

US 16 - Empresas de um mesmo segmento oscilam de maneira semelhante dado um período de tempo?

Como investidor gostaria de saber se existe uma correlação ou tendência seguida por algum segmento (considerar 15 segmentos com mais empresas) que mereça ser observada.

1) Definição das cotações

Devido a existência de mais de uma cotação por empresa, definimos os seguintes critérios para seleção de qual cotação utilizar por empresa:

- Selecionar apenas as Cotações do Lote Padrão
- Selecionar apenas as empresas com Cotações no período definido¹
- Selecionar apenas a maior Cotação por Empresa

Após essa filtragem foram selecionadas as empresas pertencentes aos 15 maiores segmentos. A Tabela 1 mostra o segmento e a quantidade de empresas em cada um em ordem decrescente.

Tabela 1. Nº de Empresas por Segmento

Segmento	Nº de Empresas
Energia Elétrica	41
Bancos	26
Construção Civil	20
Fios e Tecidos	17
Material Rodoviário	13
Exploração de Imóveis	11
Outros	11
Holdings Diversificadas	8
Tecidos. Vestuario e Calçados	7
Artefatos de Ferro e Aço	6
Carnes e Derivados	6
Eletrodomesticos	6
Exploracao e/ou Refino	6
Telefonia Fixa	6
Construcao Pesada	5

¹ Foi utilizado todo o período do Real: 1994 até 2013

2) Problema: Gaps nas cotações

Para resolver o problema dos gaps internos em cada cotação (i.e. dias sem dados), todas as cotações foram preenchidas (interpoladas) com o último valor existente (interpolação constante), seguindo a ideia de que caso a cotação não abra o preço se mantém.

3) Adequação ao Contexto de Análise Grafista de Dados Financeiros

Adequamos a análise de correlação de acordo com os estudos no site do [Quantitative Finance](#). As cotações foram suavizadas através da média móvel (*Moving Average*) de 7 dias e calculou-se o Retorno de acordo com a fórmula a seguir (retirada desse [post](#)):

$$R_t = (P_t - P_{t-1}) / P_{t-1} = (P_t / P_{t-1}) - 1$$

Onde P_t é o preço no tempo $t - 1$. Essas foram indicações dadas pelos especialistas em análises grafistas nas respostas a essas duas perguntas do site acima: [Time Series similarity measures](#) e [Correlation between prices or returns?](#).

4) Correlação entre cotações por segmento

Para analisar a similaridade entre as cotações de um mesmo segmento calculou-se a correlação de pearson (por ser a mais comumente utilizada segundo essa [resposta](#)) entre cada par de cotações por segmento, no período de tempo em que ambas existiam (interseção dos períodos). Foram calculadas também as auto-correlações (i.e. uma cotação A com ela mesma), para que fosse possível comparar visualmente com as demais correlações, sendo este um *upper-bound*.

5) Visualizações

A Figura 1 mostra a comparação das correlações agrupada por segmento em boxplots e os pontos (coloridos por segmento) mostram exatamente os valores dos pares de correlação com um *jitter* para evitar que todos fiquem concentrados em um ponto só.

O primeiro gráfico foi gerado com todos os dados de correlações inclusive as auto-correlações (valores de correlação máxima, 1). O segundo gráfico já não possui as auto-correlações, por isso é considerado filtrado. É possível perceber que as caixas dos boxplots tem sua amplitude reduzida. Por fim, o terceiro gráfico permite que tenhamos uma intuição analítica para responder a seguinte pergunta de pesquisa: **As cotações dos 15 maiores segmentos são correlacionadas entre si?**, para isso nós deixamos todas as correlações em uma escala positiva e percebemos que a grande maioria dos boxplots está bem abaixo dos 50%, ou seja, tem uma **Correlação Fraca**.

Para concluir a análise e permitir estudos futuros mais granulares geramos o gráfico de calor mostrando as correlações entre todos os pares de cotações por segmento na Figura 2.

Figura 1: Ver arquivo PDF *Figura 1 - Boxplot da Correlação entre Pares de Cotações (Retorno) por Segmento*.

Figura 2: Ver arquivo PDF *Figura 2 - Heatmap da Correlação entre Cotações (Retorno) por Segmento*.