratio (3)**

Os pesos de alocação devem ser definidos de forma a maximizar a estatística *Sharpe Ratio* dos retornos esperados. A estatística é dada por  $Sharpe\ Ratio = \mu/\theta$ , onde  $\mu$  e  $\theta$  são estimados pela média e desvio padrão dos retornos esperados, respectivamente. Para a maximização da estatística, consideram-se os  $l\%$  ativos com maiores retornos para as posições em *long* e os  $l\%$  ativos com menores retornos para as posições em *short*.

### **Benchmark - Estratégia de pesos proporcionais ao retorno esperado (4)**

Alocação é feita com pesos proporcionais à diferença dos retornos esperados em relação à média dos retornos esperados no portfólio. Isto é, os pesos são proporcionais às diferenças  $R_{t,i} - \frac{1}{N} \sum_{i=1}^N R_{t,i}$ .

### 3. Seleção de portfólio

#### 3.1 Os dados

Os dados são compostos pelos preços de fechamento diário do mercado de ações local - BMF&BOVESPA. A coleta de informações foi feita através do sistema de informações Bloomberg. O histórico para o universo analisado levou em consideração o tempo de amadurecimento do mercado. Grandes marcos da economia brasileira aconteceram até o ano 2000, como o plano real implantado em 1996, ou a política monetária com metas para a inflação introduzida em 1999. Assim, dados mais antigos não seriam uma amostra representativa do mercado atual. O tamanho total da amostra foi definido em nove anos: janeiro de 2007 a dezembro de 2015. As observações dos primeiros quatro anos formam a pré-amostra e foram usadas unicamente para o primeiro ajuste do modelo, enquanto as estratégias foram simuladas de 2011 a 2015 utilizando uma janela móvel para estimação. Este período é particularmente interessante para a realização das simulações, já que o IBOVESPA do período teve resultado negativo e pode ser considerado um cenário adverso para investimentos em renda variável. Como a estratégia proposta pretende ser neutra ao mercado, isto não deveria ser um impedimento para conseguir retornos positivos.

Os ativos considerados na amostra foram selecionados levando em consideração os seguintes critérios: 1) o ativo deve existir para negociação durante os nove anos incluídos na amostra, 2) ter pelo menos uma transação por dia para todo o histórico, 3) ter boa liquidez para contratos de aluguel, visto que é uma ferramenta essencial das estratégias *long-short* e 4) ter pertencido, em algum momento dos 9 anos da amostra, ao índice IBX100.

O critério 3, referente à liquidez de contratos de aluguel, foi analisado de forma julgamental devido à falta de informações sobre esses contratos. Ativos que não tiveram nenhum registro de contrato de aluguel por semanas seguidas ao longo de todo o histórico foram desconsiderados. O critério 4 é um critério adicional de liquidez, e optou-se pelo índice IBX100 em vez do IBOVESPA apenas pelo número maior de ativos, já que ambos os índices são compostos apenas por ativos líquidos. Depois de aplicar os quatro critérios, foi determinado um universo de 76 ativos.

### 3.2 Determinando a existência de tendências comuns

Para conhecer as tendências comuns da amostra, foram extraídos os fatores latentes pela técnica de componentes principais. Há dois objetivos primordiais para esta análise: usar o conhecimento sobre tendências comuns entre ativos para montar portfólios cointegrados, e posteriormente definir o número de vetores de cointegração para serem usados na formulação do VECM.

Inicialmente, é preciso identificar o total de fatores não estacionários, que representam tendências comuns nos dados, e também verificar se o total destas tendências mantém-se invariável ao longo do tempo. O ideal para poder trabalhar com apenas um tipo de parametrização no modelo ocorre quando os dados mantêm uma estrutura semelhante em todo o histórico. Desse modo, os fatores comuns foram extraídos para cada ano separadamente, e o número de fatores não estacionários ao longo dos anos foi estudado através de testes univariados KPSS e de raiz unitária. O resultado do teste KPSS para os fatores extraídos por PCA para intervalos de um ano está especificado na Tabela 1. O teste foi aplicado a todos os N fatores latentes, mas por simplicidade mostram-se aqui apenas os resultados dos primeiros oito fatores. Observa-se que, para um nível de significância de 5%, os três primeiros fatores são não estacionários para a maioria dos anos na amostra. A análise também mostrou que os três primeiros fatores estimados por PCA são suficientes para explicar 80% da variância para a maioria dos anos.

Pelo teste de raiz unitária, nenhum dos três primeiros fatores mostrou-se estacionário. Considerando também outras análises como técnicas de componentes independentes, concluiu-se que podem existir até três tendências comuns no universo estudado.

### 3.3 Seleção de portfólios candidatos

Apesar dos testes terem detectado tendências comuns no universo considerado, é improvável que elas influenciem na mesma medida todos os ativos da amostra, e seria impraticável trabalhar com um portfólio tão numeroso. Surge assim a necessidade de identificar grupos de ativos dentro do qual as tendências estejam bem representadas. Para fazer isto, lembramos que o PCA usa uma representação

**Tabela 1**

Valor-p do teste KPSS para fatores estimados por PCA, por ano.

