

# US 7 - O que é um "solavanco"?

## *Sumário*

- [1. Dúvidas e Decisões de Análise](#)
- [2. Análise descritiva das séries](#)
- [3. Métodos de Detecção de Solavancos](#)
  - [3.1. Global Baseline Detector \(quantil limite\)](#)
    - [3.1.1. Análise das Séries](#)
  - [3.2. Local Baseline Detector \(tamanho da janela, quantil limite\)](#)
    - [3.2.1. Análise das Séries](#)
  - [3.3. Long-Term Vision Detector# \(tamanho do horizonte, tamanho da janela de horizontes, quantil limite\)](#)
    - [3.3.1. Análise das Séries](#)

## 1. Dúvidas e Decisões de Análise

- Como lidar com a mudança na moeda de Cruzeiro (CZ\$) para Real (R\$)?
  - Usamos apenas dados do Real
  - Existem cotações com Real a partir de **04/07/1994**
  - Apenas essas cotações são adicionadas no Banco de Dados
- Existem gaps entre os dados das cotações, ou seja, existem dias sem nenhuma cotação.
  - Os gaps serão mantidos na série temporal, vamos considerar como sendo os dias que o Pregão não abre.
  - Consequência: não é possível prever se houve solavanco no(s) dia(s) do gap e no dia imediatamente posterior.
  - Problemas:
    - O que é um gap?
      - O pregão não abriu, o que acontece naturalmente (e.g. final de semana, feriado nacional)
        - Nesse caso, para a predição de solavancos deve-se ignorar os gaps, unindo a série nesses pontos, pois se existe um salto ou queda antes e depois de um dia que o pregão não abriu isso não desconfigura um solavanco, muito pelo contrário, podem ter acontecido notícias nesse intervalo que abalaram a cotação.
        - Quando o pregão não abre pode haver alguma modificação (e.g. no câmbio, etc.) que impacte os dados que analisamos (e.g. mudando a escala dos dados)?
      - Erro na coleta

- Deve manter o gap.
- E se uma empresa não tiver cotação em um dia que o pregão abriu?
  - Isso pode acontecer?
- Etc.
- Os dados de cotacao guardam o preco diario de 7 formas distintas (abaixo), qual usar?
  - preco\_abertura
  - preco\_maximo
  - preco\_minimo
  - preco\_medio
    - Foi selecionado.
    - Como o importante é a predição de solavancos, observando as séries globalmente a escolha de qualquer um desses preços só mudaria a escala do valor. Como o solavanco é medido através dos valores relativos (e.g. hoje comparado com ontem) então a escala tem pouca importância.
  - preco\_ultimo
  - preco\_melhor\_compra
  - preco\_melhor\_venda
- Depois de estudo dos dados com o cliente, ficou definido que serão analisadas apenas as séries temporais das cotações com CODIGO\_BDI igual a 02 ou 96, ou seja, do Lote Padrão ou Lote Fracionário para cada CODIGO\_ISIN existente
  - Apenas as cotações que se encaixam nessa classificação são adicionadas no Banco de Dados
- Detectamos com análise das séries temporais agrupadas por: empresa (através do NOME\_PREGAO), CODIGO\_ISIN e CODIGO\_BDI (02 ou 96); que o formato da série é similar para CODIGO\_BDI diferentes e até para CODIGO\_ISIN diferentes.
  - Essa análise foi feita superficialmente mas caso seja necessário podemos calcular a correlação entre elas
  - Caso seja verdadeiro, alcançamos uma descoberta interessante pela análise dos dados:
    - Todas as cotações relacionadas a determinada empresa tem “sempre” um comportamento semelhante (de acordo com um nível de correlação).
    - Como o objetivo é definir solavancos e quedas, podemos selecionar apenas uma cotação da empresa (um ISIN e um BDI) e estudá-la como sendo representativa de todas as outras cotações da empresa.
    - Isso está de acordo com o “sentimento” do cliente (e do projeto em questão) que afirma existir um padrão nas informações veiculadas em blogs, jornais, etc. relativo a empresa como um todo (não para cada ISIN ou BDI)

## 2. Análise descritiva das séries

Os arquivos PDF da pasta *data/time\_series/describe\_ts* contém os gráficos da análise descritiva realizada para cada uma das 10 séries temporais<sup>1</sup>.

O gráfico da página 1 mostra a série temporal completa, quando há gap o valor não é mostrado. O gráfico da página 2 mostra as diferenças entre valores consecutivos (sem gaps), quando há gap a diferença com o valor anterior e próximo não é exposta. O gráfico da página 3 mostra o histograma das diferenças que conta as diferenças. O gráfico da página 4 mostra a distribuição acumulada das diferenças. E o gráfico da página 5 mostra a distribuição acumulada do valor absoluto das diferenças.

Esses gráficos não são analisados mas podem ser utilizados para compreender as séries e para pensar sobre elas.

