

Rodrigo Yabunaka

**Mínimos quadrados flexíveis aplicado em data mining e arbitragem
estatística no mercado brasileiro**

Rodrigo Yabunaka

**Mínimos quadrados flexíveis aplicado em data mining e arbitragem
estatística no mercado brasileiro**

Dissertação apresentada à
escola Politécnica da
universidade de São Paulo
para conclusão do curso de
MBA em engenharia
financeira.

Área de Concentração: MBA
de engenharia financeira

Orientador: Prof. Oswaldo Luiz
do Valle Costa

São Paulo

2011

RESUMO

Em algoritmos de negociação, um grande número de fluxos de dados deve ser analisado a cada momento assim que um novo dado se torne disponível, gerando assim um grande esforço computacional. Através do estudo da correlação e dos padrões dos fluxos de dados pode se obter oportunidades para operações de arbitragem. Baseado nessa crença este trabalho tem como objetivo demonstrar uma aplicação de uma arbitragem estatística através do método dos mínimos quadrados flexíveis (MQF) e de sua equivalência com o filtro de Kalman aplicado em um experimento utilizando ativos listados na BM&F Bovespa e obtendo resultados promissores.

Palavras-chave: Arbitragem estatística, Filtro de Kalman, mínimos quadrados flexíveis algoritmo de negociação.

ABSTRACT

In trading algorithms, a large number of data streams have to be analyzed each time so that a new data is available, thus generating a large computational effort. It's possible to get opportunities for arbitrage operations from the study of patterns and correlation of data streams. Based on this belief this work aims to demonstrate a statistical arbitrage application using the method of flexible least squares (FLS) and their equivalence with the Kalman filter applied to an experiment using the assets listed on the BM&F Bovespa and getting promising results.

Keywords: Statistical arbitrage, Kalman filter, flexible least squares, algorithmic trading.

LISTA DE ILUSTRAÇÕES

Figura 1 - Comparação das cotações da AMBEV 3 e AMBEV4.....	10
Figura 2- Evolução dos retornos da estratégia	29
Figura 3 - Evolução do desempenho das estratégias	30

LISTA DE TABELAS

Tabela 1 - Comparação do Índice de Sharpe.....	28
--	----

Sumário

1 Introdução	7
2 Estratégia de arbitragem estatística	9
3 Mínimos quadrados flexíveis	12
4 Mínimos quadrados flexíveis e Filtro de Kalman	16
4.1 Filtro de Kalman	16
4.2 Aplicando o MQF ao Filtro de Kalman.....	19
5 Redução de dimensões.....	23
6 Sistema de negociação	25
6.1 Dados.....	25
6.2 Regra de trading.....	26
6.3 Implementação	27
7 Conclusão	31
8 Referências	32

1 Introdução

Nos últimos anos tem se visto um aumento da necessidade por parte dos bancos, clubes, fundos de investimentos, e corretoras de valores, para se automatizar as operações dos sistemas de negociações. No mercado americano e europeu já esta cada vez mais consolidado o uso de sistemas autônomos de negociação ou *algorithmic trading systems* (ATS) que efetuam operações nos mercados com pouca ou nenhuma intervenção humana (para mais detalhes consulte Bloomberg... (2010)).

Nas economias emergentes, devido à atual situação econômica mundial, há um aumento da demanda destes sistemas e principalmente estratégias de investimento que tirem o máximo proveito das particularidades de cada mercado como é caso do Brasil e outros países emergentes, para mais informações consulte High...(2011). Isto é feito através de sofisticadas plataformas de negociação baseados em modelos preditivos que servem de base para estratégias de investimento especulativo.

Em algoritmos de negociação, um grande número de fluxos de dados deve ser analisado a cada momento assim que um novo dado se torne disponível, gerando assim um grande esforço computacional. Um problema recorrente está em identificar dependências entre fluxos de dados dependentes do tempo, isso pode ser modelado como um problema de regressão. Desta maneira um sistema automatizado de negociação deve ser capaz de quantificar em que medida um fluxo depende de outros fluxos de cotação em um dado ponto no tempo.

A cada ponto no tempo, assume se que o fluxo do ativo alvo depende de forma linear de uma série de outros fluxos, porém os coeficientes dos modelos de regressão podem mudar de maneira suave ao longo do tempo.

O objetivo deste trabalho é demonstrar uma solução possível para uma estratégia de negociação conhecida como *pair trading*, onde o foco é encontrar disparidades no comportamento de ativos listados em bolsa e a fim de executar uma operação de arbitragem.

Este trabalho está organizado da seguinte maneira. No capítulo 2 trataremos brevemente dos tipos e de como funcionam as estratégias de arbitragem estatística, assim como da motivação para o método proposto. No capítulo 3 será apresentado o método dos mínimos quadrados flexíveis (MQF), um modelo de regressão linear que não requer uma suposição probabilística para seu uso, onde tal característica é muito útil para o propósito deste trabalho. No capítulo 4 será apresentada uma conexão entre o MQF e as equações do filtro de Kalman que levará ao modelo que será implementado. Será tratada no capítulo 5 uma forma de extrair os componentes principais a partir de um grande número de fluxos de cotação, complementando o MQF na obtenção no algoritmo da estratégia de investimento. A implementação da arbitragem estatística e bem como seu resultado será apresentado no capítulo 6. Por último, considerações finais serão tratadas no capítulo 7.

2 Estratégia de arbitragem estatística

Este trabalho terá foco na arbitragem estatística, uma gama de estratégias que tem como principal característica a procura por disparidades de preços de um ou mais ativos baseados no valor esperado destes ativos, tais estratégias são amplamente utilizadas em bancos, *Assets*, em fundos de investimentos e *traders* independentes.

O exemplo mais simples de arbitragem estatística é uma estratégia de investimento conhecida como *pair trading*, onde um o investidor deve escolher dois ativos que tendem a se mover de forma correlacionada ao longo de um grande período. Desta forma o investidor pode obter oportunidades com as disparidades momentâneas dos preços destes ativos, posicionando se comprado no ativo em baixa e vendido no ativo que estiver sobrevalorizado, e revertendo esta posição quando ambos os ativos voltarem a um estado próximo ou igual ao original. Este tipo de estratégia tenta capturar a tendência que o mercado irá seguir e isto é feito através da identificação de padrões presentes nas séries de preços.

As tendências dos preços de ativos são comumente relacionadas com correlações nas variações de suas séries, uma tendência é uma série de preços de ativos que se movimenta constantemente em uma direção ao longo de um determinado intervalo de tempo, onde as séries exibem uma correlação positiva.