2015	2007	2008	2009	2010	2011	2012	2013	2014
Fator 1 0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*
Fator 2 0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*
Fator 3 0.01*	0.02*	0.03*	0.01*	0.02*	0.01*	0.02*	0.01*	0.01*
Fator 4 0.08	0.04*	0.07	0.09	0.02*	0.10	0.01*	0.03*	0.05
Fator 5 0.10	0.02*	0.10	0.10	0.10	0.08	0.10	0.09	0.09
Fator 6 0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Fator 7 0.10	0.10	0.05	0.10	0.10	0.10	0.10	0.10	0.10
Fator 8 0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10

\*Resultados que rejeitam  $H_0$  para um nível de significância de 5%, indicando que o fator é não estacionário.

dos dados da forma  $X^s = F\Lambda'$ , onde  $X^s$  é o conjunto de séries padronizadas,  $F$  é a matriz de fatores latentes e  $\Lambda$  é a matriz de carregamentos (melhor conhecidos como *factor loadings*). Ao trabalhar com uma amostra padronizada, os *factor loadings* refletem a influência de cada ativo na composição das tendências e é razoável supor que ativos com maiores pesos  $\Lambda$  são os mais influenciados pelos verdadeiros fatores latentes. Por isso, é possível identificar grupos de ativos cointegrados selecionando os  $N$  ativos com maiores valores de  $\Lambda$  correspondentes aos fatores não estacionários. Com o intuito de não utilizar informação futura, a matriz  $\Lambda$  estimada na pré-amostra foi usada como critério de seleção para os portfólios (a) e (b) utilizados nas simulações deste trabalho. Os portfólios são descritos a seguir.

- (a) Portfólio com um fator latente:

Para montar este portfólio, foram selecionados os  $N = 10$  ativos com os maiores elementos de  $\Lambda$  correspondentes ao primeiro fator. Os ativos selecionados foram, na ordem dos pesos dados por  $\Lambda$ : Randon (RAPT4), Banco Itaú (ITUB3), Banco do Brasil (BBAS3), Banco Bradesco (BBDC4 e BBDC3), Saraiva (SLED4), Lojas Americanas (LAME4), Bradespar (BRAP4), Cyrela Brazil Realty (CYRE3) e Localiza (RENT3).

- (b) Portfólio com três fatores latentes:

Este portfólio é composto pelos ativos correspondentes aos 10

maiores elementos de  $\Lambda$  relacionados aos  $r = 3$  fatores latentes especificados na seção anterior. O portfólio é composto pelos seguintes 30 ativos: Randon (RAPT4), Banco do Brasil (BBAS3), Banco Itaú (ITUB3), Banco Bradesco (BBDC4 e BBDC3), Lojas Americanas (LAME4), Saraiva (SLED4), Bradespar (BRAP4), Cyrela Brazil (CYRE3), Localiza (RENT3), Unipar Carbocloro (UNIP6), B2W (BTOW3), Companhia Brasileira de Distribuição (PCAR4), Embraer (EMBR3), Vanguarda Agro (VAGR3), Telefônica Brasil (VIVT4 e VIVT3), Natura Cosméticos (NATU3), Totvs (TOTS3), Companhia Transmissora de Energia Elétrica Paulista (TRPL4), Petrobras (PETR3 e PETR4), Companhia Siderúrgica Nacional (CSNA3), CSU Cardsystem (CARD3), Oi (OIBR4), Gerdau (GOAU4), Gerdau (GGBR3 e GGBR4), Eletropaulo (ELPL4) e Usiminas (USIM5).

O objetivo do primeiro portfólio é analisar se é mais vantajoso para estratégias de *trading* utilizar grupos de ativos, cuja influência depende majoritariamente de apenas um fator, do que ampliar o portfólio para representar o efeito de todas as tendências comuns. O número final de ativos foi escolhido com base em testes preliminares com amostras de  $N = 10, 20$  e  $30$  ativos, que mostraram que os melhores resultados eram obtidos quando  $N = 10$ . Sendo resultado de uma tendência comum evidente, o primeiro portfólio também visa representar outros tipos usuais de portfólios de *pairs trading* usados na literatura, como aqueles formados por ativos da mesma empresa ou ativos do mesmo setor. Já o segundo portfólio, intencionalmente mais numeroso, tem maior informação sobre as três principais tendências do mercado. Um atrativo em ambos os portfólios é a diversidade dos setores das empresas selecionadas, principalmente ao compará-los àqueles selecionados dentro de um mesmo setor de atuação. Além disso, diferentemente de outras metodologias usuais de seleção, nos portfólios selecionados através da técnica de PCA as tendências comuns detectadas não são facilmente perceptíveis ao observador.

Como mencionado anteriormente, a existência de tendências comuns já é indício de cointegração, mas para certificar-se formalmente de que os portfólios são cointegrados, foi aplicado o teste de cointegração de Engle-Granger. Para ambos os portfólios, a combinação linear das séries do primeiro autovetor forma uma série es-

tacionária. Conclui-se que os ativos que compõem cada um dos portfólios selecionados são de fato cointegrados. Outra forma de mostrar a existência de cointegração é analisar se é possível obter séries estacionárias através das combinações lineares das séries dadas pelos vetores de cointegração estimados pelo VECM. A estacionariedade de  $S_t$  também é particularmente importante pois permite o uso da estratégia de *trading* (1) sugerida. Como um total de três tendências comuns foram detectadas, a parametrização mais completa do VECM deve incluir os  $N - 3$  vetores de cointegração. No entanto, não seria necessário que todos eles sejam considerados, caso apenas um vetor seja capaz de criar uma série  $S_t$  estacionária que permita a aplicação da estratégia proposta.

Na Figura 2 são mostradas as séries  $S_t = B'_{cons}X_t$ , onde  $B_{cons}$  é a soma de todos os vetores de cointegração somados no modelo,  $\sum_{i=1}^q B_i$ . Os dados correspondem à pré-amostra (2007-2010). Em primeiro lugar, observa-se que para ambos os portfólios a série  $S_t$  é estacionária para os modelos cuja parametrização considera tanto um vetor, quanto  $N - r$  vetores de cointegração. Conclui-se que os portfólios são devidamente cointegrados, e utilizar VECM para modelar os conjuntos de séries é adequado.

