# 3
# Uma metodologia para validação estatística da análise técnica: a busca pela homogeneidade

Este capítulo tem como objetivo apresentar uma solução para as falhas observadas na metodologia utilizada por Lo *et al.* (2000) para validação estatística da análise técnica. Em particular, apresentamos uma proposta de metodologia de identificação prévia de séries homogêneas a fim de agrupá-las para posterior análise estatística.

## 3.1
## A proposta de avaliação da análise técnica de Lo, Mamaysky e Wang

O objetivo desta primeira seção é descrever as propostas de Lo *et al.* (2000) para validação da análise técnica e os problemas estatísticos observados em sua metodologia que podem vir a comprometer seus resultados.

### 3.1.1
### Descrição

Em uma parte de grande importância no trabalho de Lo *et al.* (2000), analisou-se a funcionalidade da análise técnica através de uma verificação de existência de conteúdo informativo nos padrões extraídos das séries de preços. O procedimento adotado por Lo *et al*. (2000) para a validação da análise técnica consistiu em comparar a distribuição de retornos após as formações geométricas identificadas (os *retornos condicionais*) com a distribuição empírica de retornos das séries completas (os *retornos incondicionais*). Uma vez sendo rejeitada a hipótese nula ( $H_0$ ) de que as distribuições são as mesmas, interpreta-se tal resultado como um indicativo da existência de conteúdo informativo dos padrões identificados com a análise técnica.

Este objetivo é realizado através da utilização de testes não paramétricos de aderência, em particular do teste Qui-Quadrado (cf. DeGroot, 1986). A estatística de teste correspondente é definida por:

$$Q = \sum_{i=1}^{10} \frac{\left(Y_i - (0.1)n\right)^2}{(0.1)n}\ , \qquad (3.1)$$

onde $Y_i$ representa o número total de retornos condicionais obtidos que assumem valores entre o i-1-ésimo e o i-ésimo decil da distribuição empírica proveniente dos retornos incondicionais, n é número total global de retornos condicionais e (0.1)n representa a freqüência esperada entre cada decil sob $H_0$.

O teste é realizado sobre os retornos condicionais agrupados por ocorrência de cada padrão específico extraído pela análise técnica. A Figura 15 ilustra, através do gráfico de uma série de preços, as “partes” da série de preços a partir das quais seriam extraídos os retornos condicionais[2]. Analogamente, a série de retornos incondicionais é obtida pelo agrupamento das séries totais de retornos nas quais foram identificados os respectivos padrões.

O objetivo no artigo em questão foi o de comparar tais retornos condicionais agrupados com os retornos totais agrupados (*retornos incondicionais*).

Sob a hipótese nula, a diferença entre o número de retornos condicionais e incondicionais entre cada decil será “pequena”, fornecendo um valor “baixo” para a estatística de teste em (3.1). Rejeitando-se a hipótese nula (a distribuição nula adotada é a Qui-Quadrado usual com 10-1 = 9 graus de liberdade), Lo et al. (2000) concluem que os retornos condicionais vêm de distribuição diferente da dos retornos incondicionais, o que foi interpretado, na ocasião, como indícios estatísticos sobre a existência de conteúdo informativo dos padrões identificados com análise técnica.

[2] Esta parte posterior aos padrões tem um tamanho determinado conforme a prática da análise técnica.