A figura 1 mostra um exemplo desta estratégia. Onde vemos a evolução histórica dos preços de dois ativos, a AMBEV3 e a AMBEV4 denotamos x_t e y_t respectivamente onde $t = 1, 2, 3 \dots$, note que no período apresentado as duas séries de preços exibem uma dependência. Esta estratégia poderia ser implementada em um sistema automatizado de negociação para tirar proveito das divergências temporárias antes que os preços dos ativos voltem para o seu estado de equilíbrio. O algoritmo para o *spread* é pode ser descrito como $y_t - x_t = s_t$. Por exemplo, se o valor de s_t em um dado momento t for maior que uma constante c , assume se que o ativo AMBEV3 está sobre valorizado e o investidor ou um sistema automatizado, deverá se posicionar de maneira comprada em AMBEV4 e vendido em AMBEV3, o lucro será obtido quando estas ações reverterem à média. Embora uma relação estável entre dois ativos possa persistir por algum tempo, esta pode desaparecer de

repente ou apresentar-se em diferentes padrões, como padrões periódicos ou uma tendência que pode ser capturado através da implementação de outros mais modelos refinados.

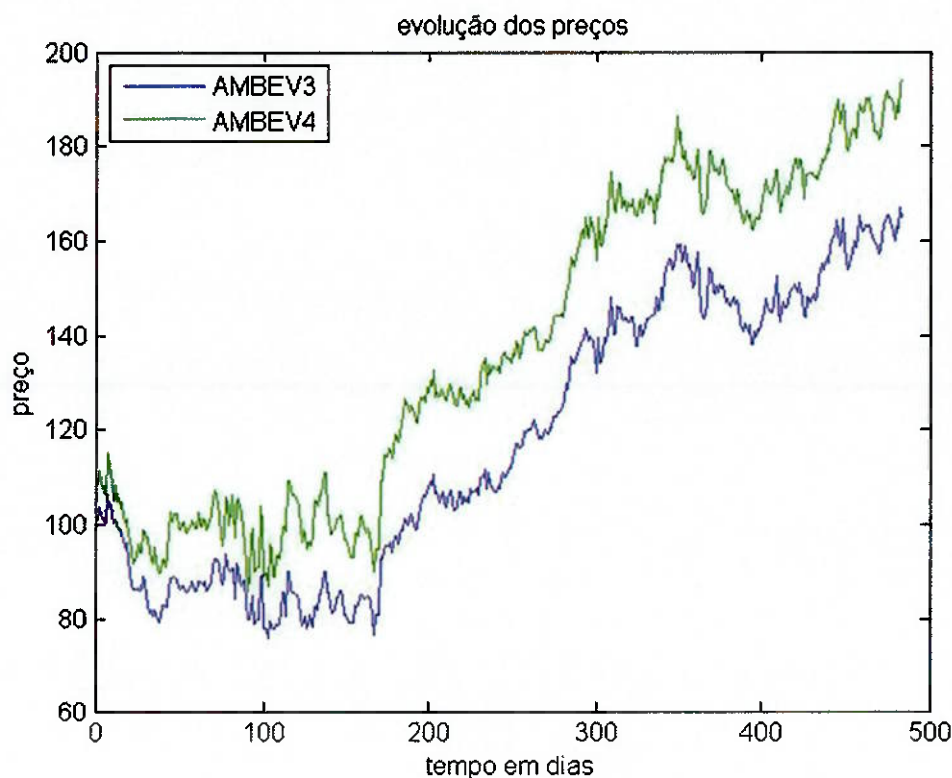


Figura 1 - Comparação das cotações da AMBEV3 e AMBEV4

As oportunidades para o uso de uma estratégia de *pair trading* como descrito acima são limitadas, pois, dependem de ativos que possuem correlação positiva, próxima a 1. Para este problema foram criadas ao longo de dois anos extensões e variações baseadas neste tipo de estratégia. Por exemplo, arbitragem de *basket*, também conhecida como arbitragem de índice, onde o investidor explora as discrepâncias temporárias entre um índice e uma carteira formada com os ativos formadores desse índice. No mercado brasileiro essa estratégia pode ser feita comprando o contrato IBOV futuro e vendendo a carteira com os papéis das 50 empresas formadoras do IBOV nas devidas proporções de cada papel participante do índice ou vice-versa.

Neste trabalho será abordada uma versão da estratégia de *pair trading* onde serão exploradas as discrepâncias temporárias de preços entre um dado ativo escolhido pelo investidor e outro ativo artificial que reproduz o ativo escolhido. O ativo artificial

é obtido pela combinação linear de um conjunto de preços no qual assumimos ter correlação com o ativo escolhido.

Segundo Montana et al.(2008), a justificativa desta abordagem parte do princípio de que se existe uma forte associação entre o ativo artificial e o ativo escolhido dure por um longo período de tempo esta associação sugere que os ativos reagem a algum componente, e não observado, subjacente de risco que explica essa dinâmica que pode incluir todas as fontes de risco relacionadas ao mercado. O objetivo dessa abordagem é minimizar todas as fontes de risco relacionadas ao mercado e obter um fluxo de dados que representa um o risco-alvo específico.

Por exemplo, suponha que y_t seja o valor de um ativo em um momento t , e \hat{y}_t seja o valor do ativo artificial calculado a partir de um conjunto de preços no mesmo momento t . Até o momento o ativo artificial pode ser interpretado como o “preço justo” do ativo escolhido, dada toda a informação disponível e condições de mercado. A diferença entre y_t e \hat{y}_t representa o risco associado ao ativo escolhido, ou seja, o descasamento do preço, e se este comportamento contenha um padrão previsível é possível explorar este comportamento em um investimento especulativo. A estratégia depende fundamentalmente da forma e da precisão de como se estima o valor do ativo artificial, isto será visto a seguir através do estudo dos mínimos quadrados flexíveis em uma abordagem apresentada por Montana et al.(2008).