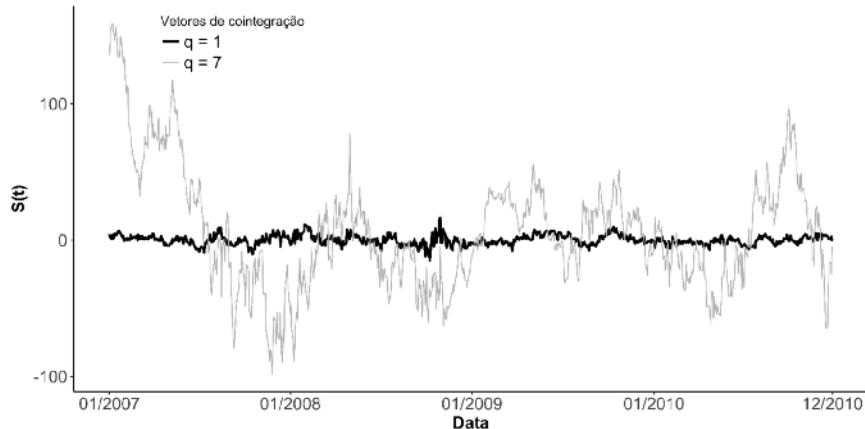
É devido se atentar também à velocidade de reversão à média de  $S_t$ . Apesar do modelo ser menos explicativo, quando apenas um vetor de cointegração é considerado, para ambos os portfólios, a série  $S_t$  tem uma menor amplitude e uma reversão mais rápida. Isto permite um número maior de transações por ano, mas com uma oportunidade de retorno por transação reduzida. Já quando  $N - r$  vetores são considerados, a série tem uma reversão mais lenta mas há uma maior distância entre valores extremos e a média. Visto que em ambas parametrizações é alcançada a estacionariedade de  $S_t$ , a parametrização para o desempenho ótimo da estratégia (1) irá depender da relação entre a frequência das transações, os custos de negociação, e os ganhos obtidos em cada transação.

#### 4. Implementação

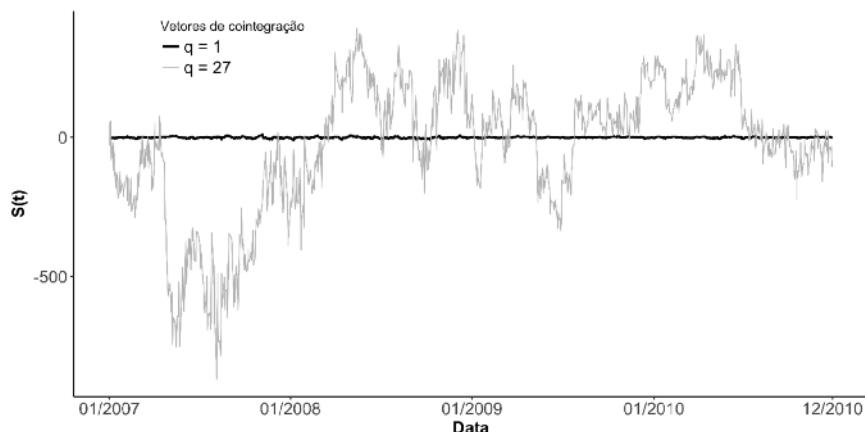
As simulações descritas a seguir foram feitas através de um código na linguagem **R** e contemplam as quatro diferentes estratégias de *trading* aplicadas nos portfólios de um e três fatores latentes.

**Figura 2**

Séries  $S_t$  para os dados da pré-amostra (2007-2010) nos portfólios de um e três fatores latentes.



(a) Portfólio de um fator



(b) Portfólio de três fatores

As negociações foram simuladas em um período de praticamente cinco anos: fevereiro de 2011 a dezembro de 2015. Para a estimação dos modelos foi utilizada uma janela móvel de quatro anos, considerando cada ano como 252 dias úteis. Para a estratégia (1), onde as quantidades para alocação são definidas pelos vetores de cointegração, o modelo foi estimado utilizando preços, enquanto que para as estratégias (2), (3) e (4), o modelo foi estimado utilizando log-preços e as previsões foram depois devidamente transformadas em retornos esperados. Para a estimação dos modelos, foi utilizado 1 *lag*, como sugerido pelo critério de Akaike e critério Bayesiano (AIC e BIC respectivamente) na pré-amostra.

A quantidade de vetores de cointegração que devem ser incluídos no modelo poderia ser definida também por critérios de informação, mas o número ótimo para o modelo não necessariamente será o número ótimo de vetores a ser considerado para a composição da estratégia de *trading* proposta. Como foi mencionado anteriormente, a viabilidade estratégia (1) depende do comportamento de  $S_t$ , de modo que a quantidade de transações e os retornos por transação sejam superiores aos custos de negociação. É completamente plausível que este cenário aconteça usando um vetor de cointegração apenas. No entanto, espera-se que o melhor resultado seja obtido ao montar um portfólio completo, representativo das principais tendências do mercado, e ao utilizar toda a informação disponível para detectar corretamente desvios do equilíbrio de longo prazo. Para mostrar esse fenômeno, as estratégias serão testadas com um e com  $N - r$  vetores de cointegração, sendo  $N$  o total de ativos no portfólio.

Ao observar as séries  $S_t$  e após testes preliminares, ficou evidente que limites maiores para definir os percentis extremos de  $S_t$  permitem uma maior flexibilidade para gerar sinais de negociação e consequentemente aumentam o número de transações, beneficiando a performance da estratégia mesmo que isto acarrete em um aumento de custos. Por isso, a estratégia (1) foi simulada utilizando o parâmetro  $k = 30$ .

Para conseguir um amplo repertório de estratégias para comparação, também foram testados diferentes valores para o horizonte de previsão  $hf$  e os limites  $l$ , parâmetros que guiam as decisões de negociação das estratégias baseadas em retornos esperados. O universo de parâmetros considerados para todas as estratégias é descrito na

Tabela 2.