### 3. Métodos de Detecção de Solavancos

Todos os métodos proposto são parametrizados (ver palavras entre parênteses nos títulos). Os PDFs da pasta *data/time\_series/solavanco\_ts* contém os gráficos gerados, com os solavancos em vermelho, para cada uma das 10 séries selecionadas.

#### 3.1. Global Baseline Detector (quantil limite)

Um solavanco é definido se a variação entre dois dias com pregão (ou seja, não há gap) for maior do que o parâmetro <quantil limite> da série completa.

Exemplo:

Se o <quantil limite> for 95% então uma dada variação será um solavanco se ela for maior do que 95% de todas as outras variações.

##### 3.1.1. Análise das Séries

O gráfico na 1ª página dos PDFs são desse método. O <quantil limite> foi 95%.

Como pode-se ver os solavancos detectados são apenas nas maiores diferenças. Se observar o gráfico das diferenças (página 2) na descrição de cada série aonde estão os maiores picos estão também os solavancos.

Esse método é útil para análises a longo prazo.

#### 3.2. Local Baseline Detector (tamanho da janela, quantil limite)

Uma janela de dados em uma série temporal é um intervalo de dados. Define-se um <tamanho da janela> e uma variação entre dois dias sem gaps será um solavanco se o valor dela estiver acima do <quantil limite> dos (<tamanho da janela> - 1) dados passados.

Exemplo:

Se o <tamanho da janela> for 10 e o <quantil limite> for 90% então uma variação de ontem para hoje será um solavanco se ela for maior do que 90% das variações dos últimos 9 dias.

---

<sup>1</sup> Os dados das 10 séries temporais estão no arquivo *data/time\_series/ts\_big\_cotacoes.csv*

### 3.2.1. Análise das Séries

O gráfico na 2ª página dos PDFs são desse método. Esse método foi executado com <tamanho da janela> 30 e <quantil limite> 95%.

Os resultados são bem diferentes do método anterior pelo fato da definição ser realizada apenas para os 29 dados anteriores e por isso, a quantidade de solavancos detectados é bem maior.

Esse método é uma evolução do método anterior pois permite que o <tamanho da janela> defina qual o tamanho do passado a ser considerado.

### 3.3. Long-Term Vision Detector<sup>2</sup> (tamanho do horizonte, tamanho da janela de horizontes, quantil limite)

Esse método aplica um algoritmo de compressão na série temporal que comprime mais vezes quanto maior for o <tamanho do horizonte>, quanto maior for o horizonte, maior será a distância entre dois pontos da série comprimida com relação a série inicial. Uma particularidade importante é que o algoritmo não modifica os valores da série para comprimi-la apenas seleciona os valores mais importantes (pontos extremos) capazes de manter seu conteúdo.

Com a série comprimida aplica-se o Local Baseline Detector passando o <tamanho da janela de horizontes> e o <quantil limite> como parâmetros. O resultado será a detecção de solavancos entre dias da série comprimida, para voltar à série original, fazemos a simples consideração de que, se há um solavanco entre dois dias comprimidos existe solavanco entre todos os dias no intervalo.

Assim, o resultado desse método é bem expresso no seu nome pois ele é capaz de detectar solavancos de subida ou descida a longo prazo. Se houve um solavanco de subida, ele permite que haja pequenas descidas, mas que no final a subida seja considerável.

O prazo e a frequência dos solavancos detectados são definidos pelos valores dos parâmetros. Se o <tamanho do horizonte> cresce então esse prazo cresce de acordo. Se o <tamanho da janela de horizontes> cresce então a frequência de solavancos diminui. E o <quantil limite> define o que é um solavanco dentro da janela de valores considerados (mesma interpretação do método *Local Baseline Detector*).

#### 3.3.1. Análise das Séries

O gráfico na 3ª página dos PDFs são desse método. Os parâmetros utilizados foram <tamanho do horizonte> 30, <tamanho da janela de horizontes> 10 e <quantil limite> 95%.

Os resultados mostram uma interpretação diferenciada dos demais métodos pois por analisar a série comprimida pode detectar solavancos desconsiderando oscilações de menor amplitude focando nas oscilações maiores, os solavancos.

---

<sup>2</sup> Método baseado no algoritmo de compressão de séries temporais LMME: [http://www.joics.com/publishedpapers/2013\\_10\\_9\\_2747\\_2756.pdf](http://www.joics.com/publishedpapers/2013_10_9_2747_2756.pdf)

Para entender melhor como funciona o método geramos a Figura “*US 7/solavancos\_ts/10\_SOUZA CRUZ\_BR CRUZACNOR0\_2\_Pontos\_Extremos.png*”<sup>3</sup> com a série 10 em seu estado comprimido até o ano de 200, mostrando os pontos selecionados. Analisando os solavancos para a série 10 e comparando com os pontos da série comprimida veremos que os solavancos detectados são de fato os maiores dentre os 9 últimos (de acordo com o <tamanho da janela de horizontes> - 1).

---

<sup>3</sup> Evitamos adicionar a imagem ao documento para não perder a resolução original.