3 Mínimos quadrados flexíveis

Um modelo de regressão linear padrão contém uma variável de resposta y_t e p variáveis de predição x_1, \dots, x_p , que formam um vetor coluna $x_t = (x_{1t}, \dots, x_{pt})'$. O modelo nos diz que y_t pode ser aproximado por $x_t\beta$, onde β é um vetor de dimensão p dos parâmetros de regressão. A regressão dos mínimos quadrados estima $\hat{\beta}$ do vetor de parâmetros são encontrados como os valores que minimiza a de função de custo

$$C(\beta) = \sum_{t=1}^T (y_t - x_t'\beta)^2 \quad (1)$$

se a variável de resposta y_t e o vetor de predição x_t são observações em um momento t de um conjunto de fluxos de dados, é possível que a dependência entre estas duas variáveis mude e evolua de maneira dinâmica ao longo do tempo. O método dos mínimos quadrados flexíveis (MQF) foi introduzido nos anos 80 por Kalaba e Tesfatsion (1989) como uma generalização do modelo de regressão linear descrito acima que permite o uso de variáveis dependentes do tempo em seus coeficientes. Adicionando a suposição básica da regressão

$$(y_t - x_t'\beta_t) \approx 0 \quad (2)$$

o MQF também postula que

$$\beta_{t+1} - \beta_t \approx 0 \quad (3)$$

desta forma os coeficientes irão evoluir de maneira lenta e gradual ao longo do tempo.

Uma importante característica do MQF é a não necessidade de se especificar as probabilidades para o erro residual da form. (2), e este aspecto é muito útil para casos onde normalmente não seja possível especificar um modelo preciso para descrever os erros e que seja valido em todos os momentos, como é o caso dos fluxos de cotação. De acordo com Montana et al.(2008) o modelo do MQF funciona bem mesmo quando suposição da form.(3) é violada, em mudanças grandes e repentinas de β_{t-1} para β_t , para alguns momentos de t .

O método MQF consiste em minimizar uma versão penalizada da função de custo da regressão dos mínimos quadrados

$$C(\beta, \mu) = \sum_{t=1}^T (y_t - x_t' \beta_t)^2 + \mu \sum_{t=1}^{T-1} \xi_t \quad (4)$$

onde

$$\xi_t = (\beta_{t+1} - \beta_t)' (\beta_{t+1} - \beta_t) \quad (5)$$

e μ é um valor escalar maior que 0 a ser definido.

No algoritmo original proposto por Kalaba e Tesfatsion (1988), a minimização do para cada β_t é feita de forma seqüencial, onde todos os pontos de dados estão armazenados em uma base de dados e podem ser acessados prontamente. Na seqüência será visto o núcleo dessa abordagem.

De acordo com Montana et al.(2008) o menor custo do processo de estimação no tempo t pode ser escrito recursivamente como

$$c(\beta_{t+1}; \mu) = \inf_{\beta_t} \{ (y_t - x_t' \beta_t)^2 + \mu \xi_t + c(\beta_t; \mu) \} \quad (6)$$

além disso, este custo é assumido para ter uma forma quadrática

$$c(\beta_t; \mu) = \beta_t' S_{t-1} \beta_t - 2\beta_t' s_{t-1} + r_{t-1} \quad (7)$$

onde S_{t-1} e s_{t-1} têm dimensões $p \times p$ e $p \times 1$, respectivamente, e r_{t-1} é uma variável escalar. Substituindo a eq.(7) na eq.(6) e, em seguida, derivando a função de custo da eq.(6) respeitando se β_t e condicionado à β_{t+1} , obtém-se uma equação recursiva de atualização para o coeficiente de regressão variável no tempo

$$\hat{\beta}_t = d_t + M_t \beta_{t+1} \quad (8)$$

onde

$$d_t = \mu^{-1} M_t (s_{t-1} + x_t y_t)$$

$$M_t = \mu (s_{t-1} + \mu I_p + x_t x_t')^{-1}$$

lembrando que as recursões são iniciadas com os valores de S_0 e s_0 com algum valor. Agora, segundo Montana et al.(2008), usando a eq.(8), a função de custo pode ser escrito como

$$c(\beta_{t+1}; \mu) = \beta'_{t+1} S_{t+1} - 2\beta'_{t+1} s_t + r_t$$

onde

$$S_t = \mu(I_p + M_t) \quad (9)$$

$$s_t = \mu d_t \quad (10)$$

$$r_t = r_{t-1} + y_t^2 - (s_{t-1} + x_t y_t)' d_t$$

e onde I_p é uma matriz identidade $p \times p$. Para aplicar a eq.(8), este procedimento exige que todos os pontos de dados até T devem estar disponíveis, de modo que o vetor de coeficientes β_T deve ser calculado primeiro. Os trabalhos de Kalaba e Tesfatsion (1988) e Montana et al.(2008) mostram que o calculo da estimativa $\hat{\beta}_T$ pode ser obtida seqüencialmente como

$$\hat{\beta}_T = (S_{T-1} + x_T x_T')^{-1} (s_{T-1} + x_T y_T)$$

posteriormente, a eq.(8) pode ser usada para estimar todos os vetores de coeficientes restantes $\beta_{T-1}, \dots, \beta_1$, indo para trás no tempo.

O procedimento baseia-se no ajuste do parâmetro μ , onde esta variável regula a suavizando a influência da função de custo definida em na eq.(4). O parâmetro μ é definido como

$$\mu = \frac{(1 - \delta)}{\delta}$$

onde δ é um valor escalar, assim, com um δ próximo a 0, corresponde a maiores valores de μ , dando maior peso para a função de custo. Por outro lado um valor maior de δ resulta em uma influência menor de μ para o MQF estimar $\hat{\beta}_t$.

Tendo interesse em situações em que os pontos de observação dos fluxos de dados chegam seqüencialmente, e não em situações em que se têm todos os pontos prontamente disponíveis. De acordo com Montana et al.(2008), cada variável x_t da matriz de dimensão p representa um ponto de observação do fluxo de preços no tempo. Neste contexto, usando o método dos mínimos quadrados flexíveis o calculo de $\hat{\beta}_t$ é dado pela seguinte equação recursiva,

$$\hat{\beta}_t = (S_{t-1} + x_t x_t')^{-1} (s_{t-1} + x_t y_t) \quad (11)$$

onde S_t e s_t são

$$S_t = \mu(S_{t-1} + \mu I_p + x_t x'_t)^{-1} (S_{t-1} + x_t x'_t) \quad (12)$$

$$s_t = \mu(S_{t-1} + \mu I_p + x_t x'_t)^{-1} (S_{t-1} + x_t y_t)$$

como descrito em Montana et al.(2008) este algoritmo tem uma boa precisão, retornando ao caminho dos seus coeficientes, tanto no caso onde todos os dados estão à disposição, como no caso aonde os dados chegam de forma seqüencial, e mesmo quando a premissa da form.(3) é quebrada devido a saltos abruptos nos valores dos fluxos de dados.