#### 4.1 Custos de operação

Nas quatro estratégias descritas existe a possibilidade de ocorrem transações com frequência diária, portanto há gastos elevados de compra, venda e aluguel de ativos que não poderiam ser considerados irrisórios. Para simular corretamente o comportamento das estratégias no mercado, foram incluídos nos cálculos os custos aproximados para um investidor pessoa jurídica. Assume-se que o volume de negociação máximo C estará de acordo com este perfil.

Na prática, os preços de aluguel variam de acordo com o ativo e a data da negociação, sendo cada contrato definido de forma individual com o doador. Os dados sobre contratos de aluguel concretizados foram disponibilizados pela BOVESPA. Os custos referentes à taxa anualizada de aluguel  $a$  são calculados diariamente por:

$$\text{Aluguel} = \text{Volume negociado} * [(1 + a)^{1/252} - 1].$$

Os ativos que compõem os portfólios deste trabalho têm uma taxa de aluguel média de 3,5% ao ano para o período de 2011 a 2015. Uma suposição pertinente nas simulações é que o contrato de aluguel nunca pode ser terminado pelo doador, e que este sempre estará disposto a estender o contrato de aluguel a critério do tomador. Além da taxa de aluguel paga ao doador, paga-se a taxa de liquidação para a BMF&BOVESPA. Considera-se que o empréstimo é feito de forma automática, de modo que o custo de liquidação dos aluguéis para cada dia cobrado pela BMF&BOVESPA é dado por:

$$\text{Liq. fin. do empréstimo} = \text{Volume negociado} * (1,005^{1/252} - 1).$$

Os emolumentos e liquidação da compra de ações foram fixados em 0,025% por transação. A corretagem operacional foi fixada em 0,5% do volume da operação, com devolução de 80%. Assim, a taxa de corretagem final resulta em  $(0,5 * 0,2)\% = 0,1\%$ . A taxa de custódia é desconsiderada, pois ela não está diretamente relacionada à operação, nem foram considerados descontos referentes a impostos.

Assume-se que não há reinvestimento dos retornos dentro da estratégia: quando uma posição é desfeita, os ganhos são recolhidos e investidos na taxa livre de risco. Caso haja prejuízo, o valor será tomado emprestado também pela taxa livre de risco. O conceito de taxa livre de risco no mercado brasileiro para investidores institucionais é 100% do CDI.

**Tabela 2**  
Combinacão de parâmetros testados para cada estratégia

Nr.	Nome da estratégia	Descrição	Vetores de cointegração	Outros parâmetros
1	Extremos de $S_t$	Pesos proporcionais aos vetores de cointegração.	$q = 1, N - r$	$k = 30$
2	<i>Dummy</i>	Pesos iguais para os $\frac{1}{N}$ ativos com maiores e menores retornos.	$q = 1, N - r$	$l = 30, 50, hf = 1, 10, 30$
3	<i>Max Sharpe</i>	Otimização <i>Max Sharpe</i> para os $\frac{1}{N}$ ativos com maiores e menores retornos.	$q = 1, N - r$	$l = 30, 50, hf = 1, 10, 30$
4	<i>Benchmark</i>	Pesos proporcionais às diferenças $R_{t,i} - \frac{1}{N} \sum_{i=1}^N R_{t,i}$	$q = 1, N - r$	$hf = 1, 10, 30$

## 5. Resultados

Os resultados das simulações estão detalhados a seguir para cada um dos portfólios estudados ao longo dos anos 2011 a 2015, especificando os parâmetros de cada simulação:  $q$  (total de vetores de cointegração considerados no modelo),  $k$  (os percentis que determinam as negociações na estratégia de valores extremos de  $S_t$ ,  $l$  (limites para as posições compradas e vendidas das estratégias definidas em função de retornos esperados) e  $hf$  (horizonte de previsão para a estratégia definida em função de retornos esperados). O comportamento dos retornos da estratégia (1) aqui proposta também é comparado à evolução do CDI do período, tratada aqui como taxa livre de risco. Deve ser lembrado, no entanto, que um investimento em renda fixa indexado ao CDI requer um capital próprio aplicado. A estratégia proposta, por outro lado, precisa do capital próprio apenas para cobrir as chamadas de margem da BOVESPA, permitindo que o investidor aplique este montante na própria taxa livre de risco e assim potencializar o rendimento.

### Portfólio de um fator latente

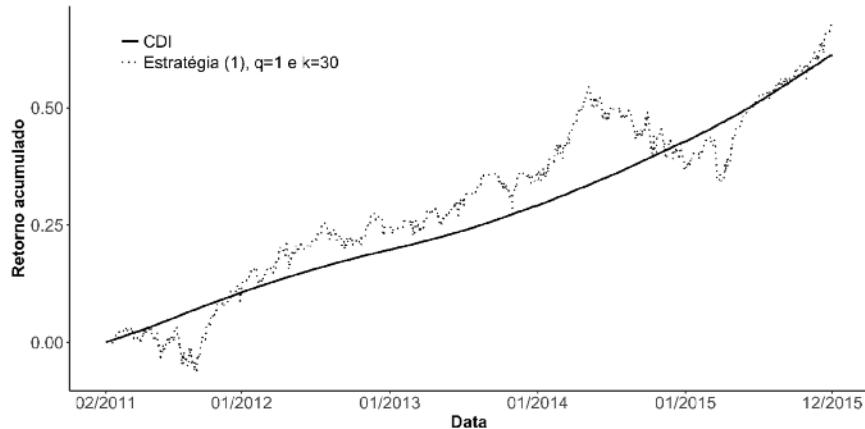
Os resultados das simulações para todas as estratégias neste portfólio são detalhados na Tabela 3. Os detalhes da estratégia (1) ao longo dos anos estão detalhados na Figura 3.