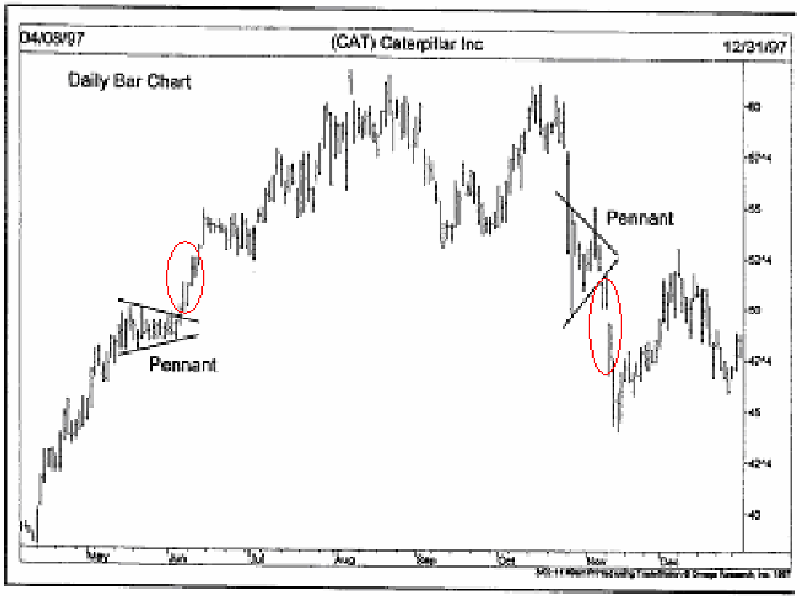


Figura 15: Série de preços com destaque para as figuras (padrões técnicos) em preto e para as séries de preços pós-padrão, circundados em vermelho. O agrupamento das séries de retorno de preços pós-padrão dá origem às séries de retornos condicionais.

### 3.1.2
### Problemas estatísticos

A premissa sobre a qual aqueles autores se basearam para obter suas conclusões possuem incongruências importantes. Com efeito, o teste Qui-Quadrado tem como pressupostos básicos as condições de independência e de mesma distribuição (de forma mais curta, i.i.d) dos dados. Este fato põe em xeque todos os resultados obtidos naquele estudo por dois motivos. O primeiro motivo advém do fato conhecido de que séries financeiras possuem dependência serial considerável nos momentos condicionais de ordens superiores, destacando-se a variância condicional (a *volatilidade condicional* ou, de forma mais simples, *volatilidade*), o que é vastamente difundido na literatura; veja Hamilton (1994), Morettin e Toloi (2004) e Fernandes (2004). Ainda neste quesito, também se observa, principalmente em mercados financeiros cuja eficiência é parcialmente violada, evidências de que os retornos apresentam, ainda que em grau bastante reduzido, correlação serial – em outras palavras, há evidências de efeito AR(1) na média condicional. Esta discussão é didaticamente retomada em Fernandes (2004).

O segundo motivo, e mais importante, está no agrupamento de séries de diferentes ativos, o que fere o pressuposto da homogeneidade. Embora aqueles autores tentassem atenuar a heterogeneidade dos dados padronizando-os, este procedimento não soluciona o problema uma vez que uma suposta heterogeneidade dos momentos de ordem superior não estaria sendo eliminada. O fato é que, todos os demais momentos, caso existissem, estariam sendo negligenciados.

## 3.2
## Metodologia alternativa

Nesta seção, inicialmente, serão apresentadas as técnicas estatísticas utilizadas para a contribuição principal desta dissertação, qual seja, a proposta de uma metodologia que vise à homogeneidade dos ativos escrutinados para validação da análise técnica, via testes Qui-Quadrado.

As técnicas descritas envolvem: a identificação dos processos estocásticos geradores dos retornos dos ativos do tipo AR-GARCH, o agrupamento (ou *análise de cluster*) das séries de ativos semelhantes através da utilização de componentes principais e, por fim, a análise de presença, ou não, de informação advinda dos padrões de preços através da aplicação do teste Qui-Quadrado.

### 3.2.1
### Modelos AR-GARCH

A família de processos AR(1)-GARCH(1,1) descreve de forma realista o processo estocástico de retornos de preços de diversas séries financeiras. Um processo auto-regressivo AR(1) descreve uma dependência linear entre observações sucessivas no tempo dado pela Eq.(3.2), onde $R_t$ é o retorno diário de um ativo e $\eta_t$ é o erro do processo, dado pela Eq.(3.3). O choque GARCH (descrito em um modelo AR(1)-GARCH (1,1) ) assume que a variância condicional $h_t$ dos erros do processo AR(1) são dependentes dos erros passados e das variâncias condicionais destes erros. Esta dependência é descrita pela Eq.(3.4). Este modelo foi escolhido porque um modelo auto-regressivo puro, seja de que ordem for, não reflete a realidade quanto aos retornos provenientes de séries financeiras que sabidamente possuem dependência na variância, apresentando o chamado *clustering* de volatilidade. Tal dependência na variância dos retornos nas séries financeiras são bem capturados por um modelo AR-GARCH. Devido à complexidade de identificação dos modelos GARCH, a literatura recomenda a utilização de modelos de ordem baixa como (1,1), (1,2) ou (2,1). Segue a definição do modelo AR(1)-GARCH(1,1):

$$R_t = \phi_0 + \phi_1 R_{t-1} + \eta_t \tag{3.2}$$

$$\eta_t = h_t^{1/2} \varepsilon_t, \qquad \varepsilon_t \sim NID(0,1) \tag{3.3}$$

$$h_t = \omega_0 + \alpha\, \eta_{t-1}^2 + \beta h_{t-1}, \tag{3.4}$$

com $|\phi_1| < 1$, $\omega_0 > 0$, $\alpha \geq 0$ e $\beta \geq 0$ e $\alpha + \beta \leq 1$.