4 Mínimos quadrados flexíveis e Filtro de Kalman

O método MQF puro apresentado na última seção nos impõe alguns “desafios” tanto do ponto de vista do algoritmo como computacional, como estimar a matriz de covariância e inversão das matrizes das eq.(11) e eq.(12). Desta forma, se assumirmos que os coeficientes de regressão podem conter valores aleatórios sem perder o a flexibilidade e generalidade do método original.

Embora a hipótese geral exija apenas que a evolução dos coeficientes de regressão de $\beta_{t+1} - \beta_t$ seja pequena a cada passo e todas às vezes, não há nenhuma restrição ao fato de que cada vetor β_t possa ser um vetor randômico, tanto que no trabalho original de Kalaba e Tesfatsion (1988), β_t é tratado como uma seqüência de incógnitas a serem estimadas.

Com essa suposição em mãos, é possível fazer uma ligação algébrica entre as equações do método MQF e as equações do filtro de Kalman (FK), dando assim uma solução para os “desafios” descritos há pouco, além de dar um sentido para o parâmetro μ da função de custo. Neste capítulo será mostrado um método proposto por Montana et al.(2008) que conecta o MQF com FK.

4.1 Filtro de Kalman

O filtro de Kalman é um poderoso método para a estimação de β_t , onde o coeficiente de regressão em $t + 1$ é modelado como uma versão com ruído do coeficiente anterior no tempo t . Segundo Montana et al.(2008) foi introduzido um vetor aleatório ω_t com média zero e uma matriz de covariância V_ω , de modo que

$$\beta_{t+1} = \beta_t + \omega_t \quad t = 0, 1, \dots, T - 1$$

da mesma maneira, foi introduzida uma variável ϵ_t com média zero e variância V_ϵ ,

$$y_t = x'_t \beta_t + \epsilon_t \quad t = 0, 1, \dots, T - 1$$

considerando ambas as equações acima formam um modelo de estado- espaço no qual se assume que as séries $\{\omega_t\}$ e $\{\epsilon_t\}$ são mutuamente e individualmente não correlacionadas, ou seja ω_a não tem correlação com ω_b assim como ϵ_a não tem

correlação com ϵ_b , e ω_z não tem correlação com ω_c , para qualquer valor de z e c ou $a \neq b$. Ainda se assume que para todo t , ω_t , ϵ_t e β_0 inicial não são correlacionados. Nenhuma suposição sobre a distribuição de ϵ_t e ω_t é feita, assume se apenas que ela é desconhecida. Desta maneira, a única diferença entre o sistema especificado acima e o MQF é a suposição de aleatoriedade da β_t .

Para se estabelecer uma conexão entre o FK e o MQF, Montana et al.(2008) construiu uma forma alternativa do filtro de Kalman que não faz suposições sobre as distribuições de ϵ_t e ω_t , não se baseiam em suposições probabilísticas, e ainda não necessitando de inversões de matrizes para o calculo da estimação.

No tempo t , definimos $\hat{\beta}_t$ como o valor estimado de β_t e $\hat{y}_{t+1} = E(y_{t+1})$, o resultado do valor esperado de y_{t+1} , a variância de y_{t+1} é conhecida para cada momento t e dada por $Q_t = Var(y_{t+1})$, o erro para cada passo é definido como $e_t = y_t - E(y_t)$. A matriz de covariância de $\beta_t - \hat{\beta}_t$ como P_t e R_t é a matriz de covariância de $\beta_t - \hat{\beta}_{t-1}$, assim podemos escrever $P_t = Cov(\beta_t - \hat{\beta}_t)$ e $R_t = Cov(\beta_t - \hat{\beta}_{t-1})$ e assumindo a linearidade do sistema, podemos ver que, no tempo $t - 1$,

$$\begin{aligned} R_t &= P_{t-1} + V_\omega \\ \hat{y}_t &= x'_t \hat{\beta}_{t-1} \\ Q_t &= x'_t R_t x_t + V_\epsilon \end{aligned}$$

onde P_{t-1} e $\hat{\beta}_{t-1}$ assumimos como conhecidos, a equação recursiva FK atualizará P_t e $\hat{\beta}_t$ em função de P_{t-1} e $\hat{\beta}_{t-1}$.

Imagine que queremos obter um estimador de β_t que seja linear em y_t , que é $\hat{\beta}_t = a_t + K_t y_t$, para um a_t e K_t que serão definidos mais a diante Assim podemos escrever

$$\hat{\beta}_t = a_t^* + K_t e_t \quad (16)$$

onde

$$\begin{aligned} a_t^* &= \hat{\beta}_{t-1} \\ e_t &= y_t - x'_t \hat{\beta}_{t-1} \end{aligned}$$

desta maneira mostraremos que para um K_t , e se $\hat{\beta}_t$ é necessário para a seguinte minimização da soma dos quadrados

$$C = \sum_{t=1}^T (y_t - x_t' \beta_t)^2 \quad (17)$$

para provar isso, escrevemos

$$Y = (y_1, \dots, y_T)'$$

$$X = (x_1', \dots, x_T')'$$

$$B = (\beta_1', \dots, \beta_T')'$$

$$\varepsilon = (e_1, \dots, e_T)'$$

$$K = \begin{pmatrix} K_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & K_T \end{pmatrix}$$

assim podemos escrever a eq. (17) da seguinte forma

$$C \equiv C(B) = (Y - XB)'(Y - XB)$$

e $\hat{B} = A^* + K\varepsilon$, onde $A^* = ((a_t^*)', \dots, (a_T^*)')'$.