Para a especificação das negociações desta estratégia (1), o uso de apenas 1 vetor de cointegração acarreta em maiores retornos que o uso de 7 vetores. Observa-se na Tabela 3 que são 40.4 e 28.8 negociações ao ano nas simulações com  $q = 1$  e  $q = 7$ , respectivamente, valores que representam em média uma transação a cada 10 e 7 dias úteis aproximadamente. A parametrização com apenas um vetor de cointegração apresentou rendimentos de 69.4% do capital máximo de negociação, acima dos 61.3% referente ao CDI do período, e mostra-se como uma alternativa de investimento resiliente mesmo em um mercado adverso, dado que o IBOVESPA do período acumulou resultado negativo (-36%). Além disso, a simulação com um vetor tem uma volatilidade anualizada de 13.4%, a menor entre todas as estratégias testadas neste portfólio. Este resultado mostra que a estratégia (1) é viável mesmo em portfólios reduzidos, mas, como será visto mais à frente, pela pouca informação contida nos ativos do portfólio, não é atingido todo seu potencial.

Para as estratégias baseadas em retornos esperados, por outro

**Figura 3**

Retornos da estratégia (1) no portfólio de um fator para as diferentes parametrizações.



lado, o parâmetro de maior influência é o horizonte de previsão  $hf$ . Diferentemente da estratégia (1), com um horizonte fixo para realização de transações há custos recorrentes mesmo quando a negociação não é favorável. Isto está refletido na baixa performance em todas as simulações com  $hf = 1$ . Em particular, nota-se que estratégia *Max-Sharpe* tem perdas constantes para o horizonte de um dia, já que transações diárias não são adequadas para uma otimização que depende da variância. As simulações apresentam em sua maioria uma volatilidade anualizada acima dos 20%. Mesmo assim, em concordância com as conclusões de Focardi et al. (2016), as estratégias *long-short* baseadas em retornos esperados podem performar bem em um portfólio com uma tendência comum predominante quando bem calibradas.

Dentre as estratégias baseadas em retornos esperados, os melhores resultados são obtidos pelo VECM com  $q = N - r$  vetores de integração. Conclui-se que as previsões feitas por um modelo mais completo (com um número maior de vetores de cointegração) são as mais adequadas. No entanto, o uso de todos estes vetores para identificar oportunidades de arbitragem pela estratégia (1) de valores extremos de  $S_t$  não se mostrou eficiente.

**Tabela 3**  
Resumo das simulações no portfólio de um fator

Estratégia	Vet. de cointegração	Parâmetros	Retorno total	Retorno a.a.	Vol a.a.	Custos a.a.	Transações a.a.
CDI	-	-	0.613	0.123	0.010	-	-
IBOVESPA	-	-	-0.361	-0.074	0.232	-	-
(1)	$q = 1$	$k = 30$	0.694	0.144	0.134	0.069	40.4
	$q = 7$	$k = 30$	0.597	0.124	0.280	0.074	28.8
	$l = 30, hf = 1$	$l = 30$	-0.137	-0.028	0.238	0.411	252.0
	$l = 50, hf = 1$	$l = 50$	-0.498	-0.103	0.173	0.309	24.8
	$l = 30, hf = 10$	$l = 30$	0.043	0.009	0.233	0.062	24.8
	$l = 50, hf = 10$	$l = 50$	-0.178	-0.037	0.179	0.058	25.1
	$l = 30, hf = 30$	$l = 30$	-0.237	-0.049	0.238	0.043	8.1
	$l = 50, hf = 30$	$l = 50$	-0.402	-0.083	0.180	0.042	8.1
(2)	$l = 30, hf = 1$	$l = 30$	0.270	0.056	0.241	0.298	252.0
	$l = 50, hf = 1$	$l = 50$	0.173	0.036	0.173	0.221	252.0
	$l = 30, hf = 10$	$l = 30$	1.118	0.232	0.246	0.058	24.8
	$l = 50, hf = 10$	$l = 50$	0.533	0.110	0.183	0.049	25.1
	$l = 30, hf = 30$	$l = 30$	-0.184	-0.038	0.254	0.046	8.1
	$l = 50, hf = 30$	$l = 50$	0.270	0.056	0.184	0.038	8.1
	$max\ width =$						
	$l = 30, hf = 1$	$l = 30$	-2.223	-0.460	0.278	0.546	252.0
	$l = 50, hf = 1$	$l = 50$	-2.408	-0.499	0.222	0.488	252.0
	$l = 30, hf = 10$	$l = 30$	0.484	0.100	0.284	0.075	25.1
	$l = 50, hf = 10$	$l = 50$	-0.239	-0.049	0.219	0.071	25.1
	$l = 30, hf = 30$	$l = 30$	0.393	0.081	0.267	0.043	8.3
	$l = 50, hf = 30$	$l = 50$	0.153	0.032	0.207	0.043	8.3
(3)	$l = 30, hf = 1$	$l = 30$	-3.457	-0.716	0.277	0.510	252.0
	$l = 50, hf = 1$	$l = 50$	-3.315	-0.687	0.223	0.464	252.0
	$l = 30, hf = 10$	$l = 30$	0.743	0.154	0.297	0.084	25.1
	$l = 50, hf = 10$	$l = 50$	0.455	0.094	0.224	0.075	25.1
	$l = 30, hf = 30$	$l = 30$	0.242	0.050	0.275	0.054	8.3
	$l = 50, hf = 30$	$l = 50$	0.807	0.167	0.226	0.049	8.3
	$hf = 1$						
	$hf = 10$	$hf = 10$	-0.033	-0.007	0.248	0.414	252.0
	$hf = 30$	$hf = 30$	-0.376	-0.078	0.241	0.063	25.1
(4)	$hf = 1$	$hf = 1$	0.908	0.188	0.241	0.303	8.3
	$hf = 10$	$hf = 10$	0.771	0.160	0.249	0.058	25.1
	$hf = 30$	$hf = 30$	0.239	0.049	0.256	0.045	8.3