### 3.2.2
### Análise de componentes principais – Clusterização

A *clusterização* através da *análise de componentes principais* (veja Johnson, 1998; e Johnson e Wichern, 1998) consiste no agrupamento de unidades do experimento com valores de componentes principais próximas. O objetivo é obter-se uma redução na dimensionalidade das variáveis em análise através do cálculo de componentes principais – que serão as novas variáveis – desde que estas consigam representar as variáveis originais sem muita perda de informação. Para reduzir-se de forma relevante a dimensionalidade, deve ser utilizado o menor número possível de componentes principais que representem de forma satisfatória o conjunto das variáveis originais. Este cálculo é feito escolhendo-se as componentes que representam a maior proporção da variância das variáveis originais.

Na formalização desta técnica, verifica-se que seus objetivos são alcançados quando as variáveis são pelo menos moderadamente correlacionadas duas a duas.

Considere que existam n unidades de experimento (ativos) e p variáveis observadas (os coeficientes originais). O vetor de variáveis correspondente a cada unidade de experimento é dado por: $\mathbf{X}_i = (X_{i1},..., X_{ip})'$, i = 1,..., n. Considere também o auto-vetor normalizado da matriz de covariância amostral das variáveis originais, $\mathbf{e}_j = (e_{j1},..., e_{jp})$, j = 1,..., p.

Identificam-se as componentes principais como o produto escalar de cada autovetor da matriz de covariância amostral das variáveis originais com o conjunto de variáveis de cada unidade de experimento. Assim, a j-ésima componente principal da i-ésima unidade de experimento $Y_{ij}$ é obtida pelo produto escalar entre o autovetor $\mathbf{e}_j$ e o vetor $\mathbf{X}_i$ :

$$Y_{ij} = \mathbf{e}'_{\mathbf{j}}\mathbf{X}_{\mathbf{i}} = e_{j1}X_{i1} + \ldots + e_{jp}X_{ip} . \tag{3.5}$$

A variância da j-ésima componente é o autovalor do respectivo autovetor:

$$\hat{V}(Y_j) = \lambda_j , \quad j = 1, \ldots, p . \tag{3.6}$$

### 3.2.3
### Teste Qui-Quadrado

Considera-se a estatística de teste em (3.1) utilizada por Lo *et al.* (2000) com o objetivo de comparar as distribuições empíricas provenientes dos retornos condicionais e incondicionais.

A seleção dos retornos condicionais após cada padrão identificado foi feita obedecendo ao o critério de pertencerem ao objetivo mínimo da figura, conforme descrito na seção 2.6.

## 3.3
## A Metodologia *per si*

A metodologia proposta para a solução do problema da heterogeneidade dos dados da análise de Lo *et al.* (2000) pode ser resumida em 7 passos, quais sejam:

(i) Obtenção de série financeiras de diferentes tipos de ativos.

(ii) Estimativa dos coeficientes do modelo AR(1)-GARCH(1,1) para os ativos analisados.

(iii) Análise de componentes principais dos coeficientes estimados para as séries de retornos dos ativos. Uma vez obtidas as componentes principais, será selecionado o menor numero possível de componentes (de preferência menos que três) que representem a maior parte da *variância total* (isto é, a soma das variâncias amostrais das variáveis originais).

(iv) Agrupamento das unidades amostrais de acordo com a proximidade entre suas componentes principais. O objetivo deste passo está na seleção das unidades amostrais que apresentam evidências de homogeneidade (ou seja, de serem provenientes de uma mesma distribuição empírica).

(v) Após selecionados os ativos homogêneos, identificam-se os padrões geométricos nas séries de preços com a ajuda de uma analista de mercado.

(vi) Os retornos após a formação de padrões específicos, pertencentes ao objetivo mínimo da figura, são selecionados e agrupados sob o título de *retornos condicionais* para cada tipo de padrão. Assim, haverá tantas séries de retornos condicionais quanto o número de padrões distintos encontrados nas séries financeiras. A totalidade dos retornos das séries de ativos das quais foram extraídos e agrupados os retornos condicionais, também são agrupados formando as séries de *retornos incondicionais*.