Mostraremos que $A^* = B^*$, onde $B^* = (\hat{\beta}_1', \dots, \hat{\beta}_T')'$. Com o \hat{B} acima exposto, a soma dos quadrados pode ser escrita como

$$\begin{aligned} S(\hat{B}) &= (Y - XA^* - XK\varepsilon)'(Y - XA^* - XK\varepsilon) \\ &= (Y - XA^*)'(Y - XA^*) - 2(Y - XA^*)'XK\varepsilon + \varepsilon'K'X'XK\varepsilon \end{aligned}$$

que é minimizado quando $Y - XA^*$ ou quando $0 = E(Y - XA^*)$, levando a $A^* = B^*$, assim, $a_t^* = \hat{\beta}_{t-1}$ da eq.(16) temos

$$\hat{\beta}_t = \hat{\beta}_{t-1} + K_t e_t \quad (18)$$

para um K_t (será definido mais a frente) e da definição de P_t temos:

$$\begin{aligned} P_t &= Cov(\beta_t - (\hat{\beta}_{t-1} + K_t(x_t' \beta_t + \varepsilon_t - x_t' \hat{\beta}_{t-1}))) \\ &= Cov((I_p - K_t x_t')(\beta_t - \hat{\beta}_{t-1}) - K_t \varepsilon_t) \\ &= (I_p - K_t x_t')(I_p - x_t K_t') R_t - V_\varepsilon K_t K_t' \\ &= R_t - K_t x_t' R_t - R_t x_t K_t' + Q_t K_t K_t' \end{aligned} \quad (19)$$

assim, podemos escolher um K_t que minimiza

$$E(\beta_t - \hat{\beta}_t)'(\beta_t - \hat{\beta}_t) \quad (20)$$

que é o mesmo que minimizar o traço de P_t , e, portanto, K_t é a solução da equação matricial

$$\frac{\partial \text{traço}(P_t)}{\partial K_t} = -2(x_t' R_t)' + 2Q_t K_t = 0$$

onde $\frac{\partial \text{traço}(P_t)}{\partial K_t}$ é a derivada parcial do vestígio de P_t , que respeita K_t . Resolvendo esta equação encontraremos $K_t = \frac{R_t x_t}{Q_t}$, onde a quantidade K_t , também conhecida como o “ganho de Kalman”, é ideal entre todos os estimadores lineares de $\hat{\beta}_t$ que minimiza a form.(20) junto com $K_t = \frac{R_t x_t}{Q_t}$, a partir da eq.(19) a mínima matriz de covariância P_t é

$$P_t = R_t - Q_t K_t K_t' \quad (21)$$

o filtro de Kalman é formado basicamente das equações eq.(18) e eq.(21), acompanhados de

$$\begin{aligned} K_t &= \frac{R_t x_t}{Q_t} \\ R_t &= P_{t-1} + V_\omega \\ Q_t &= x_t' R_t x_t + V_\epsilon \\ e_t &= y_t - x_t' \hat{\beta}_{t-1} \end{aligned}$$

os valores iniciais de $\hat{\beta}_0$ e P_0 devem ser dados (normalmente iniciando ambos com 0), note que da recursão de P_t e R_t temos

$$R_{t+1} = R_t - Q_t K_t K_t' + V_\omega \quad (22)$$

4.2 Aplicando o MQF ao Filtro de Kalman

De acordo com Montana et al.(2008), as equações do filtro de Kalman são normalmente derivadas considerando uma distribuição normal para ϵ_t e ω_t

respectivamente, está hipótese sobre a distribuição permite a derivação da função de verossimilhança, assim, quando a verossimilhança normal está disponível, podemos constatar que a sua maximização é equivalente a minimizar a quantidade

$$\sum_{t=1}^T (y_t - x_t' \beta_t)^2 + \frac{1}{V_\omega} \sum_{t=1}^{T-1} \xi_t$$

respeitando β_t, \dots, β_T onde ξ_t , definido na equação (5), a prova está em Jazwinski (1970), observando a expressão acima vemos que ela é a função de custo definida na eq.(4) apenas substituindo $\frac{1}{V_\omega}$ por μ .

Para o caso deste trabalho, e para um cenário mais geral, nenhuma distribuição pode ser assumida. Podemos chegar neste resultado, através de uma reorganização da eq.(11) na forma da eq.(18), que é o estimador FK de β_t , note que da eq.(12) podemos escrever

$$(S_{t-1} + x_t x_t')^{-1} = \mu S_t^{-1} (S_{t-1} + \mu I_p + x_t x_t')^{-1}$$

substituindo na eq.(11) nós temos $\hat{\beta}_t = S_t^{-1} s_t$, então teremos

$$\begin{aligned} \hat{\beta}_t - \hat{\beta}_{t-1} &= S_t^{-1} s_t - S_{t-1}^{-1} s_{t-1} \\ &= (S_{t-1} + x_t x_t')^{-1} (s_{t-1} + x_t y_t) - S_{t-1}^{-1} s_{t-1} \\ &= S_{t-1}^{-1} x_t y_t - \frac{S_{t-1}^{-1} x_t x_t' S_{t-1}^{-1} (s_{t-1} + x_t y_t)}{x_t' S_{t-1}^{-1} x_t + 1} \\ &= \frac{S_{t-1}^{-1} x_t}{x_t' S_{t-1}^{-1} x_t + 1} (y_t x_t' S_{t-1}^{-1} x_t + y_t - x_t' S_{t-1}^{-1} s_{t-1} - x_t' S_{t-1}^{-1} x_t y_t) \\ &= \frac{S_{t-1}^{-1} x_t}{x_t' S_{t-1}^{-1} x_t + 1} (y_t - x_t' \hat{\beta}_{t-1}) \\ &= K_t e_t \end{aligned}$$

onde

$$K_t = \frac{R_t x_t}{Q_t}$$

$$R_t = S_{t-1}^{-1}$$

$$Q_t = x_t' R_t x_t + 1$$

$$V_\epsilon = 1$$

resta provar que a recursividade de S_t como em (12) se comunica com a recursão da eq.(22), para $R_{t+1} = S_t^{-1}$, partindo se da eq.(12) e usando o lema de inversão de matriz, obtemos

$$\begin{aligned} R_{t+1} &= S_t^{-1} = \mu^{-1}(S_{t-1} + x_t x_t')^{-1}(S_{t-1} + \mu I_p + x_t x_t') \\ &= \mu^{-1}(I_p + \mu(S_{t-1} + x_t x_t')^{-1}) \\ &= \mu^{-1}I_p + (S_{t-1} + x_t x_t')^{-1} \\ &= S_{t-1}^{-1} - \frac{S_{t-1}^{-1}x_t x_t' S_{t-1}^{-1}}{x_t' S_{t-1}^{-1} x_t + 1} + \mu^{-1}I_p \\ &= R_t - Q_t K_t K_t' + V_\omega \end{aligned}$$

que é a recursão FK descrita na eq.(22), onde $V_\omega = \mu^{-1}I_p$.