## Portfólio de três fatores latentes

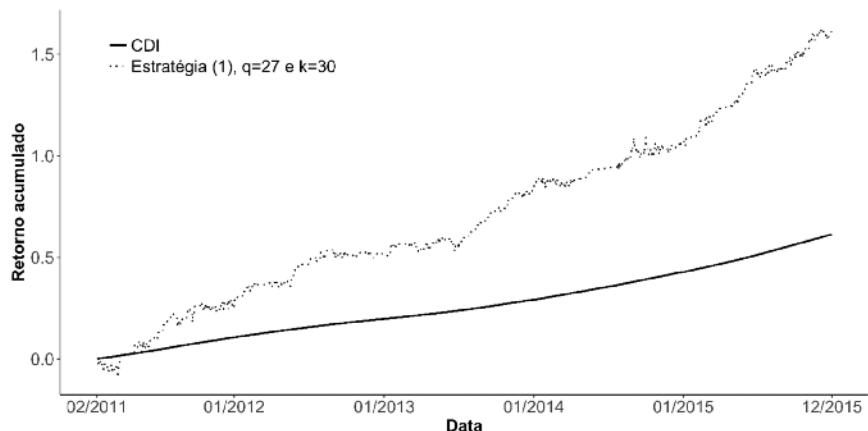
Ao contrário de técnicas usuais de *pairs trading* que procuram isolar os efeitos da cointegração em um grupo pequeno de ativos, com este portfólio a intenção era conseguir uma boa representação das principais tendências do mercado. Ele concentra uma maior quantidade de informação, já que é composto por um número grande de ativos ( $N = 30$ ), escolhidos de modo a representar de forma igualitária as três principais tendências detectadas no mercado. Detalhes de todas as estratégias são encontrados na Tabela 4, e o histórico da estratégia (1) comparado ao CDI é desenhado na Figura 4. Como era esperado, a estratégia (1) alcança seu potencial máximo ao utilizar os 27 vetores de cointegração. O desempenho excelente é evidente pelas 50 transações ano (aproximadamente uma transação a cada sete dias úteis), baixa volatilidade (em torno de 18%) e ganhos praticamente contínuos. O retorno total desta simulação é 2.6 vezes o CDI do mesmo período.

A estratégia (1) é viável inclusive considerando apenas um vetor de cointegração. Neste caso, o retorno total foi de 47.8% (78% do CDI) e a volatilidade anualizada foi de apenas 7.9%. No entanto, este portfólio permite que mais vetores de cointegração sejam corretamente estimados e à medida que eles são incluídos no modelo, a performance da estratégia vai aumentando. Assim, nota-se uma diferença em relação ao portfólio de um fator, onde não há vantagem em tentar estimar mais vetores. Em poucas palavras, o uso de mais informação do mercado no portfólio mais completo permite criar uma estratégia altamente rentável e de baixa volatilidade.

A performance das estratégias que dependem de retornos esperados, por outro lado, ficam aquém da estratégia (1). A maior parte das realizações apresentou retorno negativo, e apenas uma realização, referente à estratégia (3), gerou retornos acima do CDI. Novamente a baixa performance está relacionada ao alto número de transações e o alto custo de negociação que isso gera.

**Figura 4**

Retornos da estratégia (1) no portfólio de três fatores para as diferentes parametrizações.



**Tabela 4**  
Resumo das simulações no portfólio de três fatores

Estratégia	Vet. de cointegração	Parâmetros	Retorno total	Retorno a.a.	Vol a.a.	Custos a.a.	Transações a.a.
CDI	-	-	0.613	0.123	0.010	-	-
IBOVESPA	-	-	-0.361	-0.074	0.232	-	-
(1)	$q = 1$	$k = 30$	0.478	0.099	0.079	0.140	86.1
	$q = 27$	$k = 30$	1.633	0.338	0.182	0.083	48.9
	$l = 30, hf = 1$	$l = 30, hf = 1$	0.140	0.029	0.199	0.470	252.0
	$l = 50, hf = 1$	$l = 50, hf = 1$	-0.045	-0.009	0.146	0.348	252.0
	$l = 30, hf = 10$	$l = 30, hf = 10$	0.295	0.061	0.204	0.082	25.1
	$l = 50, hf = 10$	$l = 50, hf = 10$	0.302	0.062	0.152	0.070	24.8
	$l = 30, hf = 30$	$l = 30, hf = 30$	-0.261	-0.054	0.203	0.057	8.1
	$l = 50, hf = 30$	$l = 50, hf = 30$	-0.175	-0.036	0.148	0.050	8.1
(2)	$l = 30, hf = 1$	$l = 30, hf = 1$	-0.078	-0.016	0.201	0.337	252.0
	$l = 50, hf = 1$	$l = 50, hf = 1$	-0.196	-0.041	0.147	0.250	252.0
	$l = 30, hf = 10$	$l = 30, hf = 10$	0.098	0.020	0.210	0.073	24.8
	$l = 50, hf = 10$	$l = 50, hf = 10$	-0.122	-0.025	0.154	0.063	24.8
	$l = 30, hf = 30$	$l = 30, hf = 30$	-0.458	-0.095	0.203	0.055	8.3
	$l = 50, hf = 30$	$l = 50, hf = 30$	-0.625	-0.129	0.150	0.050	8.1
	$max\ width =$						
	$l = 30, hf = 1$	$l = 30, hf = 1$	-2.799	-0.580	0.242	0.572	252.0
	$l = 50, hf = 1$	$l = 50, hf = 1$	-2.518	-0.521	0.187	0.521	252.0
	$l = 30, hf = 10$	$l = 30, hf = 10$	-0.990	-0.205	0.226	0.096	25.1
	$l = 50, hf = 10$	$l = 50, hf = 10$	-1.181	-0.245	0.169	0.088	25.1
	$l = 30, hf = 30$	$l = 30, hf = 30$	-0.010	0.002	0.210	0.061	8.3
(3)	$l = 50, hf = 30$	$l = 50, hf = 30$	-0.611	-0.126	0.164	0.056	8.3
	$l = 30, hf = 1$	$l = 30, hf = 1$	-3.567	-0.739	0.242	0.555	252.0
	$l = 50, hf = 1$	$l = 50, hf = 1$	-3.143	-0.651	0.185	0.510	252.0
	$l = 30, hf = 10$	$l = 30, hf = 10$	-0.641	-0.133	0.245	0.092	25.1
	$l = 50, hf = 10$	$l = 50, hf = 10$	-0.937	-0.194	0.173	0.087	25.1
	$l = 30, hf = 30$	$l = 30, hf = 30$	-0.754	-0.156	0.253	0.057	8.3
	$l = 50, hf = 30$	$l = 50, hf = 30$	-0.102	-0.021	0.175	0.055	8.3
	$hf = 1$	$hf = 1$	0.307	0.064	0.209	0.473	252.0
(4)	$hf = 10$	$hf = 10$	0.501	0.104	0.213	0.083	25.1
	$hf = 30$	$hf = 30$	-0.158	-0.033	0.207	0.058	8.3
	$hf = 1$	$hf = 1$	0.184	0.038	0.215	0.345	252.0
	$hf = 10$	$hf = 10$	-0.257	-0.053	0.229	0.074	25.1
	$hf = 30$	$hf = 30$	-0.350	-0.073	0.211	0.054	8.3