(vii) Efetua-se o teste Qui-Quadrado, cuja hipótese nula é a de que os retornos condicionais seguem a mesma distribuição dos respectivos retornos incondicionais. Caso a hipótese nula seja negada, infere-se que o padrão específico advindo da análise técnica contém conteúdo informativo.

Na figura 16, a metodologia é resumida em um simples, porém elucidativo, fluxograma.

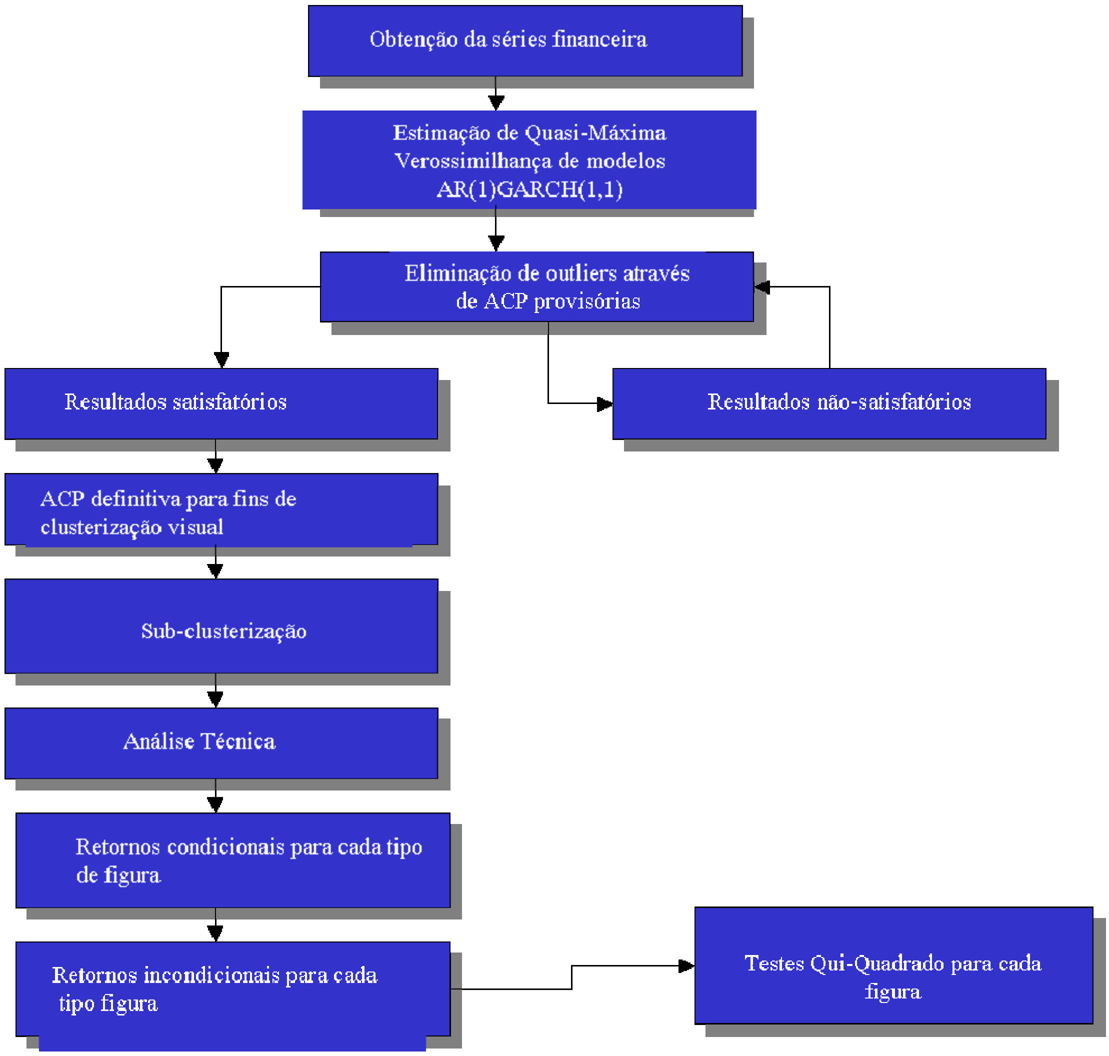


Figura 16: Fluxograma dos passos da metodologia proposta