Claramente, vemos que o $\hat{\beta}_t$ estimado pelo MQF visto na eq. (11) é o mesmo que o estimador $\hat{\beta}_t$ do filtro de Kalman na eq.(18), com isso temos

$$\text{Cov}(\beta_{t+1} - \beta_t) = \frac{1}{\mu} I_p$$

desta forma temos o papel do parâmetro de suavização μ definido na eq.(4), com μ tendendo ao infinito a matriz de covariância de $\beta_{t+1} - \beta_t$ tende ao infinito, o que significa que $\beta_{t+1} = \beta_t$, para todo t , reduzindo o modelo uma regressão com coeficientes constantes. Quando μ tende a zero, a matriz de covariância de $\beta_{t+1} - \beta_t$ terá uma diagonal com valores de muito elevados dos seus elementos fazendo com que os valores estimados por β_t flutuem erráticamente.

Um efeito importante da correspondência estabelecida entre a MQF e o FK do ponto de vista computacional, como dito antes, é que o filtro de Kalman não necessita de nenhuma inversão de matriz, para estimar cada $\hat{\beta}_t$ em qualquer t . Isso é muito útil, pois p pode ser muito grande.

De maneira resumida será apresentado abaixo todas formulas para a estimação de $\hat{\beta}_t$, adiantando um pouco da implementação descrita mais a diante, x_t é o vetor coluna dos retornos observados e y_t é o valor da cotação corrente do ativo alvo.

$$\hat{\beta}_t = K_t e_t + \hat{\beta}_{t-1}$$

$$K_t = \frac{S_{t-1}^{-1} x_t}{x_t' S_{t-1}^{-1} x_t + 1}$$

$$e_t = y_t - x_t' \hat{\beta}_{t-1}$$

$$Q_t = x_t' S_{t-1}^{-1} x_t + 1$$

$$S_t^{-1} = S_{t-1}^{-1} - Q_t K_t K_t' - \mu^{-1} I_p$$

5 Redução de dimensões

Os sistemas de negociações de alta frequência precisam lidar com uma grande quantidade de dados que chegam a todo o momento, atualmente em frações de segundo. E o tratamento da regressão linear deste grande fluxo de dados gera uma grande demanda computacional.

Se pudéssemos reduzir a necessidade de consumo dos fluxos de dados poderíamos ter um sensível aumento no processamento dos mesmos. Neste capítulo iremos ver uma solução para este problema, extraindo assim os componentes principais dos fluxos de cotação. A seguir veremos uma abordagem vista em Montana et al.(2008) do algoritmo de Weng et al.(2003).

De acordo com o algoritmo proposto por Weng et al.(2003) e considerando que $R_t = E(r_t r_t')$ é a matriz de covariância desconhecida de um conjunto de fluxos de dados disponível em um momento $t = (1, \dots, T)$, será mostrado um método para se estimar de maneira incremental o valor de R_{t+1} , em termos práticos, é possível extrair os fluxos de cotação que melhor representam o seu conjunto.

A seguir será mostrado em linhas gerais o algoritmo proposto por Weng et al.(2003). O autovetor g_t de R_t satisfaz a seguinte equação, onde λ_t é o autovalor.

$$h_t = \lambda_t g_t = R_t g_t \quad (23)$$

Onde \hat{h}_t é o valor estimado atual a partir de h_t que usa todos os dados até o tempo $t = (1, \dots, T)$, note que

$$\frac{h_t + \dots + h_T}{T} = \frac{1}{T} \sum_{i=1}^T R_i g_i$$

o valor estimado $\hat{h}_t = \frac{h_t + \dots + h_T}{T}$ e substituindo R_t por $r_t r_t'$ temos

$$\hat{h}_t = \frac{1}{t} \sum_{i=1}^t r_i r_i' g_i \quad (24)$$

até o momento vemos que \hat{h}_t é a média de $r_i r'_i g_i$. Observando a função vemos g_t pode ser estimado como $\frac{\hat{h}_{t-1}}{\|\hat{h}_{t-1}\|}$, desta maneira \hat{h}_0 deve ser ter o valor de r_0 na primeira iteração, substituindo na equação temos

$$h_t = \frac{1}{t} \sum_{i=1}^t r_i r'_i \frac{\hat{h}_{i-1}}{\|\hat{h}_{i-1}\|} \quad (25)$$

no cenário onde não se dispõe de todos os dados prontamente, mas sim chegando seqüencialmente, é necessário rearranjar a função para ela consiga trabalhar de maneira recursiva utilizando se apenas \hat{h}_{t-1} . Note que influência de valores antigos é controlada através dos pesos $\frac{t-1}{t}$ e $\frac{1}{t}$ para as estimativas de \hat{h}_t .

$$\hat{h}_t = \frac{1}{t} \sum_{i=1}^{t-1} r_i r'_i \frac{\hat{h}_{i-1}}{\|\hat{h}_{i-1}\|} + \frac{1}{t} r_t r'_t \frac{\hat{h}_{t-1}}{\|\hat{h}_{t-1}\|} = \frac{t-1}{t} \hat{h}_{t-1} + \frac{1}{t} r_t r'_t \frac{\hat{h}_{t-1}}{\|\hat{h}_{t-1}\|}$$

6 Sistema de negociação

Neste trabalho o experimento foi feito de duas maneiras, diferenciando apenas na quantidade de papéis (fluxos de cotação) usados para se construir o valor corrente do ativo artificial, a primeira utilizando se todos os ativos listados na Bovespa e a segunda utilizando apenas os componentes principais encontrados pelo método de Weng et al.(2003).

Ambas as simulações utilizaram a mesma amostra de cotação histórica dividida em duas partes, uma para o treinamento do algoritmo e a outra para o experimento em si. Como veremos mais adiante a massa de treinamento será utilizado para se escolher o melhor valor para o parâmetro δ e também para a escolha dos componentes principais. A seguir veremos os detalhes desta implementação que utiliza os mesmos métodos descritos por Montana et al.(2008).

6.1 Dados

Neste trabalho utilizaremos na arbitragem estatística papéis listados na bolsa de valores de São Paulo, a Bovespa. Foi usado o preço de fechamento de 04/06/2008 até 31/08/2011 das 433 empresas listadas na bolsa no início do período. Para os casos de dados faltantes, como dias em que um ou mais papéis não foram negociados, ou simplesmente se uma dada empresa fechou seu capital, foi utilizado o último valor conhecido de cada fluxo.

Todos os dados dos fluxos foram transformados em retornos logaritmos para t maior que 1 e a forma utilizada foi

$$r_t = \log(p_t) - \log(p_{t-1})$$

onde p_t é o preço corrente de um fluxo. Assim temos uma forma de comparar ativos com preços diferentes de maneira direta e com valores próximos de zero.