### 5.1 Teste de arbitragem estatística

Para determinar se houve arbitragem estatística nas estratégias testadas foi utilizado o teste descrito em Jarrow et al. (2012), daqui em diante chamado de JTTW. De acordo com a definição especificada, para se ter arbitragem estatística os seguintes itens devem ser conjuntamente verdadeiros:

1.  $v(0) = 0$
2.  $\lim_{t \rightarrow \infty} E^P[v(t)] > 0$
3.  $\lim_{t \rightarrow \infty} P(v(t) < 0) = 0$
4.  $\lim_{t \rightarrow \infty} \text{Var}[\Delta v(t)|v(t) < 0] = 0$ .

Em síntese, estes itens determinam que a estratégia deve ter custo inicial zero, esperança de retornos acumulados positiva, e uma probabilidade de perda e variância ponderada no tempo convergindo a zero. Para formular o teste, dada uma estratégia de *trading* com ganhos totais acumulados  $v_t$ , calcula-se os incrementos  $\Delta v_t = v_t - v_{t-1}$  para um horizonte  $[t-1; t]$  fixo, e são trazidos a valor presente pela taxa livre de risco. Assume-se que os incrementos dos retornos podem ser expressados por um processo da forma

$$\Delta v_t = \mu t^\theta + \sigma t^\lambda z_t, \quad (2)$$

sendo  $z_t \sim N(0, 1)$  *i.i.d.*. Os parâmetros  $\mu$  e  $\theta$  determinam a esperança dos incrementos enquanto os parâmetros  $\sigma$  e  $\lambda$  definem a variância do processo. Nota-se que o processo é não estacionário se  $\theta \neq 0$  ou  $\lambda \neq 0$ . Já que a maioria das simulações apresentou valores de  $\theta$  não significativos, aqui será utilizado o caso que os autores do teste denominam de média restrita (CM), onde  $\theta = 0$ .

A hipótese a ser testada é  $H_0$ : Não há arbitragem estatística. Para rejeitar  $H_0$  e concluir que há arbitragem estatística precisam ocorrer, simultaneamente:

1.  $R_1 : \mu \geq 0$
2.  $R_2 : -\lambda > 0$  or  $\theta - \lambda > 0$
3.  $R_3 : \theta - \lambda + 1/2 > 0$  and
4.  $R_4 : \theta + 1 > 0$

Como o teste deve ser feito sob uma união de hipóteses, a menor

das estatísticas t serve como critério de rejeição. Por isso, o teste utiliza a estatística Min-t definida por

$$\text{Min-t} = \min \left\{ t(\hat{\mu}), t(\hat{\theta} - \hat{\lambda} + 1/2), t(\hat{\theta} + 1), \max \left[ t(-\hat{\lambda}), t(\hat{\theta} - \hat{\lambda}) \right] \right\}. \quad (3)$$

Os parâmetros são estimados por máxima verossimilhança e o t-crítico para um determinado nível de significância  $\alpha$ , é obtido através de simulações Monte Carlo. Para a amostra deste estudo, o t-crítico para um nível de significância de 5% é 0.827.

No portfólio de um fator latente, nenhuma simulação apresentou arbitragem estatística. No portfólio de três fatores latentes, apenas uma realização, a estratégia (1) com 27 vetores de cointegração e  $k = 30$ , apresentou arbitragem estatística com um valor de Min-t de 4.235.

## 6. Considerações finais

O principal objetivo deste trabalho é identificar falhas de precificação nos ativos de um portfólio, e propor uma estratégia *long-short* que aproveite estas possíveis oportunidades de arbitragem. O foco da estratégia é gerar sinais de *trade* quando os ativos se desviam do seu equilíbrio de longo prazo, e capturar retornos quando os ativos retornam ao dito equilíbrio. A estratégia foi simulada em dois portfólios que visam representar as principais tendências do mercado e os resultados foram analisados em função do retornos, volatilidade, custos de operação e existência de arbitragem estatística. Para fins de comparação, também foram simuladas outras três estratégias *long-short*, onde os pesos de alocação são determinados pelos retornos esperados.

Mesmo considerando que o mercado brasileiro possui uma taxa de juros extremamente elevada para os padrões internacionais, a magnitude dos retornos obtidos nas estratégias simuladas é muito atraente, ainda mais ao considerar que se trata de um investimento com alavancagem. A estratégia proposta destaca-se de todas as outras pela excelente performance e por apresentar arbitragem estatística pelo teste JTTW. Os resultados ótimos e com menor volatilidade são obtidos no portfólio com uma boa representatividade de todas as tendências do mercado.

O ponto de atenção em relação a esta estratégia é a imprevisibilidade da frequência de negociação. Como tomar e desfazer as posições

não dependem de um prazo fixo, a estrutura do portfólio e o momento do mercado podem fazer com que uma posição demore para ser desfeita. Para contornar esse problema, medidas de *stop-loss* e *stop-gain* por exemplo, podem ser implementadas para agilizar as negociações. Além disso, assim como todas as estratégias *long-short*, é importante identificar cuidadosamente os ativos para montar portfólios cointegrados.

Os resultados obtidos são promissores dentro do escopo de gestão de portfólios, visto que todas as estratégias testadas são capazes de agregar valor para o investidor a um baixo custo de capital próprio. A estratégia sugerida identifica, com sucesso, oportunidades de arbitragem e tem a vantagem de somente assumir risco nestes momentos. Já as estratégias baseadas em retornos esperados são autofinanciáveis e de fácil monitoramento, dado seu cronograma fixo de negociações. O uso de VECM mostrou-se adequado para os fins deste trabalho, podendo modelar as séries de preços e logpreços e ainda gerar boas estimativas dos vetores de cointegração e retornos esperados. Os resultados podem ser aperfeiçoados com o uso de outras técnicas de seleção de portfólio e determinação de tendências além de PCA, como recentes extensões de modelos de fatores dinâmicos para séries cointegradas, ou modelos de espaço estado.

## Referências

- Alexander, C. (1999). Optimal hedging using cointegration. *Philosophical Transaction of The Royal Society A*, 357:2039–2058.
- Alexander, C. and Dimitriu, A. (2002). The cointegration alpha: Enhanced index tracking and long-short equity market neutral strategies. *The Business School for Financial Markets*.
- Alexander, C. and Dimitriu, A. (2004a). A comparison of cointegration and tracking error models for mutual funds and hedge funds. *ISMA Centre Discussion Papers in Finance*.
- Alexander, C. and Dimitriu, A. (2004b). Sources of outperformance in equity markets. *The Journal of Portfolio Management*, 30(4):170–185.
- Alexander, C. and Dimitriu, A. (2005a). Indexing and statistical arbitrage. *The Journal of Portfolio Management*, 31(2):50–63.

Alexander, C. and Dimitriu, A. (2005b). Indexing, cointegration and equity market regimes. *International Journal of Finance and Economics*, 10(3):213–231.

Amengual, D. and Watson, M. W. (2006). Consistent estimation of the number of dynamic factors in a large n and t panel. Master's thesis, Princeton University.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1).

Caldeira, J. F. and Moura, G. V. (2013). Seleção de uma carteira de pares de ações usando cointegração: uma estratégia de arbitragem estatística. *Revista Brasileira de Finanças*, 11(1):49–80.

Caldeira, J. F. and Portugal, M. S. (2010). Estratégia long-short, neutra ao mercado, e index tracking baseadas em portfólios cointegrados. *Revista Brasileira de Finanças*, 8(4):469–504.

Chang, Y., Jiang, B., and Park, J. Y. (2008). Using Kalman filter to extract and test for common stochastic trends. In *Midwest Econometrics Conference*.

Chang, Y., Miller, J. I., and Park, J. Y. (2005). Extracting a common stochastic trend: Theory with some applications. *Symposium on Econometric Theory and Applications*, 150(2):231–247.

Elliott, R. J., Hoek, J. V. D., and Malcolm, W. P. (2005). Pairs trading. *Quantitative Finance*, 5(3):271–276.

Focardi, S. M., Fabozzi, F. J., and Mitov, I. K. (2016). A new approach to statistical arbitrage: Strategies based on dynamic factor models of prices and their performance. *Journal of Banking & Finance*, 65:134–155.

Gregoriou, G. N., Hübner, G., Papageorgiou, N., and Rouah, F. D. (2011). *Hedge Funds: Insights in Performance Measurement, Risk Analysis, and Portfolio Allocation*. John Wiley & Sons.

- Harris, D. (1997). Principal components analysis of cointegrated time series. *Econometric Theory*, 13(4):529–557.
- Jarrow, R., Teo, M., Tse, Y. K., and Warachka, M. (2004). Testing market efficiency using statistical arbitrage with applications to momentum and value trading strategies. *Journal of Economics*, 73(3):525–565.
- Jarrow, R., Teo, M., Tse, Y. K., and Warachka, M. (2012). An improved test for statistical arbitrage. *Journal of Financial Markets*, 15(15):47–80. An improved test for statistical arbitrage.
- Peña, D. and Poncela, P. (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, 136(4):1237–1257.
- Pucciarelli, A. J. (2014). Estratégia de cointegração dinâmica empírica para arbitragem estatística e trading. Escola de Economia de São Paulo, Fundação Getúlio Vargas, São Paulo. Dissertação de Mestrado.
- Roll, R. (1992). A mean/variance analysis of tracking error. *The Journal of Portfolio Management*, 18(4):13–22.
- Rudd, A. (1980). The “market model” in investment management. *The Journal of Finance*, 35(2):597–607.
- Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association*, 83:1097–1107.
- Triantafyllopoulos, K. and Montana, G. (2011). Dynamic modeling of mean-reverting spreads for statistical arbitrage. *Computational Management Science*, 8(1):23–49.