6.2 Regra de trading

Para esta simulação contamos com investimento inicial de R\$100000,00, chamado de w , e não foram incluídos nos cálculos da estratégia os custos atrelados a negociação como corretagem ou emolumentos.

O papel escolhido para ser o ativo alvo em todas as simulações foi a PETR4, Petrobras PN, apenas por ser uma empresa com um papel de grande liquidez e conseqüentemente por ter negócios em todos os dias de pregão. O número de contratos a serem negociados em cada t é dado por

$$\pi_t = \frac{w}{p_t}$$

onde π_t é uma razão entre o montante w e o preço do ativo alvo p_t no momento t . A função de *spread* é dada por

$$s_t = a_t - r'_t \hat{\beta}_t \quad (26)$$

onde s_t é o *spread*, a_t é o preço atual do ativo alvo onde neste caso terá será o mesmo que p_t e r'_t é o vetor de retornos observados que junto com $\hat{\beta}_t$ estimado pelo FK formam o valor escalar do ativo artificial.

A posição corrente de um dia para outro é dada por

$$\vartheta_t(s_t) = -\text{sign}(s_t) \pi_t$$

a estratégia é muito simples, pois apenas depende do comportamento de reversão a média, podemos ver isso na evolução dos retornos da figura 2, sugerindo que a hipótese é verdadeira. Para calcularmos o retorno de cada momento t foi utilizado o seguinte método

$$f_t = (p_t - p_{t-1}) \vartheta_{t-1}(s_t) \quad (27)$$

6.3 Implementação

Nesta seção será mostrado como foi implementado o modelo e serão apresentados os resultados do experimento.

Primeiramente, os dados foram separados em duas partes, de 04/06/2008 a 02/09/2010 foi usado para o treino do algoritmo de negociação e para medir qual seria o melhor peso a ser usado no parâmetro δ , para isso foi utilizado o índice de Sharpe, que é definido como a razão entre a média dos retornos monetários e seu desvio padrão.

O procedimento usado para se escolher melhor parâmetro δ foi feito da seguinte maneira; evoluindo-se o valor do parâmetro δ de 0,1 até 1, onde δ é constante dentro de cada simulação. Para a estimativa do ativo artificial foram utilizados todos os 433 papéis. Os retornos calculados pela eq.(27) a cada momento t foram armazenados em um vetor $V = (f_t, \dots, f_T)$ que serviu de base para o cálculo do índice de Sharpe de cada simulação. O melhor índice de Sharpe obtido foi 1,129 com δ igual a 0,5.

A tabela 1 mostra o desempenho de cada simulação usando seu respectivo parâmetro δ dentro da base de treino. Para efeito de comparação, foi montada uma estratégia de comprar e manter a posição até o final do período. A compra foi feita utilizando todo montante w ao preço de p_0 e o desmonte da posição foi feita ao preço de p_T e o índice de Sharpe observado para esta estratégia foi de -2,807.

O mesmo período usado para se escolher o peso do parâmetro δ foi usado também para se escolher os principais fluxos de preços através do algoritmo de Weng et al.(2003) e Montana et al.(2008) para a redução da dimensão do vetor de retornos, assim, a partir dos 433 fluxos de dados foi possível chegar a apenas 2 ativos que melhor representam o todo. Tanto os 433 como os 2 fluxos foram usados na massa de dados do experimento de 6/9/2010 a 31/6/2011 onde foi possível observar um aumento da velocidade do processamento e também houve uma melhora no índice de Sharpe quando usados os 2 fluxos em relação aos 433 fluxos.

Com a massa de dados de 6/9/2010 a 31/6/2011, foram feitas três simulações. A primeira foi usar o filtro de Kalman (FK) com os 433 fluxos para construir o ativo

artificial, a segunda utilizando apenas os 2 fluxos encontrados anteriormente, e a terceira foi apenas comprar o ativo e manter a posição até o final do período. Para as duas primeiras simulações δ foi mantido com 0,5 e utilizando a mesma maneira de se avaliar o desempenho do parâmetro δ , o índice de Sharpe para a simulação montada com os 433 fluxos foi de -1,02, a simulação com 2 fluxos teve o desempenho de -0,07 e a estratégia de comprar e manter a posição obteve o índice de -1,08. A evolução do desempenho de cada estratégia pode ser visto na figura 3 que construído a partir da soma acumulada dos retornos da eq.(27) de cada estratégia.

δ	Índice Sharpe
0,1	-0,510
0,2	1,386
0,3	0,774
0,4	1,129
0,5	1,129
0,6	0,815
0,7	0,815
0,8	0,691
0,9	0,255
1	0,946

Tabela 1 - Comparação do Índice de Sharpe

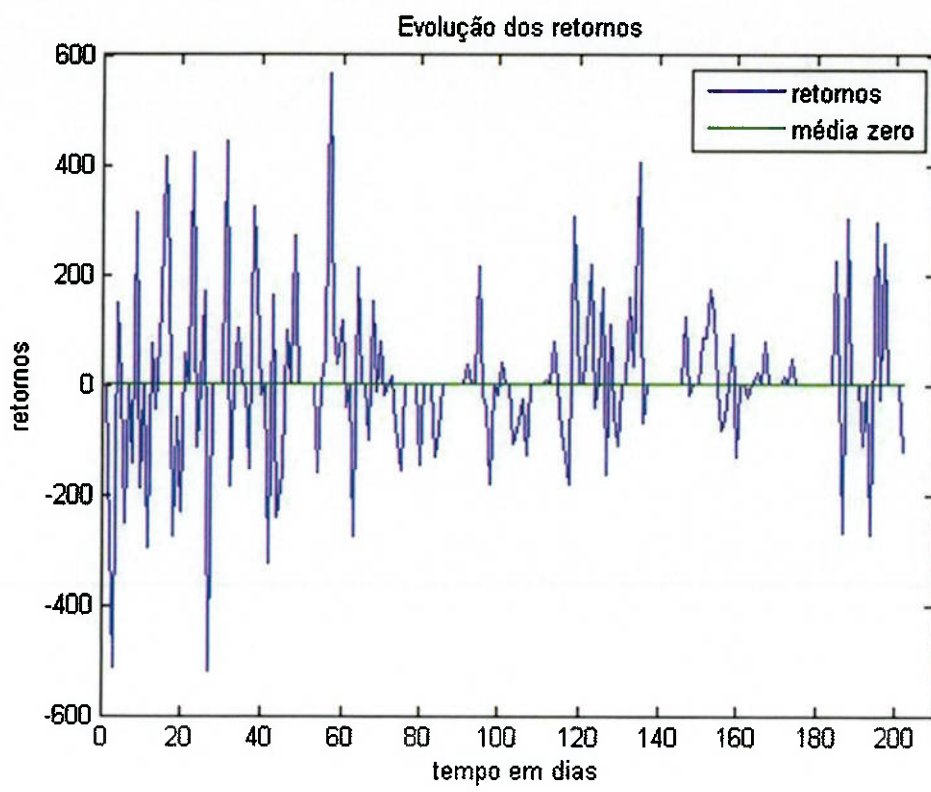


Figura 2- Evolução dos retornos da estratégia

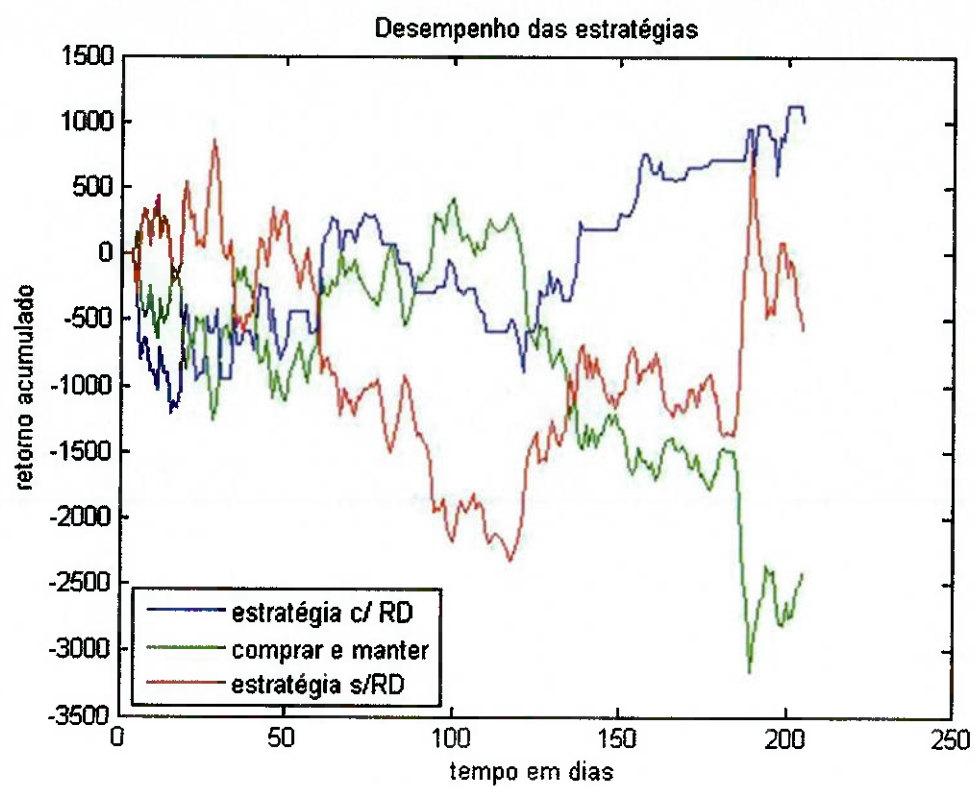


Figura 3 - Evolução do desempenho das estratégias

7 Conclusão

Neste trabalho vimos que é possível construir uma estratégia de arbitragem a partir do MQF interligado com o filtro de Kalman. Também vimos ser possível reduzir a quantidade dos fluxos de cotação sem perder a precisão do *trade*. Tanto a estratégia MQF-FK como o algoritmo para redução de dimensões poderiam ser implementadas como componentes em um sistema de *trading* automatizado como o Streambase stream...(2011).

Nesta simulação não consideramos os custos relacionados à operação como a taxa de corretagem, taxa de custódia e emolumentos. Para um cenário ou simulação mais realista os custos devem ser considerados. Uma maneira de se fazer isso seria modificar as regras de *trading* do capítulo 6 para que o algoritmo modifique sua posição apenas quando o valor do retorno for maior que os custos.

A seleção dos fluxos de dados poderia acontecer de forma automática e dinâmica, não dependendo assim de todos os fluxos de preços da Bovespa, mas sim na similaridade do comportamento do ativo alvo. Uma maneira de se responder a esta questão é recorrer a ferramentas utilizadas em data mining a fim de encontrarmos fluxos de cotação correlacionados. Um exemplo disso é o SPIRIT construído por Sun et al. (2006), um algoritmo que detecta padrões de forma automática a partir de *streams* de dados em tempo real.

Por último, o método poderia ser aplicado a uma carteira com várias ações e um modelo de previsão de mercado poderia ser acoplado à regra de arbitragem, evitando-se assim que mudanças no comportamento dos ativos causem prejuízos à estratégia, como visto em Bollen et al.(2010).

8 Referências

Bloomberg Markets. New York, Nov. 2010, p 110.

HIGH FREQUENCY TRADING REVIEW. Londres. Disponível em <http://www.hftreview.com/pg/blog/mike/read/5390/high-frequency-trading-in-india>. Acesso em: 09 jul. 2011.

KALABA, R.; TEFATSION, L. The flexible least squares approach to time-varying linear regression. *Journal of Economic Dynamics and Control*, v. 12, n. 1, p 43, 1988.

KALABA, R.; TEFATSION, L. Time-varying linear regression via flexible least squares. *Computers & Mathematics with Applications*, v. 17, n. 1, p 1215, 1989.

MONTANA, G; TRIANTAFYLLOPOULOS, K.; TSAGARIS, T. Flexible least squares for temporal data mining and statistical arbitrage *Expert Systems with Applications*, v. 36, p. 2819, 2009.

JAZWINSKI, A. H. *Stochastic Processes and Filtering Theory*. New York, Academic Press, 1970.

WENG, J.; ZHANG, Y. ; HWANG, W. S. Candid covariance-free incremental principal component analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v 25, p 1034, 2003.

STREAMBASE. Lexington, Massachusetts. Disponível em <http://www.streambase.com/algorithmic-trading.htm>. Acesso em: 09 jul. 2011.

Bollen, J.; Mao, H ; Zeng, X. Twitter mood predicts the stock market. *Journal of Computational Science*, v. 2, p. 1, 2011.

SPIRIT: STREAMING PATTERN DISCOVERY. Singapore. Disponível em <http://www.cs.cmu.edu/afs/cs/project/spirit-1/www/>. Acesso em: 09 jul. 2011.